

Wetenschappelijke verantwoording Entreetoets groep 7

Herman van Boxtel, Wobbe Zijlstra en
Marie-Anne Keizer-Mittelhaëuser



Wetenschappelijke verantwoording

Entreetoets groep 7

Herman van Boxtel

Wobbe Zijlstra

Marie-Anne Keizer-Mittelhaeuser

Cito

Arnhem, 2017

© Cito B.V. Arnhem (2017)

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

Inhoud

1	Inleiding	5
2	Uitgangspunten van de toetsconstructie	9
2.1	Meetpretentie	9
2.2	Doelgroep	10
2.3	Gebruiksdoel en functie	10
2.4	Theoretische inkadering	12
2.4.1	Inhoudelijk	12
2.4.2	Psychometrisch	13
2.4.2.1	Opgavenbanken en constructieprocedures	13
2.4.2.2	Het gehanteerde meetmodel	15
2.5	Constructie en proeftoetsing van opgaven voor de Entreetoets groep 7	18
2.5.1	Het construeren van opgaven	18
2.5.2	De proeftoetsing	19
3	Beschrijving van de toets	29
3.1	Opbouw en structuur	29
3.2	Inhoudsverantwoording	30
3.2.1	De basistoets en de module Verdieping	32
3.2.1.1	Rekenen	32
3.2.1.2	Lezen	37
3.2.1.3	Taalverzorging	46
3.2.2	De module Verbreding	53
3.2.2.1	Luisteren	53
3.2.2.2	Schrijven	57
3.2.2.3	Woordenschat	61
3.2.2.4	Wereldoriëntatie	64
3.3	Kenmerken van de toets: beschrijvende gegevens	68
4	Kalibratie en normering	75
4.1	Kalibratie en het welslagen van de kalibraties op de vaardigheidsschalen	76
4.1.1	Het kalibratieproces: de stappen in de kalibratie	76
4.1.2	Het welslagen van de kalibraties: passing van de opgaven; model-fit	78
4.2	Relatieve normering: toekenning van percentielscores	87
4.2.1	Evaluatie van de representativiteit van het normeringsbestand	88
5	Betrouwbaarheid en meetnauwkeurigheid	93
5.1	Betrouwbaarheid	93
5.2	Nauwkeurigheid: lokale meetfouten	95
5.3	Classificaties	97
5.3.1	Rapportage op referentieniveau: nauwkeurigheid van de classificatie	98
5.3.2	Rapportage van het meest geschikte brugklatype: nauwkeurigheid van de classificatie	100

6	Validiteit	103
6.1	Inhoudsvaliditeit	103
6.2	Begripsvaliditeit	105
6.2.1	Passing van het meetmodel	105
6.2.2	Structuur	106
6.2.3	Soortgenootvaliditeit	111
6.2.4	De Entreetoets groep 7: Relevante verschillen tussen subgroepen	111
6.2.5	Verschillen tussen subgroepen: intercorrelatiematrix en DIF-analyses	114
6.3	Criteriumvaliditeit	118
6.3.1	Het predictieve gebruik van de Entreetoets groep 7 in historisch perspectief	119
6.3.2	Equipercntiel-equivalering	119
6.3.3	Resultaten van het toelatings- en doorstroomonderzoek	119
7	Samenvatting	123
8	Literatuur	129
Bijlage 133		
1	Rechteroverschrijdingskansen van S-toetsen en R1c per toetsonderdeel	134
2	Normtabellen (omzetting van ruwe naar percentielscores)	138

1 Inleiding

Deze wetenschappelijke verantwoording betreft de Entreetoets groep 7. De Entreetoets groep 7 (in deze tekst soms afgekort tot ET7 of ET7 2016) is een onderdeel van het Cito Volgsysteem primair en speciaal onderwijs. De Entreetoetsen voor groep 6 en 7 vormen samen met de LVS-toetsen een krachtig instrument om leerlingen in het reguliere basisonderwijs te volgen.

De Entreetoets groep 7 is ten opzichte van de vorige versie die in 2010 verscheen volledig vernieuwd. De belangrijkste verandering betreft de modulaire opbouw. De toets bestaat nu uit een verplichte basistoets en twee optionele modules, Verdieping en Verbreding. Alle opgaven zijn nieuw geconstrueerd. Dit maakte het mogelijk om deze nieuwe versie aan te laten sluiten op de meest recente ontwikkelingen in het basisonderwijs, zoals de invoering van de referentieniveaus Nederlandse taal en Rekenen.

De *basistoets* is beperkt van omvang en bevat uitsluitend de onderdelen Rekenen, Lezen en Taalverzorging. Door de basistoets af te nemen, krijgt de leerkracht zicht op de mate waarin leerlingen de referentieniveaus Rekenen, Lezen en Taalverzorging behalen zoals deze zijn omschreven in het referentiekader Nederlandse Taal en Rekenen (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009a). Daarnaast kan de leerkracht de prestaties van de leerlingen en de school vergelijken met landelijke normgroepen. Tevens krijgt de leerkracht per leerling via het rapport Vooruitblik een indicatie van het (de) best passende brugklatype(n).

De module *Verdieping* bevat extra opgaven voor de drie onderdelen uit de basistoets. Dit maakt het mogelijk om verdiepende analyses uit te voeren en om voorspelde vaardigheidsscores te gebruiken ter vervanging van de resultaten van LVS-toetsen voor Rekenen-Wiskunde, Begrijpend lezen en Taalverzorging eind groep 7.

De module *Verbreding* bevat taken voor andere vaardigheden. Op de eerste plaats zijn dat de extra taalonderdelen Luisteren, Schrijven en Woordenschat; onderdelen die ook in de vorige versie van de Entreetoets groep 7 waren opgenomen. Ook van het onderdeel Woordenschat is het mogelijk om een voorspelde vaardigheidsscore te gebruiken ter vervanging van de LVS-toets. Volledig nieuw is het onderdeel Wereldoriëntatie, dat de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek omvat. De leerkracht kan zelf bepalen welke onderdelen uit de module Verbreding hij afneemt.

Het onderdeel Studievaardigheden is als zelfstandig onderdeel verdwenen uit de Entreetoets groep 7. Dit is het gevolg van de keuze van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen om onderdelen van Studievaardigheden toe te wijzen aan Lezen (domeinen Opzoeken en Samenvatten) en Rekenen (domein Verbanden). Het onderdeel Kaartlezen is opgenomen bij Wereldoriëntatie in de module Verbreding.

Leerlingen met een vertraagde ontwikkeling kunnen in plaats van de Entreetoets groep 7 de Entreetoets groep 6 maken. Voor deze leerlingen is een aparte taak *Spelling werkwoorden* gemaakt. Door deze taak af te nemen naast de Entreetoets groep 6 is het mogelijk ook voor deze leerlingen een compleet overzicht te verkrijgen van de referentieniveaus Rekenen, Lezen en Taalverzorging en het rapport Vooruitblik.

Materialen

De Entreetoets groep 7 bestaat uit materialen voor de leerkracht, leerling en ouders.

Voor de *leerkracht* is er de leerkrachtmap met daarin de volgende informatie:

- Handleiding
- Stappenplannen voor voorbereiding en interpreteren en analyseren
- Afnamekaarten
- Kennismakingsboekje Entreetoets groep 7
- Bronnenboekje Kennismakingsboekje Entreetoets groep 7
- Antwoordblad en goede antwoorden Kennismakingsboekje Entreetoets groep 7
- Opgavenboekje Basis Entreetoets groep 7

- Bronnenboekje Basis Entreetoets groep 7
- Opgavenboekje Verdieping Entreetoets groep 7
- Bronnenboekje Verdieping Entreetoets groep 7
- Opgavenboekje Verbreding Entreetoets groep 7
- Bronnenboekje Verbreding Entreetoets groep 7
- Ouderfolder

Voor de *leerkracht* staat op Cito Portal:

- Mp3 voor de taken Luisteren en gesproken versie en download tekst naar spraak
- Aanvullende informatie met betrekking tot voorbereiding, afname, interpretatie en analyse van resultaten
- Rapportageprogramma Entreetoetsen (RET)

Voor de *leerling* zijn er de volgende opgavenboekjes, bijbehorende bronnenboekjes en antwoordbladen beschikbaar:

- Kennismakingsboekje Entreetoets groep 7
- Bronnenboekje Kennismakingsboekje Entreetoets groep 7
- Antwoordblad Kennismakingsboekje Entreetoets groep 7
- Opgavenboekje Basis Entreetoets groep 7
- Bronnenboekje Basis Entreetoets groep 7
- Opgavenboekje Verdieping Entreetoets groep 7
- Bronnenboekje Verdieping Entreetoets groep 7
- Opgavenboekje Verbreding Entreetoets groep 7
- Bronnenboekje Verbreding Entreetoets groep 7
- Antwoordblad Entreetoets groep 7

Voor *leerlingen met speciale behoeften* zijn de volgende materialen beschikbaar:

- Audio via mp3 van het kennismakingsboekje, de basistoets en module Verdieping
- Tekst naar spraak van het kennismakingsboekje, de basistoets en module Verdieping
- Zwart-witversie van opgavenboekje en bronnenboekje Verbreding
- Vergrote versie van het kennismakingsboekje, de basistoets, module Verdieping en module Verbreding
- Brailleversie van al het leerlingmateriaal

Voor *leerlingen met een vertraagde ontwikkeling* zijn beschikbaar:

- Opgavenboekjes Regulier en Optioneel Entreetoets groep 6
- Extra taak Spelling-werkwoorden niveau Entreetoets groep 6
- Antwoordblad Entreetoets groep 6 met extra taak Spelling werkwoorden

Om *ouders* te informeren over de Entreetoets is er een ouderfolder beschikbaar.

In deze wetenschappelijke verantwoording ligt het accent op de theoretische en psychometrische verantwoording van de toetsen, inclusief de opzet en de resultaten van de uitgevoerde onderzoeken. Uitgebreide informatie die nodig is om de toetsafname voor te bereiden en de toets af te nemen is te vinden in de handleiding (Cito, 2016). In de handleiding staat ook beschreven hoe leerkrachten, intern begeleiders en schooldirecteuren in het kader van onderwijs op maat de Entreetoets groep 7 kunnen benutten in het cyclische proces van signaleren, analyseren, plannen en handelen.

Deze verantwoording bevat alle informatie die nodig is voor een beoordeling van de kwaliteit van de Entreetoets groep 7. Naast deze verantwoording dient men bovengenoemde producten in de beschouwingen te betrekken.

Het genoemde materiaal maakt een beoordeling van de Entreetoets groep 7 mogelijk op de volgende aspecten:

- Uitgangspunten van de toetsconstructie
- De kwaliteit van het toetsmateriaal
- De kwaliteit van de handleiding
- Normen
- Betrouwbaarheid
- Validiteit

Deze wetenschappelijke verantwoording heeft met name betrekking op de uitgangspunten van de toetsconstructie en verantwoording van de inhoud (hoofdstuk 2 en 3), de normen (hoofdstuk 4), de betrouwbaarheid en meetnauwkeurigheid (hoofdstuk 5) en de validiteit (hoofdstuk 6). Hierbij past de kanttekening dat de Entreetoets niet alleen een beschrijvende functie heeft, maar ook een predictieve. Dit impliceert dat zowel inhouds- en begripsvaliditeit als criteriumvaliditeit van toepassing zijn. De uitgevoerde analyses hebben hoofdzakelijk betrekking op data die in 2016 met dit instrument zijn verzameld; ook de normering voor 2016 en later is gebaseerd op deze data. De kwaliteit van het toetsmateriaal en de handleiding is vast te stellen door kennis te nemen van de inhoud van de overige opgesomde producten.

2 Uitgangspunten van de toetsconstructie

2.1 Meetpretentie

De Entreetoets groep 7 is een zogenoemde school- of leervorderingentoets: de toets meet wat een kind in vergelijking met andere kinderen tot het moment van afname (eind groep 7) in het basisonderwijs geleerd heeft. De inhoud van de Entreetoets sluit daarom nauw aan bij het onderwijsprogramma voor de basisvaardigheden taal, rekenen en wereldoriëntatie. De opgaven zijn een operationalisering van doelstellingen die door het gehele basisonderwijs worden nagestreefd, zogeheten communale doelstellingen.

In vergelijking met de specifieke LVS-toetsen is de Entreetoets groep 7 een brede leervorderingentoets. De toets heeft betrekking op een brede range aan vaardigheden die enerzijds grotendeels overeenkomen met vaardigheden die in specifieke LVS-toetsen worden geoperationaliseerd, anderzijds zoveel mogelijk aansluiten bij wat in de Toetswijzer (zie CvTE, 2012) is beschreven over de eisen waaraan de inhoud van eindtoetsen moet voldoen.

De basistoets vormt het hart van de Entreetoets. De inhoud van de basistoets is geconstrueerd op basis van drie belangrijke elementen in het Referentiekader Nederlandse Taal en Rekenen, te weten Rekenen en de talige onderdelen Lezen en Taalverzorging. Bij Rekenen gaat het om verschillende deelvaardigheden van rekenvaardigheid en getalbegrip, zoals het inzicht in getallen en verhoudingen, het kunnen uitvoeren van operaties met getallen en ruimtelijk inzicht. Deze vaardigheden hebben betrekking op de volgende vier domeinen: Getallen, Meten en meetkunde, Verbanden en Verhoudingen. Bij *Lezen* staan de domeinen Begrijpend lezen, Opzoeken en Samenvatten centraal. Begrijpend lezen gaat over het (kunnen) toekennen van betekenis aan geschreven taal, waarbij de lezer zowel tekst- als kennisgestuurde verwerkingsprocessen in onderlinge wisselwerking toepast. Een goede lezer is in staat doelgericht te zoeken naar informatie door keuze en combinatie van de juiste informatiebronnen en door de structuur van de informatie te doorgronden. Bij Samenvatten gaat het voornamelijk om het structureren van informatie (zoals het onderscheiden van hoofd- en bijzaken, het inkorten en beschikbaar maken van informatie). Bij alle drie deze talige domeinen ligt de focus op tekstbegrip en semantiek. Bij *Taalverzorging* is dit niet of veel minder het geval. Het gaat bij Taalverzorging vooral om het kunnen produceren van een verzorgde tekst, door het correct kunnen toepassen van regels en strategieën. Dat geldt voor alle vier de domeinen die we hierbij kunnen onderscheiden: Grammatica, Interpunctie, Spelling niet-werkwoorden en Spelling werkwoorden. Aangeleerde automatisen spelen hierbij een grote rol.

Door afname van de module Verdieping is het mogelijk op een reeks domeinen gedifferentieerd, objectief en betrouwbaar te rapporteren, gedifferentieerder dan mogelijk is bij afname van alleen de basistoets. Per onderdeel op de basistoets, eventueel aangevuld met Verdieping of Verbreding wordt getoond of de leerlingen de referentieniveaus 1F of 2F/1S hebben behaald. Ook wordt door middel van de Vooruitblik een indicatie gegeven van het best passende brugklatype.

Door afname van de verbredende module is het mogelijk om dit gedetailleerde beeld van de leervorderingen nog wat verder uit te breiden. Uitbreiding is mogelijk naar de talige domeinen Luisteren, Schrijven en Woordenschat. Deze domeinen kennen allen een semantisch karakter; betekenisverlening is bij alle drie van belang. *Luisteren* heeft betrekking op het actieve proces waarbij de luisteraar betekenis toekent aan het gesproken woord. De luisteraar reconstrueert deze betekenis, net als bij begrijpend lezen, door het – in onderlinge wisselwerking – inzetten van zowel tekstgestuurde als kennisgestuurde verwerkingsprocessen. *Schrijven* is een complexe vaardigheid waarbij de schrijver zelf betekenis verleent aan zijn tekst. De schrijver doet dit onder meer door te formuleren, te monitoren en te evalueren, en te reviseren. Hij moet bovendien in staat zijn daarbij rekening te houden met zijn schrijfdoel en het lezerspubliek dat hij voor ogen heeft. *Woordenschat*, ten slotte, verwijst naar de mate waarin taalgebruikers kunnen beschikken over de verzameling labels die zij kunnen inzetten voor het begrijpen en gebruiken van taal. Deze labels (woorden

en woordbetekenissen) zijn ingebed in een begrippennetwerk dat zowel door breedte (kwantitatief) als door diepte (kwalitatief) wordt gekenmerkt.

Verbreding is ook mogelijk in de richting van Wereldoriëntatie, kort te omschrijven als 'kennis van de wereld'. Bij het verwerven van deze kennis doet de leerling een beroep op bepaalde talige en grafische studievaardigheden (zoals het lezen van studieteksten, het opzoeken van informatie, kaartlezen, het lezen van schema's, tabellen, diagrammen en grafieken) en bouwt hij deze vaardigheden verder uit. In de toetsonderdelen met betrekking tot Wereldoriëntatie valt onderscheid te maken naar de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek.

Hoofdstuk 3 geeft een uitgebreide inhoudelijke beschrijving en verantwoording van de gehele toets en alle onderscheiden taken en onderdelen. Daar gaan we ook verder in op de wijze waarop de meetpretentie van de toets is geoperationaliseerd en uitgewerkt.

2.2 Doelgroep

De Entreetoets groep 7 is bedoeld voor leerlingen in groep 7 van het primair en speciaal onderwijs. Voor leerlingen die nog maar pas in Nederland verblijven en de Nederlandse taal onvoldoende beheersen, zijn de toetsen ongeschikt. Voldoende beheersing van het Nederlands is een voorwaarde om de Entreetoets groep 7 te kunnen maken.

Voor leerlingen met een vertraagde ontwikkeling (achterstand van een jaar of meer) is deze toets eveneens minder geschikt. Deze leerlingen kunnen beter de Entreetoets groep 6 maken; deze toets past beter bij hun vaardigheidsniveau. Aanwijzingen voor afname van de toets op het niveau van groep 6 zijn te vinden in de handleiding. Op voorwaarde dat een leerling alle domeinen van een bepaald onderdeel op zijn eigen niveau maakt (bijvoorbeeld alle onderdelen van het onderdeel Lezen in plaats van bijvoorbeeld alleen Begrijpend lezen) wordt een rapportage op het niveau van groep 7 gemaakt. Deze rapportage bevat de scores van de leerling omgezet naar scores die hij zou hebben behaald als hij de Entreetoets groep 7 had gemaakt.

Voor leerlingen met een visuele beperking of dyslexie zijn de volgende aangepaste versies beschikbaar:

- auditieve ondersteuning
- tekst naar spraak
- zwart-wit
- vergroot
- braille

Afname vindt in de regel plaats aan het eind van het schooljaar in de maanden april, mei en juni. Eventueel kan de toets ook onmiddellijk na de zomervakantie in de maand september worden afgenomen als de meeste leerlingen inmiddels in groep 8 zitten. De normering is dan nog van toepassing.

2.3 Gebruiksdoel en functie

Het hoofddoel van de Entreetoets groep 7 is het *beschrijven* van het niveau van leerlingen op het gebied van Rekenen, Lezen en Taalverzorging zoals omschreven in het referentiekader Nederlandse Taal en Rekenen. Met de basistoets kan de leerkracht in kaart brengen in welke mate de leerlingen de referentieniveaus Rekenen, Lezen en Taalverzorging beheersen.

Een nadere analyse op de domeinscores die deel uitmaken van Rekenen, Lezen en Taalverzorging is mogelijk wanneer ook de module Verdieping is afgenomen. De uitbreiding van het aantal opgaven maakt het mogelijk om betrouwbaar scores per domein te rapporteren.

Afname van de module Verdieping maakt het tevens mogelijk om de LVS-toetsen voor Rekenen-Wiskunde, Begrijpend lezen en Taalverzorging voor afnamemoment 'eind groep 7' te vervangen.

De vaardigheidsscores op deze LVS-toetsen zijn afdoende te voorspellen op basis van de opgaven van de basistoets en die van de verdiepende module tesamen. Dit zorgt voor extra functionaliteit die de toetslast

voor leerlingen en leerkracht ten goede komt. De geschatte vaardigheidsscores voor de genoemde toetsen kunnen worden verwerkt via het Computerprogramma LOVS. Op deze manier past afname van de Entreetoets groep 7 ook in de volgfunctie van het Cito Volgsysteem primair en speciaal onderwijs. Het *volgen* van leerlingen in hun groei, ook wel aangeduid als progressiebepaling, is een van de belangrijkste functies van dit volgsysteem. De toetsen van het LVS geven de leerkracht (en ouders en leerlingen zelf) informatie over de ontwikkeling van de vaardigheden van de leerlingen, individueel en als groep, gedurende (vrijwel) de gehele basisschoolperiode. De toetsen geven antwoord op vragen als: is er sprake van vooruitgang, achteruitgang of van stabilisering? Is de vooruitgang – gelet op de gemiddelde vooruitgang in de populatie – volgens verwachting? Door toepassing van een IRT-model (OPLM) is het mogelijk toetsresultaten van leerlingen te vergelijken met toetsresultaten van andere leerlingen op hetzelfde meetmoment en met hun eigen toetsresultaten op eerdere of latere meettijdstippen. Meer informatie over de volgfunctie van LVS-toetsen is te vinden in de wetenschappelijke verantwoordingen van de betreffende toetsen.

Het gedifferentieerde overzicht van de beheersing leervorderingen dat ontstaat bij afname van de verdiepende module kan nog verder worden uitgebreid, en wel door de verbredende module af te nemen. Het overzicht van de leervorderingen kan hiermee worden uitgebreid naar de talige domeinen Luisteren, Schrijven en Woordenschat en daarnaast naar de drie domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek die deel uitmaken van het onderdeel Wereldoriëntatie. Wanneer het onderdeel Woordenschat is afgenomen, is het ook mogelijk om hiervan een vaardigheidsscore op de LVS-toets Woordenschat te schatten.

De beschrijvende functie van de Entreetoets groep 7 wordt ondersteund door normgerichte interpretatie. Met andere woorden, het is mogelijk de toetsresultaten van leerlingen te interpreteren door deze – in de vorm van percentielscores – te vergelijken met die van andere leerlingen. Op het individueel Leerlingprofiel worden deze percentielscores ook uitgedrukt in vaardigheidsniveaus die met Romeinse cijfers I tot en met V worden aangeduid. Elk vaardigheidsniveau betreft een groep leerlingen die 20% van de leerlinpopulatie omvat. Als een score (voor bijvoorbeeld Rekenen) valt in vaardigheidsniveau V, betekent dit dat de betreffende leerling behoort tot de 20% laagst scorende leerlingen op het onderdeel Rekenen.

Met het bovenstaande is de belangrijkste functie van de Entreetoets beschreven, namelijk het op objectieve wijze antwoord te geven op de vraag hoe elke leerling afzonderlijk ervoor staat. Door middel van rapportages is in één oogopslag te zien waar de leerling goed in is en waar ondersteuning nodig is. Daarmee kan de leerkracht hiaten op het spoor komen op een moment dat deze nog kunnen worden 'bijgespijkerd' zo lang het kind nog op de basisschool zit. In de communicatie met de ouders is het van belang een objectief beeld van de leerling te kunnen schetsen.

Door de uitgebreide rapportages op groepsniveau (zie voor een uitgebreid overzicht de handleiding) is het mogelijk het eigen onderwijs te evalueren. Er zijn diverse groepsoverzichten in percentielen, in vaardigheidsscores en naar zo genoemde functioneringsniveaus. Daarnaast wordt er gerapporteerd op referentieniveaus. Deze functie komt het sterkst tot zijn recht in de schoolrapportages die via het digitale Rapportageprogramma EntreeToets (RET) zijn uit te draaien. Deze schoolrapporten zijn een belangrijk hulpmiddel bij schoolzelfevaluatie. Af te lezen is hoe de school presteert ten opzichte van voorgaande jaren. Ook kan men de eigen prestaties vergelijken met die van andere scholen. In de schoolrapporten worden gemiddelden van de school vergeleken met gemiddelden van andere scholen. Omdat rechtstreekse vergelijking niet altijd zonder meer zinvol is, kan men ook kiezen voor gecorrigeerde schoolrapporten. Er zijn schoolrapporten die gecorrigeerd zijn voor leerlinggewicht (sociaal-culturele achtergrond van de leerlingpopulatie). Door op deze manier te corrigeren voor systematische factoren die het onderwijsprestatieniveau van een school sterk beïnvloeden zonder dat de school daar zelf veel aan kan doen, wordt een meer realistische vergelijkingsbasis gecreëerd.

Ten slotte is het mogelijk om de resultaten op de Entreetoets te gebruiken als informatie die een bijdrage kan leveren aan het formuleren van een advies voor het brugklatype dat het beste past bij het niveau van leervorderingen van elke leerling. Dit impliceert dat men, ook al vindt de afname van de toets die dit niveau van leervorderingen moet vaststellen al in groep 7 plaats, een zekere predictieve waarde toekent. Dit naar

analogie van de eindtoets in groep 8. Sinds 2002 is voorspellend gebruik van de Entreetoets mogelijk. Aanvankelijk was dit in de vorm van een pilotproject en sinds 2003 in een meer definitieve vorm. Op basis van de totaalscore op de Entreetoets groep 7 werd een *verwachte standaardscore* op de Cito Eindtoets Basisonderwijs van een jaar later berekend inclusief een onder- en bovengrens waarbinnen de eindtoets-score naar hoge waarschijnlijkheid zou liggen. De onder- en bovengrens geven de onzekerheid in de voorspelling aan: ongeveer 90% van de leerlingen met een bepaalde (specifieke) standaardscore zou op de Eindtoets Basisonderwijs een standaardscore halen die ligt tussen deze grenzen. De voorspellende waarde van de Entreetoets groep 7 werd vastgesteld op basis van onderzoek naar de samenhang van de score op de Entreetoets groep 7 met de score op de eindtoets. Dit onderzoek naar de criteriumvaliditeit kent, hoe waardevol ook, zijn beperkingen. Daarom is met de invoering van deze nieuwe Entreetoets aanvullend onderzoek gedaan, waarin rechtstreeks de samenhang werd onderzocht tussen de Entreetoetsscore en de feitelijke plaatsing in de brugklas bij intrede in het voortgezet onderwijs.

2.4 Theoretische inkadering

2.4.1 Inhoudelijk

Historische achtergrond: van regionale toets tot landelijk instrument

In 1982 is de eerste landelijke Entreetoets uitgebracht in een versie die jarenlang zou worden gebruikt. De aanleiding was de vraag om in het voorlaatste jaar van het basisonderwijs ook een toets te ontwikkelen zoals de Eindtoets. In 1998 kwam een nieuwe Entreetoets voor groep 7 uit, waarna in 2000 en 2012 vernieuwde versies van de Entreetoets groep 6 volgden en in 2002 een Entreetoets voor groep 5. In 2003 en 2010 zagen wederom nieuwe versies van de Entreetoets groep 7 het licht. Deze wetenschappelijke verantwoording heeft betrekking op de versie uit 2016, waarmee het instrument aansluiting vindt bij het referentiekader Nederlandse taal en rekenen, en de predictieve waarde van het instrument nog beter onderbouwd is.

Begripsmatige achtergrond

Een leervorderingen toets als de Entreetoets moet een deugdelijke afspiegeling vormen van de onderwijsdoelstellingen. Cito heeft zich er daarom bij de constructie van opgaven altijd sterk voor gemaakt dat de afstemming op de communale onderwijsinhouden geborgd was. Dit geschiedt door middel van de analyse van doelstellingen en methoden die gehanteerd worden in het basisonderwijs, door het publiceren van inhoudelijke verantwoordingen zoals de inhoudsverantwoording bij de Entreetoets groep 7 en door het betrekken van ervaren leerkrachten in de commissies waarin de constructie van de opgaven plaatsvindt. Deze focus op representativiteit zorgt ervoor dat de toetsinhoud een adequate afspiegeling vormt van de communale onderwijsinhouden en –doelen. Vrij recent is het ‘Referentiekader Taal en Rekenen’ beschreven, zijn referentieniveaus voor het basisonderwijs gedefinieerd en doorlopende leerlijnen opgesteld. Daarmee is (opnieuw) vastgelegd wat leerlingen op het einde van het basisonderwijs moeten kennen en kunnen op het gebied van taal en rekenen. Deze versie van de Entreetoets sluit aan bij dit referentiekader en is daarmee up to date.

Bovengenoemde toetsinhoud is gestructureerd in de vorm van opgavenbanken voor de onderscheiden basisvaardigheden en – voor zover nodig – voor de daarbinnen onderscheiden domeinen. De opgavenbanken hebben door de toegepaste technologie in de vorm van itemresponsmodellen zoals OPLM een sterk unidimensioneel karakter (zie hierover verderop in dit hoofdstuk meer).

In hoofdstuk 3 volgt een uitgebreide inhoudsverantwoording van de samenstelling van de Entreetoets groep 7. Voor nu kunnen we volstaan met de conclusie dat zowel de domeinen van de Entreetoetsen als de LVS-toetsen gebruikmaken van, naar inhoud en meetmodel, zorgvuldig samengestelde opgavenbanken. Er is een sterke inhoudelijke afstemming tussen de verschillende instrumenten die van het Cito Volgstelsel primair en speciaal onderwijs, waardoor dit een zeer consistent geheel vormt. Binnen dat geheel hebben de LVS-toetsen vooral een beschrijvende en volgende functie, terwijl de functie van de Entreetoets groep 7 naast beschrijvend ook een voorspellende is. Voor alle functies is het even

noodzakelijk dat representativiteit ten aanzien van communale onderwijsdoelen en –inhouden gewaarborgd is en dat de onderscheiden scores en onderdelen verwijzen naar unidimensionele vaardigheidsdimensies. Voor een beschrijving van de onderdelen en de domeinen, en de inhoudelijke verantwoording ervan, verwijzen we naar hoofdstuk 3.

2.4.2 Psychometrisch

In deze paragraaf gaan we allereerst in op de procedures die Cito bij de constructie van de Entreetoets hanteert. Eerst bespreken we in paragraaf 2.4.2.1 de eigenschappen van opgavenbanken en de wijze waarop deze worden geconstrueerd. Hierbij spelen IRT-meetmodellen een belangrijke rol. In paragraaf 2.4.2.2 wordt uitvoerig op deze meetmodellen ingegaan.

2.4.2.1 Opgavenbanken en constructieprocedures

Bij de constructie van opgaven wordt in de regel een veelvoud geproeftoetst van het aantal dat uiteindelijk in de toets moet worden ingezet. Er moet immers rekening worden gehouden met uitval, bijvoorbeeld wegens meer of minder triviale fouten in de constructie of extreme moeilijkheid of gemakkelijkheid. Een nieuwe toets (in dit geval de Entreetoets groep 7) wordt samengesteld uit een aantal nieuw geproeftoetste opgaven en soms ook uit opgaven die al eerder op hun deugdelijkheid zijn onderzocht. De verzameling opgaven waaruit constructeurs een keuze kunnen maken bij de samenstelling van een toets wordt opgavenbank genoemd.

Een belangrijk kenmerk van zo'n opgavenbank is dat deze gekalibreerd is met een unidimensionaal IRT-model (meestal OPLM; Verhelst en Eggen, 1989; zie verder paragraaf 2.4.2.2), waardoor niet alleen de psychometrische kenmerken (parameters) van de opgaven worden geschat, maar waarbij tevens wordt nagegaan of de opgaven van een onderdeel kunnen worden beschreven met een unidimensionale onderliggende vaardigheid. Een gedetailleerde verantwoording van deze werkwijze voor bijvoorbeeld het onderdeel Taal is te vinden in Staphorsius, Verhelst en Kleintjes (2001) en in Staphorsius, Krom, Kleintjes en Verhelst (2001). Ook Staphorsius en Verhelst (2001) geven een uitvoerige beschrijving van de gevolgde werkwijze, in dit geval voor de onderdelen van de destijds uitgegeven Toelatingstoets, die op dezelfde manier zijn gekalibreerd. In paragraaf 2.5 zullen we stilstaan bij een aantal belangrijke aspecten van de opgavenconstructiefase voor de Entreetoets. In het navolgende gaan we eerst in op enkele algemene principes bij de constructie van opgavenbanken en (in paragraaf 2.4.2.2) op de eigenschappen van het gehanteerde meetmodel.

Een opgavenbank is nadrukkelijk niet eenvoudigweg een verzameling opgaven of items waaruit een toetsconstructeur min of meer naar willekeur een aantal items selecteert om een nieuwe toets te construeren. Hieronder wordt beschreven wat de vereisten zijn om van een deugdelijke en psychometrisch goed gefundeerde opgavenbank te kunnen spreken.

– *Unidimensionaal continuüm en latente vaardigheid*

Het algemene uitgangspunt is dat de vaardigheden die in de onderdelen en onderliggende domeinen van de Entreetoets worden gemeten kunnen worden opgevat als een unidimensionaal continuüm (de reële lijn), en dat elke leerling voorgesteld kan worden als een punt op die lijn, met andere woorden: als een getal. Het getal drukt de mate van vaardigheid uit, waarbij een groter getal wijst op een grotere vaardigheid. Het doel van de meetprocedure – het afnemen van de toets – is de plaats van de leerling op dit continuüm zo nauwkeurig mogelijk te bepalen. De uitkomst van de meetprocedure bestaat strikt genomen uit twee grootheden: de eerste is de schatting van de plaats van de leerling op het vaardigheidscontinuüm. De tweede grootheid geeft aan hoe nauwkeurig die schatting is, en heeft dus de status van een standaardfout, te vergelijken met de standaardmeetfout uit de klassieke testtheorie. De antwoorden van een leerling op de items worden beschouwd als indicatoren van de vaardigheid, hetgeen ruwweg betekent dat men verwacht dat alle items in de bank deze zelfde vaardigheid meten. De vaardigheid zelf wordt als niet-observeerbaar beschouwd, en daarom gewoonlijk omschreven als een latente vaardigheid.

– *'Moeilijkheid' in de Item Respons Theorie*

Hoewel items dezelfde vaardigheid meten, kunnen ze toch systematisch van elkaar verschillen. Het belangrijkste verschil tussen de items is hun moeilijkheidsgraad. In de klassieke testtheorie wordt de moeilijkheidsgraad uitgedrukt met een zogenaamde p-waarde, de proportie correcte antwoorden op het item in een welbepaalde populatie van leerlingen. In de Item Respons Theorie (IRT) die voor het construeren van de opgavenbanken wordt gebruikt, hanteert men echter een andere definitie van moeilijkheid: ruwweg gesproken is het de mate van vaardigheid die nodig is om het item goed te kunnen beantwoorden. Dit verschil in definitie van de moeilijkheidsgraad tussen klassieke theorie en IRT is uitermate belangrijk: men kan verwachten dat de p-waarde van een item in groep 7 groter zal zijn dan in groep 6, waardoor duidelijk wordt dat de p-waarde een relatief begrip is: ze geeft de moeilijkheid aan van een item in een bepaalde populatie. Binnen de IRT is de moeilijkheid van een item gedefinieerd in termen van de onderliggende vaardigheid, zonder enige verwijzing naar een bepaalde populatie van leerlingen. Zo kan men ook de uitspraak begrijpen dat in de IRT vaardigheid en moeilijkheid op eenzelfde schaal liggen.

– *Kansmodel*

De ruwe omschrijving van de moeilijkheidsgraad die in de vorige alinea werd gehanteerd (de mate van vaardigheid die nodig is om het item goed te kunnen beantwoorden) heeft enige verdere uitwerking. Men zou deze omschrijving kunnen opvatten als een drempel: heeft een leerling die mate van vaardigheid niet, dan is hij niet in staat het item juist te beantwoorden; heeft hij die drempel wel gehaald, dan geeft hij (gegarandeerd) het juiste antwoord. Deze interpretatie weerspiegelt een deterministische kijk op het antwoordgedrag van de leerling, die echter in de praktijk geen stand houdt, omdat eruit volgt dat een leerling die een moeilijk item correct beantwoordt geen fout kan maken op een gemakkelijk(er) item. Daarom wordt in de IRT een kansmodel gebruikt: hoe groter de vaardigheid, des te groter de kans dat een item juist wordt beantwoord. De moeilijkheidsgraad van een item wordt dan gedefinieerd als de mate van vaardigheid die nodig is om met een kans van precies een half, een juist antwoord te kunnen produceren (zie verder ook de volgende paragraaf over het meetmodel).

– *Kalibratie*

In het voorgaande zijn nogal wat veronderstellingen ingevoerd (unidimensionaliteit; alle items zijn indicatoren voor dezelfde vaardigheid; kansmodel) die niet zonder meer voor waar kunnen worden aangenomen; er moet aangetoond worden dat al die veronderstellingen deugdelijk zijn. Dit 'aantonen' gebeurt met statistische gereedschappen waarop later nog dieper wordt ingegaan. Maar vóór de items in een toets gebruikt kunnen worden, moet ook geprobeerd worden de waarden van de moeilijkheidsgraden te achterhalen. Dit gebeurt met een statistische schattingsmethode die wordt toegepast op de itemantwoorden die bij een steekproef van leerlingen zijn verzameld. Het hele proces van moeilijkheidsgraden schatten en verifiëren of de modelveronderstellingen houdbaar zijn, wordt kalibratie of ijking genoemd. De steekproef van leerlingen die hiervoor wordt gebruikt (bijvoorbeeld bij het proeftoetsen van opgaven in de constructiefase) heet kalibratiesteekproef.

– *Afnamedesigns*

Meestal bevat een opgavenbank meer items dan een doorsnee taak of set taken, zodat het praktisch niet doenlijk is om alle items aan alle leerlingen voor te leggen. Elke leerling in de kalibratiesteekproef krijgt derhalve slechts een (klein) gedeelte van de items uit de opgavenbank voorgelegd. Er is dan sprake van een zogenoemd onvolledig design. Dit gedeeltelijk voorleggen moet met de nodige omzichtigheid gebeuren wil men iets zinnigs kunnen zeggen over de itemkenmerken in hun onderlinge verband. Of, anders gezegd, de informatie over afzonderlijke (groepen) items moet men met elkaar kunnen verbinden. Voor meer informatie over afnamedesigns die voor de kalibratie kunnen worden gebruikt, wordt de geïnteresseerde lezer verwezen naar Eggen (1993). Afnamedesigns kunnen ook benut worden om andere doelstellingen te realiseren, zoals het overbrengen van cesuren voor het bepalen van referentieniveaus en het equivaleren van toetsversies (zie paragraaf 2.5).

– *Implicaties van gekalibreerde opgavenverzameling*

Als de kalibratie met succes uitgevoerd is, is het resultaat een zogenoemde gekalibreerde itembank. In het kalibratieproces worden de items die niet passen bij de verzameling uit de collectie verwijderd. De opgavenbank bevat voor elk item niet alleen zijn feitelijke inhoud, maar ook zijn psychometrische eigenschappen, en de statistische zekerheid dat alle items dezelfde vaardigheid aanspreken.

Meer over de kalibreringsprocedure en een bespreking van de resultaten daarvan voor de Entreetoets groep 7 is te vinden in hoofdstuk 4 over de normering van de toets.

2.4.2.2 Het gehanteerde meetmodel

In de Entreetoets groep 7 is gebruikgemaakt van een op de itemresponstheorie (IRT) gebaseerd meetmodel zoals dat bij Cito gebruikelijk is, namelijk van het One Parameter Logistic Model (OPLM). Wij zullen dit model hieronder op hoofdlijnen bespreken.

OPLM: het One Parameter Logistic Model

IRT-modellen verschillen in een aantal opzichten nogal sterk van de klassieke testtheorie (Verhelst, 1993; Verhelst & Kleintjes, 1993; Verhelst & Glas, 1995). Bij de klassieke testtheorie staan de toets en de toetsscore centraal. Het theoretisch belangrijkste begrip in deze theorie is de zogenoemde ware score, de gemiddelde score die de persoon zou behalen indien de test een oneindig aantal keren onder dezelfde condities zou worden afgenomen. In de IRT staat het te meten begrip of de te meten vaardigheid centraal. IRT-modellen hebben belangrijke voordelen boven de klassieke testtheorie. Zo is het bijvoorbeeld mogelijk in de toetsconstructie te werken met een onvolledig design en kunnen item- en populatieparameters onafhankelijk van elkaar worden geschat (voor een overzicht van de voordelen van IRT-modellen boven de klassieke testtheorie verwijzen we naar Hambleton, Swaminathan en Rogers, 1991).

De IRT beschouwt het antwoord op een item als een indicator voor de mate waarin die vaardigheid aanwezig is. Het verband tussen vaardigheid en itemantwoord is van probabilistische aard en wordt weergegeven in de zogenoemde itemresponsfunctie. Die geeft aan hoe groot de kans is op een correct antwoord als functie van de onderliggende vaardigheid. Formeler: zij X_i de toevalsvariabele die het antwoord op item i voorstelt. X_i neemt de waarde 1 aan in geval van een correct antwoord en 0 in geval van een fout antwoord. Als symbool voor de vaardigheid wordt θ (theta) gekozen. De vaardigheid θ is niet rechtstreeks observeerbaar. Dat zijn alleen de antwoorden op de opgaven. Dat is de reden waarom θ een 'latente' variabele wordt genoemd¹. De item-responsfunctie $f_i(\theta)$ is gedefinieerd als een conditionele kans:

$$f_i(\theta) = P(X_i = 1 | \theta) \quad (2.1)$$

Een IRT-model is een speciale toepassing van (2.1) waarbij aan de functie $f_i(\theta)$ een meer of minder specifieke functionele vorm wordt toegekend. Een eenvoudig en zeer populair voorbeeld is het zogeheten Raschmodel (Rasch, 1960) waarin $f_i(\theta)$ gegeven is door

$$f_i(\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (2.2)$$

waarin β_i de moeilijkheidsparameter van item i is. Dat is een onbekende grootte die geschat wordt uit de observaties. De grafiek van (2.2) is weergegeven in figuur 2.1 voor twee items, i en j , die in moeilijkheid verschillen.

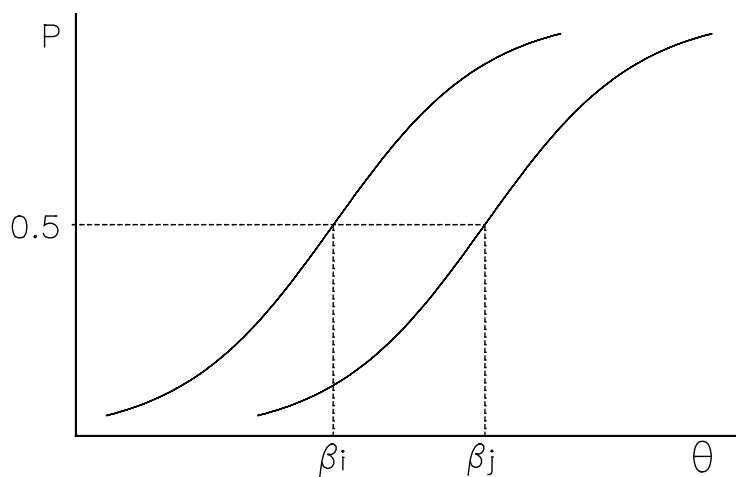
¹ Dit maakt duidelijk waarom men de modellen die ressorteren onder de IRT, ook wel aanduidt met 'latente trek'-modellen.

Deze figuur illustreert dat de itemresponsfunctie een stijgende functie is van θ : hoe groter de vaardigheid, des te groter de kans op een juist antwoord. Indien de latente vaardigheid precies gelijk is aan de moeilijkheidsparameter β_i , volgt

$$f_i(\beta_i) = \frac{\exp(\beta_i - \beta_i)}{1 + \exp(\beta_i - \beta_i)} = \frac{1}{1 + 1} = \frac{1}{2} \quad (2.3)$$

Daaruit volgt onmiddellijk een interpretatie voor de parameter β_i : het is de 'hoeveelheid' vaardigheid die nodig is voor de kans van precies een half om het item i juist te beantwoorden. Uit de figuur blijkt duidelijk dat voor item j een grotere vaardigheid nodig is om diezelfde kans te bereiken, maar dit is hetzelfde als te zeggen dat item j moeilijker is dan item i . De parameter β_j kan dus terecht omschreven worden als de moeilijkheidsparameter van item j . De implicatie van het bovenstaande is dat 'moeilijkheid' en 'vaardigheid' op dezelfde schaal liggen.

Figuur 2.1 Twee itemresponscurven in het Raschmodel



Formule (2.2) is geen beschrijving van de werkelijkheid, het is een hypothese over de werkelijkheid die getoetst kan worden op haar houdbaarheid. Hoe zo'n toetsing grofweg verloopt, is te verduidelijken aan de hand van figuur 2.1. Daaruit blijkt dat, voor welk vaardigheidsniveau dan ook, de kans om item j juist te beantwoorden steeds kleiner is dan de kans op een juist antwoord op item i . Hieruit volgt de statistisch te toetsen voorspelling dat de verwachte proportie juiste antwoorden op item j kleiner is dan op item i in een willekeurige steekproef van personen. Splitst men nu een grote steekproef in bijvoorbeeld twee deelsteekproeven, een 'laaggroep', met de vijftig procent laagste scores, en een 'hooggroep', met de vijftig procent hoogste scores, dan kan men nagaan of de geobserveerde p-waarden van de opgaven in beide deelsteekproeven op dezelfde wijze geordend zijn. Daarvan kan strikt genomen alleen sprake zijn als, in termen van de klassieke testtheorie uitgedrukt, alle opgaven eenzelfde discriminatie-index hebben. Dat echter blijkt lang niet altijd zo te zijn. Ook in ons geval niet. Veel van de items blijken dan ook niet beschreven te kunnen worden met het Raschmodel. Daarom is bij de Entreetoets groep 7 gekozen voor een ander IRT-model.

Alvorens dit bij de Entreetoets groep 7 voornamelijk gebruikte model te introduceren, is eerst een kanttekening nodig bij het schatten van de moeilijkheidsparameters in het Raschmodel. Een vaak toegepaste schattingsmethode is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: *Conditional Maximum Likelihood*, verder aangeduid als CML). Die maakt gebruik van het feit dat in het Raschmodel een afdoende steekproefgrootte (*sufficient statistic*) bestaat voor de latente variabele θ , namelijk de ruwe score of het aantal correct beantwoorde items. Dat betekent grofweg dat, indien de itemparameters bekend zijn, alle informatie die het antwoordpatroon over de vaardigheid bevat, kan worden samengevat in de ruwe score. Het doet er dan verder niet meer toe welke opgaven goed en welke fout zijn gemaakt. Hieruit vloeit voort dat de conditionele

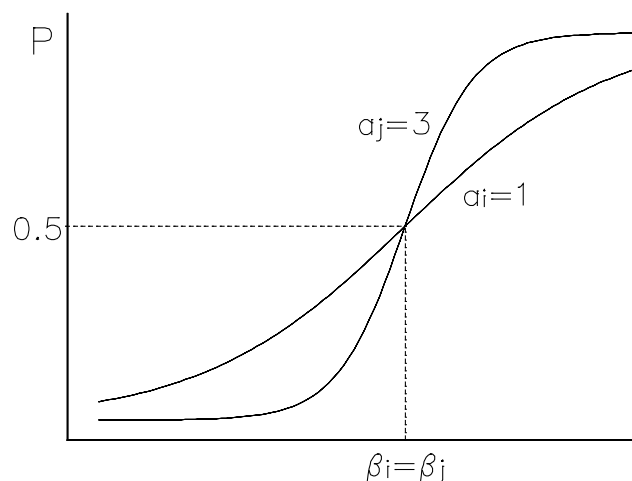
kans op een juist antwoord op item i , gegeven de ruwe score, een functie is die alleen afhankelijk is van de itemparameters en onafhankelijk van de waarde van θ . Een gedetailleerde uiteenzetting hierover is te vinden in Verhelst (1992.). De CML-schattingmethode maakt van deze functie gebruik. Deze methode maakt geen enkele veronderstelling over de verdeling van de vaardigheid in de populatie, en is ook onafhankelijk van de wijze waarop de steekproef is getrokken.

De CML-schattingmethode is echter niet bij elk meetmodel toepasbaar. In het éénparameter logistisch model (One Parameter Logistic Model, afgekort: OPLM) is CML mogelijk. Dit model is, anders dan het Raschmodel, wel bestand tegen 'omwisseling' van 'proporties juist' in verschillende steekproeven (Glas & Verhelst, 1993; Eggen, 1993; Verhelst & Kleintjes, 1993). De itemresponsfunctie van het OPLM is gegeven door

$$f_i(\theta) = \frac{\exp [a_i(\theta - \beta_i)]}{1 + \exp [a_i(\theta - \beta_i)]}, \quad (2.4)$$

waarin a_i de zogeheten discriminatie-index van het item is. Door de waarden voor discriminatie-indices te beperken tot (positieve) gehele getallen, en door ze a priori als constanten in te voeren, is het mogelijk CML-schattingen van de itemparameters β_i te maken. In figuur 2.2 is de itemresponscurve weergegeven van twee items i en j , die even moeilijk zijn maar verschillend discrimineren.

Figuur 2.2 Twee itemresponscurven in het OPLM: zelfde moeilijkheid, verschillende discriminatie-index



De schattingen worden berekend met het computerprogramma OPLM (Verhelst, Glas en Verstralen, 1995). Dit programma voert ook statistische toetsen uit op grond waarvan kan worden bepaald of het model de gegevens adequaat beschrijft. Omdat een aantal van deze toetsen bijzonder gevoelig is voor een verkeerde specificatie van de discriminatie-indices, zijn de uitkomsten van deze toetsen bruikbaar als modificatie-indices: ze geven een aanwijzing in welke richting deze discriminatie-indices moeten worden aangepast om een betere overeenkomst tussen model en gegevens te verkrijgen. Kalibratie van items volgens het OPLM is dan ook een iteratief proces waarin alternerend de modelfit van items wordt onderzocht door middel van statistische toetsing en de waarden van de discriminatie-indices worden aangepast op grond van de resultaten van deze toetsen. Deze aanpassingen geschieden in de praktijk op basis van een en hetzelfde gegevensbestand. Er kan dus kanskapitalisatie optreden. Indien een steekproef een voldoende grootte heeft, is het effect van deze kanskapitalisatie echter gering (Verhelst, Verstralen en Eggen, 1991).

Hoewel het OPLM aanzienlijk flexibeler is dan het Raschmodel, heeft het met dit model toch een nadeel gemeen, waardoor het bij het kalibreren van meerkeuzeopgaven niet zonder meer bruikbaar is. Uit de formules (2.2) en (2.4) volgt dat, indien θ zeer klein is, de kans op een juist antwoord zeer dicht in de buurt van nul komt.

Maar de items in het normeringsonderzoek zijn meerkeuze-items, zodat blind gokken een zekere kans op een juist antwoord impliceert. Er bestaan modellen die rekening houden met de raatkans (Lord & Novick, 1968), maar die laten geen CML-schattingsmethode toe. De ongeschiktheid van het Raschmodel of OPLM voor meerkeuzevragen is echter relatief: indien de items in vergelijking met de vaardigheid van de leerling niet al te moeilijk zijn, blijkt dat het effect van het raden op de overeenkomst tussen model en gegevens klein is. Door een verstandige dataverzamelingsprocedure toe te passen en met name niet te moeilijke opgaven te selecteren in de toets kan het OPLM toch toegepast worden op meerkeuzevragen, waarbij de overeenkomst tussen model en data de uiteindelijke doorslag over die geschiktheid moet geven. Ook in de normering wordt hiermee rekening gehouden.

Voor de schatting van parameters van de populatieverdeling wordt gebruik gemaakt van de 'marginale grootste aannemelijkheidsmethode' (in het Engels: *Marginal Maximum Likelihood*, verder afgekort als MML). Deze schattingsmethode veronderstelt naast (2.2) ook nog dat de vaardigheid θ in de populatie een bepaalde verdeling heeft. De meeste computerprogramma's die IRT-analyses kunnen uitvoeren, veronderstellen een normale verdeling. Bovendien stelt deze methode de voorwaarde dat de steekproef die voor de schatting gebruikt wordt uit die verdeling een aselechte steekproef is.

2.5 Constructie en proeftoetsing van opgaven voor de Entreetoets groep 7

In deze paragraaf besteden we aandacht aan de wijze waarop de opgaven werden geconstrueerd en de rol die de proeftoetsing daarbij heeft gespeeld. Allereerst bespreken we de opgavenconstructie zelf (paragraaf 2.5.1). Vervolgens gaan we (in paragraaf 2.5.2) uitgebreid in op de proeftoetsing. De proeftoetsing is om diverse redenen belangrijk. Niet alleen kan de kwaliteit van de opgaven daarmee empirisch worden onderzocht, maar ook kan – door de keuze voor een weldoordacht design – ankering tot stand worden gebracht met andere toetsen, waaronder de oude Entreetoets groep 7. Dit laatste is belangrijk om equivalentie tot stand te brengen tussen de oude en de hier verantwoorde nieuwe versie van de Entreetoets groep 7. Daarnaast is het proeftoetsonderzoek benut om de cesuren die verbonden zijn met de referentieniveaus in de referentiesets over te brengen op de toets. Vanwege de grote aantallen uit te testen opgaven is een onvolledig, maar verbonden design noodzakelijk. We lichten zo'n design toe aan de hand van een voorbeeld. Voorts geven we inzicht in de aantallen ankeropgaven die in de proeftoets werden opgenomen. Hoewel representativiteit geen belangrijk onderwerp is voor een constructieonderzoek, besteden we daar toch enige aandacht aan (men name omdat de deelnamebereidheid van scholen gering is en de nonresponse daardoor omvangrijk). Ten slotte geven we inzicht in de aantallen leerlingen die de opgaven hebben gemaakt.

2.5.1 Het construeren van opgaven

In 2014 is gestart met de voorbereidingen voor de hier verantwoorde Entreetoets groep 7. Op basis van de gespecificeerde gebruiksdoelen en functies van de toets (zie paragraaf 2.3) en de voorgenomen modulaire opzet van de toets werd een planning gemaakt voor de aantallen te construeren en te proeftoetsen opgaven. Daarbij werd rekening gehouden met een mortaliteit die per type onderdeel en vaardigheid kan verschillen. Met mortaliteit wordt hier bedoeld dat in de fase van proeftoetsing altijd opgaven verloren gaan, omdat zij niet blijken te voldoen, hetzij aan psychometrische kwaliteitseisen, hetzij aan vooraf vastgelegde inhoudelijke criteria. Daarnaast kunnen opgaven ook afvallen vanwege bezwaren die door praktijkdeskundigen naar voren worden gebracht. Schattingen van de mortaliteit zijn gebaseerd op ervaringen uit het verleden. Zo werd voor bijvoorbeeld Begrijpend lezen een mortaliteit van drie aangehouden, terwijl voor Spelling-opgaven kon worden volstaan met een mortaliteit van twee. Voor Begrijpend lezen werden dus driemaal zoveel opgaven geconstrueerd als er uiteindelijk nodig zijn, voor Spelling tweemaal zoveel. De benodigde aantallen opgaven werden geconstrueerd in zogeheten constructiegroepen (afgekort tot 'cg') die per onderdeel of domein werden samengesteld. Constructiegroepen bestaan uit drie à vier mensen 'uit het veld': leerkrachten van groep 7 of 8, PABO-docenten en/of IB'ers, onder leiding van een toetsdeskundige van Cito. De afzonderlijke cg-leden leveren conceptopgaven en uitgangsmateriaal (teksten, afbeeldingen) aan, die door de andere leden besproken en becommentarieerd worden in speciale

bijeenkomsten. Soms gaat aan de bespreking eerst een schriftelijke commentaarronde vooraf. De meeste opgaven komen van mensen uit de praktijk van het basisonderwijs, maar ook de toetsdeskundigen van Cito leveren soms opgaven aan. Hun belangrijkste taak is echter om vanuit hun expertise op vakinhoudelijk en toetstechnisch gebied commentaar te geven op de geconstrueerde opgaven. Ook kan hij of zij de opgaven aanpassen als de commentaarronde of de bespreking daartoe aanleiding geeft. De opgaven die (eventueel in aangepaste vorm) worden goedgekeurd, worden (voorlopig) opgenomen in een opgavenbank. In totaal moesten voor alle onderdelen en domeinen van de proeftoets samen 1174 nieuwe opgaven worden geconstrueerd. Al deze opgaven werden in empirisch onderzoek (de zogeheten proeftoetsing) empirisch onderzocht.

2.5.2 De proeftoetsing

De proeftoetsing vond plaats in de loop van 2015. Aan de proeftoetsing is een aantal belangrijke aspecten te onderkennen die we hieronder zullen bespreken.

Het doel van de proeftoetsing

De proeftoetsing diende verschillende doelen:

- Het verzamelen van informatie over de psychometrische kenmerken van de opgaven in termen van schaalbaarheid en moeilijkheidsgraad. Op basis van de verzamelde gegevens worden opgaven al dan niet geschikt bevonden om op te nemen in de opgavenbanken, respectievelijk de uiteindelijke onderdelen en domeinen. Zoals gezegd moet een veelvoud van de benodigde opgaven worden geconstrueerd en empirisch onderzocht omdat het inherent is aan het uittesten van opgaven dat er opgaven uitvallen. De belangrijkste functie van de proeftoetsing is het scheiden van het kaf en het koren.
- De Entreetoets groep 7 maakt deel uit van het Cito Leerlingvolgsysteem primair en speciaal onderwijs (LVS). Als zodanig verhoudt de toets zich tot andere toetsen van dit volgsysteem. Sommige deelscores kunnen met hun bijbehorende normering worden gebruikt als vervangende score voor meetmoment E7 in het LVS. Dit betekent dat er een ankering plaats moet vinden ten opzichte van sommige vaardigheidsschalen. Dit gebeurt door het opnemen van ankeropgaven. Daarnaast is er ook de Entreetoets voor groep 6. Bij (zeer) zwakke leerlingen kan deze toets of specifieke onderdelen daarvan afnemen als een alternatief voor (onderdelen van) de Entreetoets groep 7. Dit impliceert dat er ook in de richting van deze toets geankerd moet worden.
- De proeftoetsing speelt een cruciale rol in de equivalering met betrekking tot zowel de oude als nieuwe versie van de toets. Equivalering kan worden gerealiseerd door groepen leerlingen in de proeftoetsing opgaven te laten maken die in de bestaande versie van de Entreetoets groep 7 voorkomen én opgaven die bestemd zijn voor de nieuwe versie. Om uitspraken te kunnen doen over het predictieve vermogen van de nieuwe toetsversie, moet men immers kunnen bouwen op het toelatings- en doorstroomonderzoek (TDO) dat met een equivalente voorganger is uitgevoerd. Om equivalentie tot stand te brengen en empirisch te kunnen onderbouwen wordt de proeftoetsing ingezet: er vindt in de proeftoetsing ankering plaats naar de vorige versie van de Entreetoets groep 7.
- Ten slotte dient, ten behoeve van de bepaling van het referentieniveau, de cesuren van de referentiesets voor rekenen, lezen en taalverzorging die overeenkomen met de referentieniveaus 1F, 2F en (voor Rekenen) 1S te worden overgebracht op de betreffende onderdelen van de Entreetoets groep 7. Daartoe zijn ankeropgaven uit de genoemde referentiesets 'meegenomen' in de proeftoetsing.

De opzet van de proeftoetsing

Om alle genoemde doelstellingen te kunnen realiseren, is het noodzakelijk om de proeftoetsing uit te voeren in een zorgvuldig gekozen opzet. Vanwege de grote aantallen opgaven is het ondoenlijk om elke leerling alle te proeftoetsen opgaven te laten maken. De opgaven worden daarom in groepjes ('blokjes') verdeeld over verschillende boekjes (*booklets*). De opgavenblokjes moeten echter alle aan elkaar gekoppeld kunnen worden in wat genoemd wordt een verbonden design.

Het is ondoenlijk om voor alle onderdelen het betreffende design te beschrijven. We nemen daarom het design voor het domein Begrijpend lezen als voorbeeld (zie figuur 2.3). Voor sommige onderdelen

(bijvoorbeeld Rekenen) is het design nog complexer; in het algemeen geldt: hoe groter het aantal opgaven, hoe complexer het design.

We zullen dit voorbeeldesign hieronder puntsgewijs toelichten.

- Het design is ontworpen om (in totaal) 162 nieuwe opgaven (in de figuur blauw gemarkeerd) te onderzoeken op hun kwaliteit. Normaliter worden opgaven ondergebracht in blokjes van (meestal) vijf opgaven. De blokjes worden zodanig ondergebracht in *booklets* dat via het design alle opgaven onderling verbonden kunnen worden. In het geval van Begrijpend lezen doet zich de complicatie voor dat de opgaven zijn verbonden met teksten, 57 teksten in totaal. Omdat niet bij elk tekst evenveel opgaven zijn geconstrueerd, varieert bij dit onderdeel de omvang van een blokje tussen 1 en 10 opgaven. In de eerste kolom is een indicatie te vinden van de titel van elke uitgeteste tekst.
- Elke kolom in het design representeert één toetsboekje (*booklet*) met ongeveer 50 opgaven (praktisch gezien tussen 49 en 52) voor het domein Begrijpend lezen. Deze opgaven zijn verdeeld over twee taken die in verschillende dagdelen worden afgenomen. Om het design geheel verbonden te maken, zijn twaalf verschillende opgavenboekjes nodig. Gemiddeld bestaat net iets meer dan de helft van een boekje uit nieuw geconstrueerde opgaven, de overige opgaven zijn ankeropgaven.
- Het design is zo ingericht dat elk opgavenblokje in twee verschillende *booklets* is opgenomen. Hierdoor kunnen alle opgavenblokjes en opgaven in de blokjes bij het kalibreren onderling worden 'verbonden'. Opgaven voor de module Basis en voor de Module Verdieping opgaven zijn in één design ondergebracht; dat kan zonder problemen omdat beide typen opgaven niet wezenlijk van elkaar verschillen.
- Het aantal te proeftoetsen opgaven, het aantal blokjes en het aantal benodigde boekjes wordt bepaald door het aantal te realiseren opgaven in de taken van de module Basis en (Basis en) Verdieping. Voor het domein Begrijpend lezen zijn dat uiteindelijk respectievelijk 20 en 20, dus samen 40 opgaven. In het design zijn ankeritems ondergebracht. Daarvan zijn er 35 nodig om (in verband met de equivalering) te kunnen ankeren naar de oude Entreetoets groep 7 (in de figuur donkergrijs gearceerd met witte tekst) en 21 om te kunnen ankeren naar de Entreetoets groep 6 (in de figuur zwart gearceerd).

Figuur 2.3 Design voor de proeftoetsing van opgaven voor Lezen

Boekje	1	2	3	4	5	6	7	8	9	10	11	12
Lengte	49	50	51	50	50	52	51	50	49	49	50	51
waarvan nieuw	26	27	28	30	24	31	27	26	26	26	27	28
Tekst												
Avondvierdaagse 2013	5								5			
Rechercheur Waltink						6				6		
De volle villa								4		4		
Zeldzame sneeuwuil				4								4
1+1=2		5						5				
De zwaan				7	7							
Hoe lang zonder lucht	5			5								
Hoe maak je een rap			7								7	
Expl.. maiskorrels			4					4				
Uit boom gered		5					5					
Honing	3									3		
Watermuseum		7							7			
Meester van de maand										3	3	
Brugklasproof							5		5			
Zes vragen aan									2		2	
Vakantie (Oorpijn)		1								1		
Puzzelen			1		1							
Boswachter Sonja...						6	6					
Museumpodium								3				3
Pompoensoep						4						4
Apenheul									1			1
Spinder	5									5		
Taco kost fortuin	3							3				
Kom jij naar de Klokhuis									3		3	
Albert					5	5						
Ik word later...		4										4
Vragen (Mop)								2				2
Het zangfietspad			6				6					
Het huis in Frankrijk					5			5				
Friedrich Hundertwasser									3	3		
Sterren handtekeningen		3	3									
Middagje buurtavontuur					4							4
Discussieren kun je leren						3						3
Kip met drie poten		2			2							
Keepen of zelf hockeeyen				7							7	
Onze aarde				7		7						
Zakgeld	5						5					
Dagboek superpuber			7								7	
Zelf ijs maken	5	5			5	5	5	5	5	5	5	5
Is het jeugdjournaal		7					7					
Liegen	4		4	4		4		4		4		
Krokodillen praten in ei			6						6			
Op vakantie					7							7
Buigen bart				6							6	
Wijsneus	3	3										
Talent gezocht			5	5								
Smoezen		8	8									
Hoe overleef ik ...	5			5								
Allemaal vraagtekens					8	8						
Litteken							5	5				
Trouwe viervoeter									6	6		
PO_Kinderzwerfboek											5	5
Wie bedacht postzegel?							7			7		
Natuur voordeur	6										6	
Tomaat verpakt					6				6			
De rusteloze wind								10				10
Camera 6 jaar dobberen						4				4		
Toelichting type opgave	1 = ET7 nieuw 1 = ET6 anker 1 = ET7 anker 1 = referentieset 1F 1 = referentieset 2F											

- In de proeftoets is veel aandacht besteed aan het overbrengen van de referentieceresuren die horen bij de referentieniveaus 1F en 2F. Om dat verantwoord te kunnen doen zijn 29 1F-opgaven voor begrijpend lezen uit de betreffende openbare referentieset opgenomen in de proeftoetsing en 25 2F-opgaven.

Voor alle onderdelen in de proeftoetsing geldt in essentie een vergelijkbaar design. Dat design is wel complexer en het aantal benodigde *booklets* groter naarmate er meer opgaven geproeftoetst moeten worden.

Bij de feitelijke afname werden combinaties van onderdelen gemaakt die in verschillende proeftoetsboekjes werden ondergebracht. Zo nodig werd de inhoud van een *booklet* over meerdere taken verdeeld. De *booklets* voor Begrijpend lezen (zie de kolommen in figuur 2.3), bijvoorbeeld, werden over twee taken verdeeld zodat het aantal opgaven binnen een taak ongeveer overeenkomt met het aantal opgaven zoals dat in de uiteindelijk toets wordt opgenomen. Uiteindelijk werden twaalf verschillende ‘reguliere’ proeftoetsboekjes samengesteld, waarbij elke afname ongeveer overeenkomt met een volledige Entreetoetsafname. Dit aantal was nodig om alle mogelijke varianten van de *booklets* en taken te kunnen onderbrengen. Elk ‘regulier’ proeftoetsboekje bestond uit 16 verschillende taken waaronder soms meerdere taken per domein. Elk afnameboekje kent vier taken Rekenen, drie taken Begrijpend lezen, 2 taken Schrijven en 2 taken Woordenschat. Elk proeftoetsboekje is daarnaast op dezelfde manier samengesteld voor wat betreft de verdeling van de taken over het boekje (uiteraard met verschillende inhoud in termen van opgaven). Op deze manier werden in de proeftoetsfase (onderling vergelijkbare) afnames mogelijk gemaakt die in de praktijk sterk overeenkwamen met een ‘echte’ afname van de Entreetoets Groep 7.

Tabel 2.1 Overzicht van aantallen ankeropgaven naar oude Entreetoets groep 7 en referentiesets

Onderdeel	Anker ET7	Referentieset
Rekenen	32	113
Getallen	8	40
Verhoudingen	8	23
Metten	8	29
Verbanden	8	21
Lezen	61	69
Begrijpend lezen	35	54
Opzoeken	8	12
Samenvatten	18	3
Taalverzorging	80	75
Grammatica	20	20
Interpunctie	20	15
Werkwoorden	20	20
Niet werkwoorden	20	20
Module Verbreding	142	
Luisteren	25	
Schrijven	43	
Woordenschat	54	
Kaartlezen	20	
Totaal aantal ankeritems	315	257

Naast de 'reguliere' proeftoetsboekjes waren er nog een 'extra' proeftoetsboekjes nodig dat bestond uit zes verschillende taken. Deze proeftoetsboekjes zijn in zes verschillende varianten afgenomen. Elk 'extra' boekje bestond uit drie taken Luisteren en taken voor Grammatica, Interpunctie en Leestempo (deze laatste taak met het oog op de ankering naar andere versies van de Entreetoets).

De onderdelen van Wereldoriëntatie, ten slotte, werden in afzonderlijk proeftoetsboekjes ondergebracht. Hier zijn vier verschillende varianten van afgenomen.

Verderop in deze verantwoording komen we nog terug op twee belangrijke functies van de proeftoetsing, namelijk de ankering naar de oude Entreetoets groep 7 (in verband met de equivalering) en de ankering naar de referentiesets van het referentiekader Taal en Rekenen (in verband met het overbrengen van de referentiecesuur). Daarom hebben we hier een overzicht opgenomen van de aantallen ankeritems in de proeftoetsing (zie tabel 2.1). Om de equivalering met de oude Entreetoets groep 7 mogelijk te maken werden in totaal 315 opgaven uit deze toets meegenomen in de proeftoetsing. De opgaven voor 'Kaartlezen' hebben betrekking op een voormalig onderdeel van 'Studievaardigheden'. Studievaardigheden is in de nieuwe opzet komen te vervallen; de betreffende opgaven zijn ondergebracht bij Rekenen en Taal, met uitzondering van Kaartlezen dat in Wereldoriëntatie is opgenomen. Daarnaast zijn in totaal 257 opgaven opgenomen om de cesuren voor 1F en 1S Rekenen (113 opgaven), 1F en 2F Lezen (69 opgaven) en 1F en 2F Taalverzorging (75 opgaven) te kunnen bepalen.

De dataverzameling in het kader van de proeftoetsing

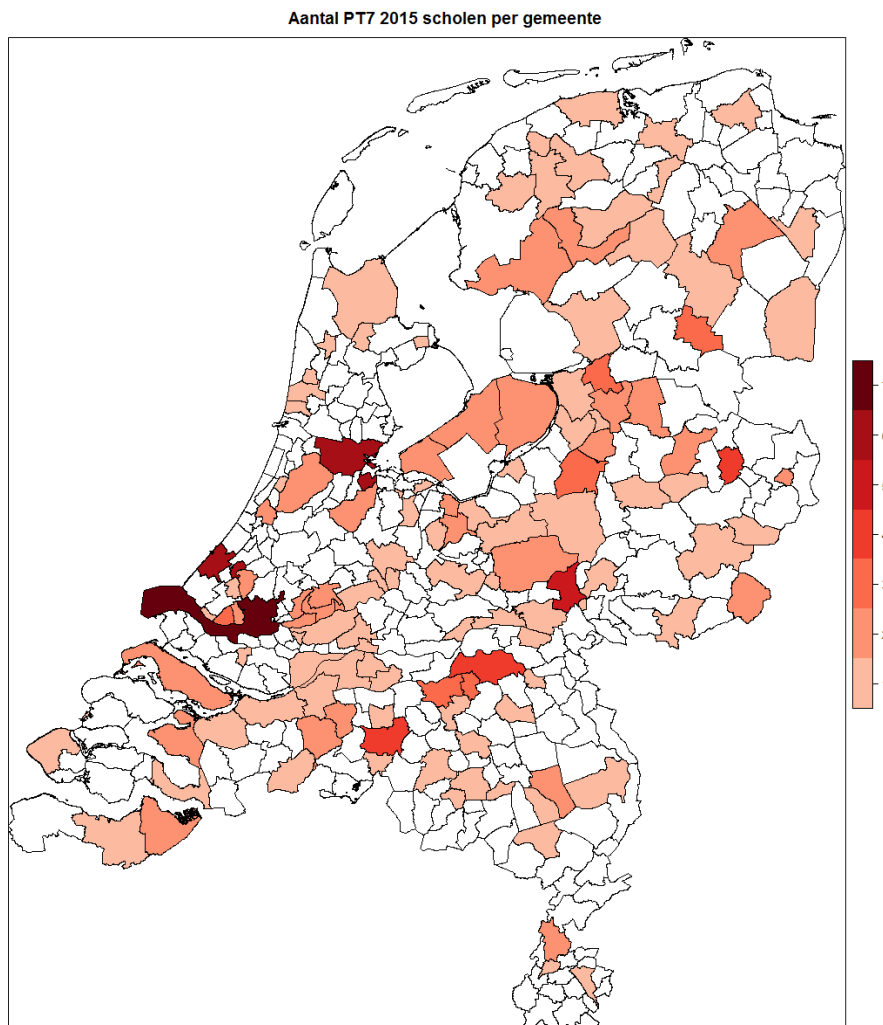
Het is van belang dat elke opgave door voldoende leerlingen is gemaakt. Er moesten voor de proeftoetsing dus voldoende scholen worden geworven. Werving van scholen geschiedt vanuit het bij Cito gebruikelijke steekproefkader (zie verder ook hoofdstuk 4 over normering en representativiteit). Dit betekent dat scholen worden benaderd overeenkomstig hun voorkomen in de populatie met betrekking tot de achtergrondkenmerken regio, mate van verstedelijking en schooltype (een combinatie van schoolgrootte en percentage gewichtenleerlingen). In dit steekproefkader zijn vooral schoolgrootte en percentage gewichtenleerlingen cruciaal. Deze kenmerken worden gecombineerd tot een stratum-indeling naar acht niveaus. De populatieverdeling voor deze strata, in combinatie met gegevens over deelnamebereidheid in eerdere jaren (variërend tussen 5% en 12%), bepaalt hoeveel scholen er per stratum worden aangeschreven. Omdat hele schoolklassen deelnemen (uit leerjaar 7) mag aangenomen worden dat het aantal jongens en meisjes bij benadering overeenkomt met de verdeling naar sekse in de populatie.

Wel maken we hier melding van de gebrekkige respons. Om het benodigde aantal scholen te realiseren, was het noodzakelijk een veelvoud daarvan aan te schrijven. De non-response was hoog. Slechts 9.7% van de benaderde scholen was bereid deel te nemen aan de proeftoetsing. Voor een deel is dit te wijten aan het feit dat scholen veelvuldig uitgenodigd worden om deel te nemen aan constructie- en normeringsonderzoeken ten behoeve van LVS-toetsen. Er waren 60 scholen die niet benaderd waren vanuit het steekproefkader, maar die zich hebben opgegeven voor deelname via de website. Uiteindelijk deden in 2015 206 scholen met 5035 leerlingen mee aan de proeftoetsing.

Voor constructieonderzoek is representativiteit van ondergeschikt belang. Een zekere mate van spreiding over de belangrijkste achtergrondkenmerken volstaat. Vanwege de gebrekkige response hebben we niettemin een aantal controleanalyses uitgevoerd.

Op de eerste plaats is bekeken hoe het gesteld was met de spreiding naar regio. Figuur 2.4 geeft hiervan een beeld. In de figuur wordt aangegeven of een gemeente al dan niet is vertegenwoordigd in de steekproef voor het proeftoetsonderzoek. Daarnaast is door middel van kleur-saturatie aangegeven om hoeveel scholen per gemeente het gaat (1 – 7 scholen per gemeente). De deelnemende gemeenten en (189) scholen zijn verspreid over het hele land. Er is geen regio of provincie aan te wijzen die geheel ongedekt is door de steekproef. Daarbij moet men verdisconteren dat het aantal gemeenten in Nederland ten tijde van de proeftoetsing ruim twee maal zo groot was als het aantal deelnemende scholen. Er is ook geen regio met een opvallend hoge concentratie van deelnemende scholen.

Figuur 2.4 Spreiding over Nederland van de scholen (gemeenten) in de proeftoetsing van 2015



Het beeld dat de spreiding naar regio laat zien, wekt vertrouwen, maar is relatief onbelangrijk omdat regio als achtergrondkenmerk over het algemeen weinig effect heeft op de scores. Bij de combinatie van achtergrondkenmerken schoolgrootte en percentage leerlingen met een afwijkend leerlinggewicht is dit wel het geval. Scholen met hoge percentages gewichtenleerlingen bijvoorbeeld scoren gemiddeld duidelijk lager dan andere scholen. Daarom zijn we nagegaan of de verdeling van de scholen over deze variabelen in de steekproef globaal overeenkomt met de populatieverdeling. Bij de genoemde achtergrondkenmerken past de volgende toelichting:

Wat betreft *schoolgrootte* kan onderscheid worden gemaakt tussen grote scholen (> 200 leerlingen) en kleine scholen (\leq 200 leerlingen).

Het *percentage leerlingen met een afwijkend leerlinggewicht* is gebaseerd op de gegevens die DUO hierover per school registreert. De formatieomvang van een school wordt in belangrijke mate bepaald door het formatiegewicht van de leerlingen die samen de schoolpopulatie vormen. Voorheen werden de leerlingen gecategoriseerd in vijf formatiegewichten die een combinatie vormden van opleidingsniveau, sociaal-economische status en etnische herkomst van de ouders. Inmiddels is echter de categorisering naar leerlinggewichten vervangen door een andere regeling waarbij nog slechts drie niveaus worden onderscheiden en waarbij etniciteit geen rol meer speelt. De definities van de categorieën zijn aangescherpt en uitsluitend gebaseerd op het opleidingsniveau van de ouders.

Gewicht	Omschrijving
0.0	één van de ouders of beide ouders heeft of hebben een opleiding gehad uit categorie 3 (zie voor de categorieomschrijving hieronder)
0.3	beide ouders of de ouder die belast is met de dagelijkse verzorging heeft of hebben een opleiding uit categorie 2 gehad
1.2	één van de ouders heeft een opleiding gehad uit categorie 1 en de ander een opleiding uit categorie 1 óf 2

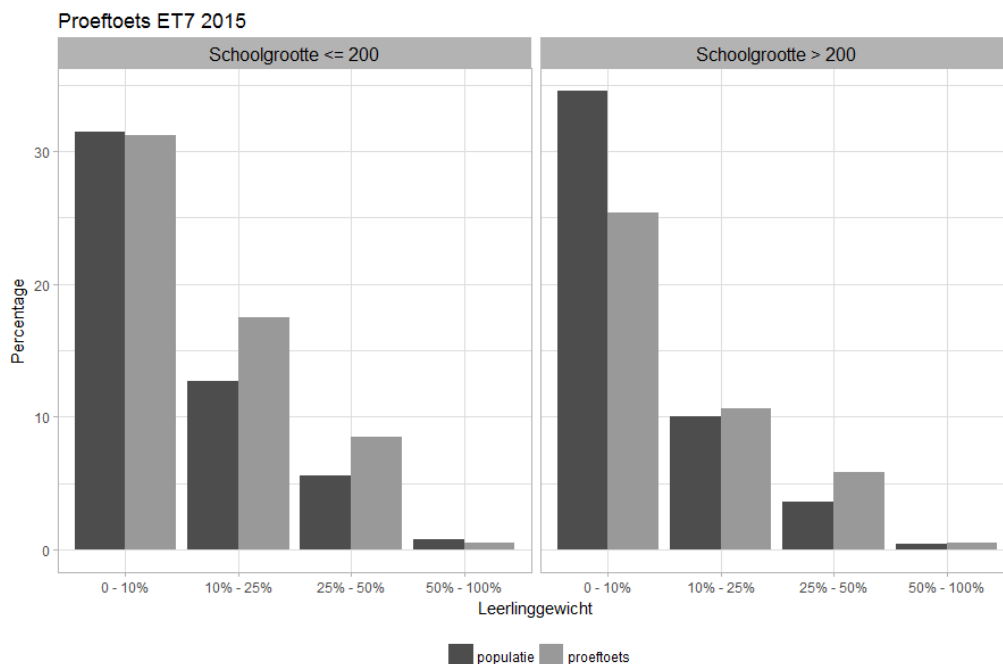
Categorie	Omschrijving
1	maximaal basisonderwijs of (v)so-zmlk
2	maximaal lbo/vbo, praktijkonderwijs of vmbo basis- of kaderberoepsgerichte leerweg
3	overig vo en hoger

Scholen zijn in te delen in een van de volgende groepen die in combinatie met schoolgrootte resulteren in acht categorieën:

Schoolgrootte		≤ 200 leerlingen	> 200 leerlingen
Categorie	(% gewichtsl leerlingen)		
0 - 10%	leerlingen met gewicht 0.3 of 1.2	(1)	(5)
10 - 25%	leerlingen met gewicht 0.3 of 1.2	(2)	(6)
25 - 50%	leerlingen met gewicht 0.3 of 1.2	(3)	(7)
50 - 100%	leerlingen met gewicht 0.3 of 1.2	(4)	(8)

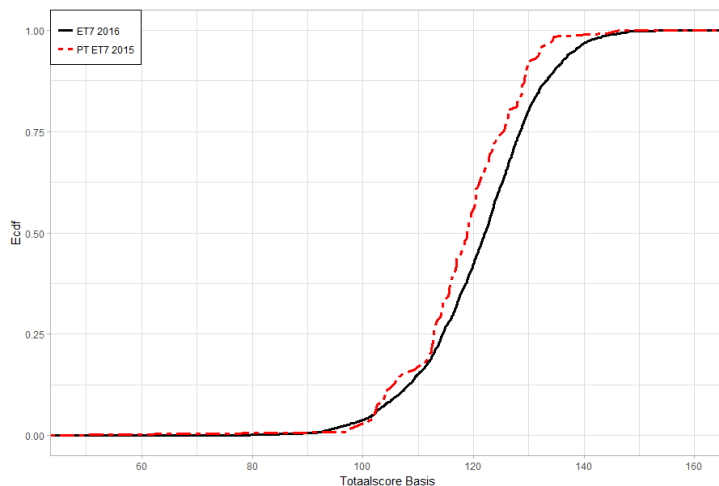
In figuur 2.5 is grafisch weergegeven of de verdeling van de 189 scholen naar schoolgrootte en percentage gewichtsl leerlingen in de steekproef van scholen (globaal) overeenkomt met de populatieverdeling. Dat is het geval. Er zijn weliswaar kleine afwijkingen, maar die vormen voor proefvoetsonderzoek geen probleem.

Figuur 2.5 Samenstelling proefvoets-steekproef naar schoolgrootte en % gewichtsl leerlingen



Ten slotte is, bij wijze van eindcontrole, een analyse uitgevoerd waarbij de empirische cumulatieve distributiefunctie (de *ecdf*) van de gemiddelde schoolscores op de Entreetoets groep 7 werd bepaald en vergeleken voor de proefoetsscholen en de gehele populatie van scholen die in 2016 deelnamen aan de Entreetoets groep 7. Aan de figuur is af te lezen dat de *ecdf* voor de steekproef (rood geplot) en die voor de verzameling van alle (ruim 2000) scholen in de afnamepopulatie van 2016 (zwart geplot) dicht in elkaars buurt liggen. Wel is het prestatieniveau in de steekproef voor de proefoetsing over het algemeen wat lager dan het prestatieniveau bij de echte afname. Dit kan wijzen op systematische afwijkingen in de steekproefverdeling van sommige achtergrondkenmerken ten opzichte van de populatieverdeling. Dat is moeilijk te zeggen omdat slechts een deel (117 van de 189) van de scholen in de steekproef een jaar later meedeed aan de 'echte' afname. Waarschijnlijker is, dat de leerlingen in het proefoetsonderzoek zich wat minder hard hebben ingespannen dan de leerlingen in de 'echte' afname. Uit eerder onderzoek (zie Keuning, van Boxtel, Lansink, Visser, Weekers & Engelen, 2015) blijkt dat leerlingen minder goed presteren naarmate er voor hen en voor hun leerkrachten minder op het spel staat. Voor de proefoetsing zijn dit soort kleine afwijkingen niet belangrijk. We kunnen daarom concluderen dat de steekproef voor de proefoetsing van 2015 qua samenstelling aan zijn doel beantwoordt.

Figuur 2.6 Empirische cumulatieve distributiefunctie voor scholen in het proefoetsonderzoek en alle scholen in de afnamepopulatie van 2016



Verder zijn, zoals gezegd, de aantallen leerlingen in het proefonderzoek zo gekozen dat er voldoende informatie per uit te testen opgave beschikbaar kon komen. In tabel 2.2 geven we aan hoe vaak een opgave werd gemaakt. We hebben daarbij niet de aantallen opgaven in de tabel opgenomen die uiteindelijk in de Entreetoets groep 7 zijn opgenomen, maar *alle* geproefoetste opgaven.

Tabel 2.2 Minimum en gemiddeld aantal leerlingen in de proeftoetsing van 2015 dat een opgave heeft gemaakt (per onderdeel)

Onderdeel/domein	Proeftoets 2015		
	Aantal opgaven	Aantal leerlingen	
		Minimum	Gemiddeld
Rekenen	413	377	932.7
Lezen	442	374	874.2
Taalverzorging	320	364	1030.9
• Grammatica	55	488	655.6
• Interpunctie	55	364	619.7
• Spelling niet werkwoorden	110	626	1175.2
• Spelling werkwoorden	100	661	1304.7
Luisteren	140	462	622.6
Schrijven	159	378	1327.6
Woordenschat	159	374	1370.6
Wereldoriëntatie	117	685	765.7

In de tabel is voor elk onderdeel/domein afzonderlijk weergegeven, welk minimum aantal leerlingen een willekeurige opgave uit dat onderdeel heeft gemaakt en het gemiddelde aantal leerlingen dat de opgaven uit een onderdeel heeft gemaakt. Voor Taalverzorging hebben we de tabel opgesplitst naar de domeinen, omdat voor de domeinen de aantallen leerlingen nogal verschilden.

In de tabel is goed te zien dat de aantallen waarnemingen per opgave minimaal bijna op vierhonderd liggen. De minimum aantallen variëren tussen 364 (Taalverzorging; Interpunctie) en 685 leerlingen (voor Wereldoriëntatie). Gemiddeld zijn de opgave door veel meer leerlingen gemaakt. Voor een aantal onderdelen ligt het gemiddelde zelfs ruim boven de duizend. Door deze grote aantallen waarnemingen kon de kwaliteit van alle opgaven met grote zekerheid worden vastgesteld, wat de zekerheid waarmee we kwaliteitsuitspraken menen te kunnen doen over de onderdelen, domeinen en opgaven van de Entreetoets groep 7 gunstig heeft beïnvloed.

3 Beschrijving van de toets

3.1 Opbouw en structuur

Inhoud en samenstelling

De Entreetoets bestaat uit een verplichte en twee optionele modules. In deze modules worden in verschillende taken, domeinen getoetst op het gebied van taal, rekenen en wereldoriëntatie. Een overzicht van alle taken en domeinen, de wijze waarop deze zijn ondergebracht in de basistoets, de verdiepende en de verbredende module, informatie over de geschatte afnametijd per taak en de manier waarop men taken kan verdelen over dagdelen is beschreven in de handleiding voor de leerkracht.

In deze wetenschappelijke verantwoording ligt het accent op de onderdelen en domeinen waarvan de opgaven leiden tot afzonderlijke scores op het individuele Leerlingprofiel. Wanneer men uitsluitend de basistoets afneemt, bestaat het Leerlingprofiel uit een score op drie onderdelen, namelijk Rekenen (90 opgaven), Lezen (40 opgaven) en Taalverzorging (50 opgaven) en een totaalscore (gebaseerd op in totaal 180 opgaven). Door de ruwe scores om te zetten in normscores (in de vorm van percentielscores) zijn deze te interpreteren: wat is hun betekenis in vergelijking met de leerprestaties van jaargroepgenoten. Percentielscores zijn ook te beschrijven in termen van niveauscores I tot en met V, een in het LVS gebruikelijk indeling in vijf groepen van steeds 20% van de normpopulatie:

Vaardigheidsniveau	Interpretatie	Percentiel
II	20% boven het landelijk gemiddelde	60-80
III	20% landelijk gemiddelde	40-60
IV	20% onder het landelijk gemiddelde	20-40
V	20% laagst scorende leerlingen	0-20

De scores voor de onderdelen Rekenen, Lezen en Taalverzorging zijn tevens te interpreteren in termen van referentieniveaus. Voor alle onderdelen wordt gerapporteerd of de leerling al dan niet voldoet aan het landelijk vastgestelde referentieniveau 1F. Daarnaast wordt voor Lezen en Taalverzorging ook gerapporteerd of referentieniveau 2F is behaald; voor Rekenen betreft dat referentieniveau 1S.

Men kan kiezen voor een meer gedifferentieerde rapportage waarbij ook percentielscores worden gerapporteerd voor de elf onderscheiden domeinen. Om dit voldoende betrouwbaar te kunnen realiseren, zijn meer opgaven nodig dan in de basistoets zijn opgenomen. Men kan hierin voorzien door de verdiepende module af te nemen. Op deze manier is het mogelijk in detail te rapporteren in de vorm van percentielscores en niveauscores over de vier domeinen van Rekenen: Getallen (50 opgaven), Meten en meetkunde (35 opgaven), Verbanden (25 opgaven) en Verhoudingen (30 opgaven). Voor Lezen zijn dit de domeinen Begrijpend lezen (40 opgaven), Opzoeken (20 opgaven) en Samenvatten (20 opgaven). Voor Taalverzorging zijn domeinscores mogelijk voor Grammatica, (20 opgaven), Interpunctie (20 opgaven), Spelling niet-werkwoorden (25 opgaven) en Spelling werkwoorden (eveneens 25 opgaven).

In de verbredende module zijn aanvullende 'talige' toetsen ondergebracht voor Luisteren (30 opgaven), Schrijven (30 opgaven) en Woordenschat (40 opgaven) en daarnaast drie deoltoetsen voor Wereldoriëntatie (met elk 20 opgaven voor de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek). Voor de verbredende module geldt, net als voor de basistoets en de verdiepende module een normering in termen van percentielscores en vaardigheidsniveaus.

Ten slotte geeft het rapport Vooruitblik een indicatie van het brugklatype dat het beste past bij het niveau van leervorderingen dat de leerling laat zien. De basis voor dit rapport is een somscore over alle opgaven in

de basistoets; de gehele basistoets moet hiervoor afgenomen zijn. De opgaven van de verbredende en verdiepende modules tellen niet mee. De vraag naar de criteriumvaliditeit van de Entreetoets groep 7 heeft dus met name betrekking op deze somscore.

In totaal telt de Entreetoets groep 7 470 opgaven.

Onderdeel	Domein	Basis	Module	
			Verdieping	Verbreding
Rekenen	Getallen	30	20	
	Meten en meetkunde	25	10	
	Verbanden	15	10	
	Verhoudingen	20	10	
Lezen	Begrijpend lezen	20	20	
	Opzoeken	10	10	
	Samenvatten	10	10	
Taalverzorging	Grammatica	10	10	
	Interpunctie	10	10	
	Spelling niet-werkwoorden	15	10	
	Spelling werkwoorden	15	10	
Luisteren				30
Schrijven				30
Woordenschat				40
Wereldoriëntatie	Aardrijkskunde			20
	Geschiedenis			20
	Natuur en techniek			20

Praktische informatie over de wijze waarop de taken zijn ondergebracht in opgavenboekjes, over kennismakingstaken en over afnametijd en inrichting van de toetsafname is te vinden in de handleiding.

3.2 Inhoudsverantwoording

In deze paragraaf wordt de inhoud van de Entreetoets besproken en verantwoord. De tekst van de paragraaf is grotendeels gelijk aan de inhoudsverantwoording die als hoofdstuk 4 is opgenomen in de handleiding in de Leerkrachtmap Entreetoets groep 7 (Cito, 2016). Er zijn echter verschillen.

De inhoudsverantwoording in de handleiding is primair beschrijvend van aard: welke (typen) opgaven zijn opgenomen in de onderdelen van de toets, wat is hun karakter en hoe groot is hun aantal. Ter illustratie zijn voorbeelden opgenomen van opgaven. In de onderhavige paragraaf ligt het accent sterker op de verantwoording van de inhoudsvaliditeit: *Waarom* is gekozen voor bepaalde typen opgaven? En waarom was het wenselijk om bij de toetsconstructie een bepaalde verdeling van de opgaven na te streven? Daarbij gaan we steeds na of we erin geslaagd zijn om de gewenste verdeling van opgaven in de toets per onderdeel ook daadwerkelijk te realiseren.

Deze inhoudsverantwoording kent, hoe uitgebreid ook, zijn beperkingen. Op de eerste plaats ontbreken voorbeelden van toetsopgaven. We verwijzen de lezer daarvoor naar bovengenoemde handleiding en natuurlijk naar de opgavenboekjes zelf. Verder hebben we ervoor gekozen om de verantwoording vooral toe te spitsen op de operationalisatie van de leerdoelen en vaardigheden die voor de betreffende doelgroep (leerlingen van groep 7) relevant zijn. De theoretische beschrijving van de wijze waarop de verschillende onderscheiden vaardigheden zich in het onderwijsleerproces ontwikkelen, hebben we hierbij beknopt gehouden. De geïnteresseerde lezer verwijzen we graag naar de uitgebreidere inhoudsverantwoording per onderdeel en domein in de wetenschappelijke verantwoordingen bij de afzonderlijke LVS 3.0-toetsen van het Cito Volgsysteem primair en speciaal onderwijs. Daarin zijn ook uitgebreidere literatuurreferenties opgenomen.

De paragraaf is als volgt ingedeeld.

In paragraaf 3.2.1 bespreken we allereerst de onderdelen van de basistoets en de module Verdieping. Dit zijn Rekenen, Lezen en Taalverzorging. We verantwoorden de basistoets en de module Verdieping in onderlinge samenhang, want de verdiepende onderdelen hebben betrekking op domeinen die binnen de basistoets worden onderscheiden. Bij het onderdeel Rekenen betreft het de domeinen Getallen, Meten en Meetkunde, Verbanden en Verhoudingen, bij Lezen de domeinen Begrijpend lezen, Opzoeken en Samenvatten, en bij Taalverzorging de domeinen Grammatica, Interpunctie, Spelling niet-werkwoorden en Spelling werkwoorden.

In paragraaf 3.2.2 verantwoorden we de onderdelen van de module Verbreding. Luisteren, Schrijven en Woordenschat zijn net als Lezen en Taalverzorging talige onderdelen. We sluiten af met het onderdeel Wereldoriëntatie waarvan Aardrijkskunde, Geschiedenis en Natuur en techniek deel uitmaken.

De bespreking van de verschillende onderdelen van de Entreetoets groep 7 kent steeds dezelfde opbouw. Eerst omschrijven we de vaardigheid die in het onderdeel wordt getoetst, gevolgd door een bespreking van de belangrijkste deelvaardigheden, facetten en aspecten ervan. Bij de onderdelen van de basistoets (en module Verdieping) doen we dit door de domeinen en referentieniveaus aan de orde te stellen. Daarna gaan we in op de ontwikkeling van de vaardigheid en de belangrijkste leerdoelen voor groep 7. Zoals gezegd doen we dit beknopt. Via een opsomming van de belangrijkste bronnen die we bij het construeren van een onderdeel hebben gehanteerd komen we terecht bij de inhoud zelf. We benoemen de belangrijkste soorten opgaven en laten zien bij welke categorie de opgaven in de verschillende taken horen. Een belangrijk element in de bewijsvoering omtrent inhoudsvaliditeit is de beargumentering van de gewenste verdeling van opgaven in de toets. Deze mondt uit in een gewenste toetsmatrijs die we steeds vergelijken met de gerealiseerde toetsmatrijs. Hierbij past de volgende toelichting.

Bij de constructie van de toets worden opgaven ontwikkeld op basis van de gewenste toetsmatrijs. Dat zijn veel meer opgaven dan er uiteindelijk in de toets kunnen worden opgenomen, in de regel twee tot vier maal de benodigde aantallen. Dat komt omdat er sprake is van mortaliteit. Ondanks alle zorg die besteed wordt aan de opgavenconstructie sneuvelen er nogal wat opgaven in de proeftoetsonderzoeken.

De psychometrische itemkenmerken blijken dan niet in overeenstemming te zijn met de verwachtingen. De opgaven zijn soms te moeilijk of te gemakkelijk, of hun discriminerende eigenschappen vallen tegen. In de keuze van de opgaven voor de uiteindelijke toets worden inhoudelijke en psychometrische eigenschappen afgewogen. Soms leidt dit ertoe dat de gewenste verdeling van opgaven niet helemaal wordt gerealiseerd, bijvoorbeeld omdat teveel opgaven van een bepaald type afvallen. Over het algemeen kunnen we stellen dat de inhoudsvaliditeit van een toets of toetsonderdeel beter is naarmate de gerealiseerde verdeling van opgaven in de toetsmatrijs meer in overeenstemming is met de gewenste verdeling. Kleine afwijkingen vormen geen probleem. Want in de regel komt de keuze voor een iets minder optimale inhoudsvaliditeit ten goede aan een optimale constructvaliditeit. In het kalibratieproces wordt immers zoveel mogelijk de voorkeur gegeven aan opgaven die uitstekend passen bij de onderliggende vaardigheidsschaal (zie hierover ook hoofdstuk 4 waarin we aandacht besteden aan de kalibratie). Opgaven die onvoldoende passen bij de te realiseren vaardigheidsschaal worden niet in de toets opgenomen. Kortom, in het constructieproces worden inhouds- en constructvaliditeit geoptimaliseerd, waarbij beide validiteitsaspecten soms tegen elkaar moeten worden afgewogen.

De onderdelen van de basistoets en optioneel de module Verdieping geven de mogelijkheid om zicht te krijgen op de prestaties van de leerlingen in relatie tot het referentiekader Nederlandse taal en rekenen (niveau 1F en 2F, respectievelijk 1F en 1S). Bij de bespreking van deze onderdelen verwijzen we naar dit referentiekader en geven we aan hoe de toetsinhoud dit referentiekader weerspiegelt. Meer informatie over de inhoudelijke verantwoording in termen van referentieniveaus, over het bepalen van de toetsscore waarbij de niveaus 1F (bij Rekenen, Lezen en Taalverzorging), 1S (bij Rekenen) en 2F (bij Lezen en Taalverzorging) worden behaald en over de betrouwbaarheid van deze scores is te vinden in de hoofdstukken 4 (over normering) en 5 (over betrouwbaarheid en nauwkeurigheid).

3.2.1 De basistoets en de module Verdieping

3.2.1.1 Rekenen

Begripsomschrijving

Het leerstofgebied Rekenen omvat verschillende vaardigheden die zich richten op getalbegrip en rekenvaardigheid. De belangrijkste vaardigheden betreffen:

- het inzicht in getallen en verhoudingen;
- maatzicht;
- ruimtelijk inzicht;
- het kunnen hanteren van tabellen en grafieken;
- het kunnen uitvoeren van operaties met getallen;
- en het kunnen toepassen van die kennis en inzichten, zonder context maar ook in uiteenlopende situaties.

We kunnen vier domeinen onderscheiden:

- Getallen
- Meten en meetkunde
- Verbanden
- Verhoudingen

Getallen

Het domein Getallen heeft enerzijds betrekking op getallen en getalsrelaties en anderzijds op het uitvoeren van de elementaire bewerkingen. Deze betreffen het optellen, aftrekken, vermenigvuldigen en delen, en combinaties van deze bewerkingen, zowel met hele getallen als met kommagetallen en breuken.

De leerling krijgt zowel opgaven met als zonder context aangeboden. Bij het oplossen van de opgaven gaat het niet alleen om feitenkennis, kennis van begrippen en de beheersing van (standaard)procedures en routines, maar ook om het doelmatig (of efficiënt) kunnen toepassen en gebruiken van deze kennis, het leggen van relaties, het inzicht hebben in reksituaties en hierover kunnen redeneren.

Meten en meetkunde

Bij het domein Meten gaat het om het meten van allerlei verschijnselen in de werkelijkheid. Leerlingen krijgen begrip van verschillende grootheden (lengte, omtrek, oppervlakte, inhoud, gewicht, tijd, snelheid en geld), leren meten met verschillende instrumenten en leren meetresultaten af te lezen en te interpreteren. Ook het omzetten van maateenheden en de opbouw en decimale structuur van het metrieke stelsel vallen hieronder. Bij het domein Meetkunde ligt de nadruk op het beschrijven van, en het greep krijgen op ruimtelijke aspecten van de werkelijkheid.

Verbanden

Binnen het domein Verbanden gaat het om het omgaan met tabellen, diagrammen en grafieken, legenda's en assenstelsels.

Verhoudingen

Leerlingen leren de structuur en samenhang van verhoudingen op hoofdlijnen te doorzien en er in praktische situaties mee te rekenen. Verhoudingen kunnen beschreven worden in verhoudingstaal (zoals bij 'één op de tien Nederlanders' of 'het aantal fietsers is twee keer zo groot als het aantal automobilisten'), maar soms ook in breukentaal (bijvoorbeeld 'driekwart van de inwoners is ouder dan 25 jaar', dat is drie op de vier) en met procenten (zoals '70 procent van de mensen is vóór de aanleg van een randweg'). Bij het domein Verhoudingen gaat het dus soms ook om het rekenen met breuken en procenten. Daar waar het om breuken als getallen gaat en om bewerkingen met breuken vallen deze onder het domein Getallen.

Referentieniveaus

Het Referentiekader Taal en Rekenen (Ministerie van Onderwijs, Cultuur en Wetenschap, 2009) schrijft voor wat leerlingen op een aantal drempelmomenten in het onderwijs moeten kennen en kunnen. Voor leerlingen in de bovenbouw van het basisonderwijs zijn voor rekenen de niveaus 1F en 1S belangrijk

Inhoudsverantwoording

De inhoud van het onderdeel Rekenen in de Entreetoets groep 7 is gebaseerd op verschillende bronnen. De belangrijkste zijn:

- De kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2004, 2006)
- De reken-wiskundemethoden die in het basisonderwijs gebruikt worden
- De TAL-publicaties (Tal-team, 1999, 2001, 2004, 2005, 2007)
- De Toetswijzer Eindtoets Basisonderwijs (College voor Examens, 2012)
- De referentieniveaus Rekenen voor het basisonderwijs (Expertgroep doorlopende leerlijnen taal en rekenen, 2008b)

In de tabellen 3.1 en 3.2 staat weergegeven hoe de opgaven zijn verdeeld over de domeinen en referentieniveaus in de basistoets en in de module Verdieping. In tabel 3.1 is in de laatste kolom te zien dat uiteindelijk 50 opgaven zijn gerealiseerd voor het domein Getallen, 35 voor Meten en meetkunde, 25 voor Verbanden en 30 voor Verhoudingen. In deze kolom zijn (tussen haakjes) ook de gewenste aantallen opgaven opgenomen. Hierbij valt op te merken dat de gewenste aantallen hoofdzakelijk worden bepaald door de omvang van de domeinen. Zoals in de nog volgende inhoudelijke bespreking van de domeinen naar voren zal komen, omvatten sommige domeinen (zoals Getallen) zeer veel (sub)onderdelen en (sub)onderwerpen, andere domeinen (zoals Verbanden) aanzienlijk minder. Hoe meer (sub)onderdelen een domein omvat, hoe meer opgaven nodig zijn om het domein inhoudelijk afdoende te kunnen dekken. De gewenste verdeling van opgaven over de domeinen heeft in eerste instantie betrekking op de basistoets. De gewenste aantallen opgaven per domein in de verdiepende module zijn globaal hiervan afgeleid, maar bij deze opgaven speelde het ook een belangrijke rol dat over elk domein afzonderlijk voldoende betrouwbaar moest kunnen worden gerapporteerd. Uit de tabel blijkt dat uiteindelijk voor elk domein precies zoveel opgaven konden worden gerealiseerd als gewenst.

Tabel 3.1 De vier domeinen met aantallen opgaven in de taken Rekenen

	Basis	aantal opgaven	module Verdieping	aantal opgaven	Totaal
Getallen	taak 2, 4, 6	10, 10, 10	taak 9, 12	10, 10	50 (50)
Meten en meetkunde	taak 2, 4, 6	8, 8, 9	taak 9, 12	5, 5	35 (35)
Verbanden	taak 2, 4, 6	5, 5, 5	taak 9, 12	5, 5	25 (25)
Verhoudingen	taak 2, 4, 6	7, 7, 6	taak 9, 12	5, 5	30 (30)

In tabel 3.2 staan de percentages en aantallen opgaven op niveau 1F en 1S die zijn opgenomen in de taken Rekenen. Daarbij zijn de aantallen opgaven uitgesplitst per domein, voor de basistaken en de verdiepende taken afzonderlijk.

Tabel 3.2 De aantallen opgaven en percentages 1F en 1S van de referentieniveaus in de taken Rekenen

	Gerealiseerde verdeling	
	1F	1S
Taken Rekenen Basis	52 (58%)	38 (42%)
Getallen	21	9
Metten en meetkunde	15	10
Verbanden	8	7
Verhoudingen	8	12
Taken Rekenen module Verdieping	30 (60%)	20 (40%)
Getallen	12	8
Metten en meetkunde	7	3
Verbanden	5	5
Verhoudingen	6	4
Totaal aantal (% gemiddeld)	82 (59%)	58 (41%)

De gewenste verdeling van opgaven over de referentieniveau 1F en 1S is vooraf en op psychometrische gronden globaal vastgelegd op tenminste 15 en bij voorkeur 20. Er werden vooraf geen streefaantallen per domein vastgelegd; de gerealiseerde aantallen per domein worden bepaald door de wijze waarop de toetsonderdelen worden samengesteld. In eerste instantie worden omwille van de inhoudsvaliditeit opgaven geselecteerd op domein, vervolgens op itemkenmerken zoals moeilijkheidsgraad en uiteindelijk wordt aan de selectie een referentieniveau toegekend. De toekenning van het referentieniveau is afhankelijk van de inhoud van de opgave en niet van de moeilijkheidsgraad. Na toekenning van de referentieniveaus wordt gecontroleerd of het gerealiseerde aantal voldoet aan de vooraf gestelde eisen. Voor de basistoets zijn er ruimschoots voldoende opgaven op beide referentieniveaus, voor basis- en verdiepende module tezamen uiteraard ook.

In tabel 3.3 is voor de basistoets en de module Verdieping aangegeven tot welk domein de specifieke opgaven behoren.

Tabel 3.3 Opgavenummers per domein per taak in de basistoets en de module Verdieping

	Getallen	Metten en meetkunde	Verbanden	Verhoudingen
Taak 2	2, 4, 7, 11, 13, 17, 19, 23, 26, 30	5, 10, 16, 18, 21, 22, 25, 28	3, 8, 14, 20, 27	1, 6, 9, 12, 15, 24, 29
Taak 4	1, 2, 6, 10, 12, 15, 18, 21, 27, 28	4, 9, 14, 16, 20, 22, 25, 29	3, 7, 13, 19, 24	5, 8, 11, 17, 23, 26, 30
Taak 6	2, 5, 7, 10, 13, 16, 19, 22, 27, 30	4, 6, 11, 15, 17, 21, 23, 26, 28	3, 9, 14, 20, 25	1, 8, 12, 18, 24, 29
Taak 9	2, 4, 6, 8, 12, 13, 16, 18, 22, 25	5, 9, 14, 19, 23	1, 10, 15, 20, 24	3, 7, 11, 17, 21
Taak 12	1, 4, 6, 7, 11, 13, 16, 18, 21, 25	3, 9, 14, 19, 23	5, 10, 15, 20, 24	2, 8, 12, 17, 22

De opgaven voor Rekenen zijn geconstrueerd aan de hand van de onderwijsdoelen die onder elk domein kunnen worden gerangschikt. Hieronder geven we steeds per domein een opsomming van de doelen.

De inhoud van het domein Getallen

De doelen voor het domein Getallen vallen uiteen in die voor getallen (in engere zin) en getalsrelaties enerzijds en bewerkingen anderzijds.

Getallen en getalsrelaties

- Uitspraak en notatie van hele, kommagetallen en breuken
- Wiskundetaal gebruiken
- Verder- en terugtellen met hele en kommagetallen
- Vergelijken van hele, kommagetallen en breuken
- Globaal en precies plaatsen en aflezen op getallenlijn
- Afronden van getallen
- Splitsen en samenvoegen van hele en kommagetallen
- Relaties tussen breuken en kommagetallen
- Doorzien structuur tientallig stelsel

Bewerkingen

Optellen en aftrekken, zowel met als zonder context ('kaal'):

- Doelmatig
- Standaard procedures
- Globaal

Vermenigvuldigen en delen, zowel met als zonder context ('kaal'):

- Doelmatig
- Standaard procedures
- Globaal

Combinaties:

- Berekeningen, zowel met als zonder context ('kaal'), exact, globaal
- Gemiddelde, zowel met als zonder context ('kaal')
- Volgorde van bewerkingen

Bewerkingen met breuken, zowel met als zonder context ('kaal'):

- Vereenvoudigen, gelijknamig maken
- Optellen, aftrekken
- Deel van geheel getal nemen, heel getal x breuk
- Breuk x breuk, deel van breuk nemen
- Heel getal / breuk of gemiddelde getal
- Breuk of gemiddelde getal / breuk

De inhoud van het domein Meten en meetkunde

De doelen voor dit domein vallen uiteen in doelen voor 'meten' en doelen voor 'meetkunde'. Bij het 'meten' gaat het om het meten van allerlei verschijnselen in de werkelijkheid. Leerlingen krijgen begrip van verschillende grootheden (lengte, omtrek, oppervlakte, inhoud, gewicht, tijd, snelheid en geld), leren meten met verschillende instrumenten en leren meetresultaten af te lezen en te interpreteren. Ook het omzetten van maateenheden en de opbouw en decimale structuur van het metrieke stelsel vallen hieronder. Bij 'meetkunde' ligt de nadruk op het beschrijven van, en het greep krijgen op ruimtelijke aspecten van de werkelijkheid.

Meten

Lengte en omtrek

- Meetinstrumenten hanteren en aflezen
- Inschatten lengte; betekenis geven en kiezen juiste lengtemaat
- Omtrek berekenen / benaderen
- Herleiden van lengtematen
- Toepassingen

Oppervlakte

- Inschatten oppervlakte; betekenis geven en kiezen juiste oppervlaktemaat
- Bepalen oppervlakte via afpassen of berekenen
- Herleiden van oppervlaktematen
- Toepassingen

Inhoud

- Meetinstrumenten hanteren en aflezen
- Inhoud bepalen (via tellen) en berekenen van rechthoekige figuren
- Herleiden van inhoudsmaten
- Toepassingen

Gewicht

- Meetinstrumenten hanteren en aflezen
- Betekenis geven en kiezen juiste maat
- Herleiden van gewichtsmaten
- Toepassingen

Tijd en snelheid

- Aflezen van tijden; digitale en analoge tijden koppelen
- Rekenen met tijdsaanduidingen
- Kalenderaanduidingen en datumaanduidingen kennen en gebruiken
- Herleidingen en samenstellingen gebruiken als km/u en m/sec

Geld

- Samenstellen van bedragen en totale waarde bepalen
- Omwisselen van munten/biljetten in andere
- Rekenen met geldbedragen in euro's (bijbetalen, wisselgeld)
- Toepassingen o.a. met wisselkoersen; samenstellingen (prijs/uur; prijs/liter)

Meetkunde

- Namen en eigenschappen van meetkundige figuren
- Relaties tussen 2D- en 3D-objecten
- Aanzichten herkennen en interpreteren
- Interpreteren van plattegronden en bouwtekeningen
- Innemen van standpunten
- Routes en plaatsen beschrijven en lezen op een kaart
- Aanduidingen op een windroos kennen
- Symmetrie herkennen; spiegelen; spiegels vinden
- Meetkundige patronen herkennen en voortzetten

De inhoud van het domein Verbanden

Hieronder staan de doelen beschreven voor dit domein.

Tabellen (lezen, analyseren, interpreteren, rekenen, voorspellen)

- Lesrooster
- Vertrektijden bussen
- Vertrektijden treinen
- Andere tabellen zoals bioscooptijden

Lezen, interpreteren, trends ontdekken, voorspellen bij diagrammen

- Beelddiagram
- Staafdiagram horizontaal
- Bevolkingspiramide
- Staafdiagram verticaal
- Spreidingdiagram
- Cirkeldiagram
- Stroomdiagram

Lezen, interpreteren, trends ontdekken, voorspellen bij lijngrafieken

- Enkelvoudige lijngrafieken
- Meervoudige lijngrafieken

Verschillende informatiebronnen combineren

- Tabel met tabel(len)
- Tabel met grafiek/diagram
- Grafiek/diagram met grafiek/diagram

De inhoud van het domein Verhoudingen

Hieronder staan de doelen voor dit domein beschreven.

- Verhoudingen (inclusief percentages en breuken) herkennen, benoemen en gebruiken
- Verhoudingsproblemen oplossen met hele getallen, kommagetallen en breuken (globaal en exact)
- Werken met schaallijn en schaal:
 - lengte bepalen met behulp van schaallijn / liniaal
 - lengte/afstand bepalen op basis van schaal
 - schaal bepalen
- Verhoudingen vergelijken
- Bij verdelingen percentages berekenen op basis van geheel is 100%
- Rekenen met percentages via rekenen met breuken
- Berekenen hoeveel procent het deel is
- Rekenen met percentages boven de 100
- Omzetten, verhoudingen – procenten, procenten – kommagetallen, verhoudingen – breuken

3.2.1.2 Lezen

Begripsomschrijving

Lezen is een complex proces waarbij de lezer betekenis toekent aan de tekst. De lezer heeft daarin een actieve rol. Hij construeert betekenis op basis van de informatie in de tekst én op basis van zijn eigen kennis. Er is hierbij sprake van een continue wisselwerking tussen verschillende verwerkingsprocessen: de verwerking van informatie die is weergegeven in de tekst, de inzet van eigen kennis en, uiteindelijk, het integreren van de informatie uit de tekst met de eigen kennis. Wanneer deze processen succesvol verlopen, vormt de lezer uiteindelijk een samenhangende mentale representatie van de tekst.

Voor het begrip van een tekst vertrouwen lezers in eerste instantie op de automatische processen die in werking treden tijdens het lezen. Wanneer dit echter niet leidt tot voldoende begrip moeten zij meer bewust sturing geven aan het proces van betekenisconstructie. Goede lezers maken in die gevallen gebruik van leesstrategieën, zoals het opzoeken en samenvatten van informatie in de tekst en de inzet van voorkennis over een onderwerp.

Lezers doen een beroep op verschillende vaardigheden:

- begrijpend lezen;
- opzoeken van informatie;
- samenvatten van teksten.

Het verwerkingsproces dat de lezer inzet om de vraag te kunnen beantwoorden, vormt bij deze indeling het uitgangspunt.

Begrijpend lezen

In het toekennen van betekenis aan gesproken taal spelen zowel tekst- als kennisgestuurde verwerkingsprocessen een belangrijke rol.

Bij tekstgestuurde verwerking staat de inhoud van de tekst centraal en verwerkt de lezer de informatie die de schrijver expliciet in de tekst vermeldt. Er is dan sprake van het 'begrijpen van geschreven teksten'. Om tot begrip van de tekst te komen, maakt de lezer gebruik van de inhoud (de betekenis van woorden, woordgroepen, zinnen, langere tekstpassages en hun onderlinge betekenisrelaties), van expliciete relaties tussen elementen in een tekst (woord- en zinsvolgorde, verwijzingen en talige structuurmarkeerders) en van de expliciete structuur van een tekst.

Kennisgestuurde verwerking gaat verder. Van werkelijk en diepgaand tekstbegrip is pas sprake wanneer de lezer informatie uit de tekst verbindt aan zijn 'eigen' kennis. Een lezer benadert een tekst niet blanco, maar zet bij de verwerking van de tekst ook zijn eigen kennis in, waaronder zijn kennis van de wereld en zijn kennis van taal en tekstkenmerken. Dit samenspel van verwerkingsprocessen noemen we het 'interpreteren van geschreven teksten'. Het onderkennen en afleiden van impliciete informatie in een tekst en het maken van inferenties zijn hierbij belangrijk.

Tussen tekst- en kennisgestuurde verwerking vindt een continue wisselwerking plaats. Pas wanneer ze in samenhang en gelijktijdig ingezet worden, is er sprake van diepgaand tekstbegrip.

Opzoeken

Bij het opzoeken van informatie gaat het om het efficiënt en systematisch opsporen van specifieke informatie in schriftelijke of elektronische informatiebronnen om deze vervolgens te gebruiken of toe te passen. Om de beoogde informatie te kunnen vinden, zal de lezer selectief, flexibel en doelgericht te werk moeten gaan. Soms zal hij de informatie grondig lezen, soms vooral scannend. Het doelgericht zoeken naar informatie vereist specifieke vaardigheden, waarbij de keuze en combinatie van de juiste informatiebronnen en het doorgronden van de structuur van de informatiebron een centrale plaats innemen. Om de veelheid van gedrukte en digitale informatiebronnen – waaronder tijdschriften, kranten, reclame- en voorlichtingsmateriaal, brochures, bedrijven- en gemeentegidsen, reisgidsen, telefoongidsen, OV-informatie, handleidingen, bijsluiters, beeldbanken, websites en woordenboeken – te kunnen hanteren is het noodzakelijk dat lezers weten dat informatie alfabetisch geordend is of op een andere logische manier is ingedeeld, bijvoorbeeld op plaats of onderwerp. Daarnaast is het kunnen formuleren van zoekvragen, trefwoorden en synoniemen hier van belang.

Samenvatten

Samenvatten is een vaardigheid die gerelateerd is aan het leren studeren. Het gaat erom dat er na het begrijpen en interpreteren van de informatie iets met die informatie gedaan moet worden: hoofd- en bijzaken moeten worden onderscheiden, de tekst moet worden verkort en de informatie moet op de een of andere manier opgeslagen worden om deze op een ander moment paraat te hebben, bijvoorbeeld voor een proefwerk, een werkstuk of een mondelinge presentatie. Daarvoor kan de lezer de informatie bewerken, in de vorm van herlezen, onderstrepen, aantekeningen maken, samenvatten of schematiseren. Leerlingen die

samenvatten, structureren de informatie en stellen de bruikbaarheid van de gelezen tekst vast. Ze zetten de tekst om in een verkorte tekst of in een schema of tabel.

Bij het vaststellen van de bruikbaarheid van een tekst zet de leerling de informatieve waarde en de betrouwbaarheid van een tekst bijvoorbeeld af tegen de achtergrond van een praktisch doel.

Referentieniveaus

Het Referentiekader Taal en Rekenen (Ministerie van Onderwijs, Cultuur en Wetenschap, 2009) brengt in kaart wat leerlingen moeten kennen en kunnen op het gebied van onder andere leesvaardigheid. Voor leerlingen in de bovenbouw van het basisonderwijs zijn de niveaus 1F en 2F belangrijk.

Inhoudsverantwoording

De inhoud van de teksten, informatiebronnen en opgaven in de taken Lezen is gebaseerd op verschillende bronnen. Ze vormen de basis voor de verschillende inhoudsaspecten en opgaventypen die in de taken Lezen zijn opgenomen. Het gaat om de volgende bronnen:

- De tussendoelen gevorderde geletterdheid voor de middenbouw. Leerlijnen voor groep 4 tot en met groep 8 (Aarnoutse & Verhoeven, 2003);
- De kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006);
- De methoden taal/lezen die in het basisonderwijs gebruikt worden;
- De referentieniveaus Taal voor het basisonderwijs (Expertgroep doorlopende leerlijnen taal en rekenen, 2009a)
- Het document 'Over drempels met Taal en Rekenen' (Expertgroep doorlopende leerlijnen taal en rekenen 2008a, 2008b);
- Het document 'Een nadere beschouwing' (Expertgroep doorlopende leerlijnen taal en rekenen, 2009b).

In de tabellen 3.4 en 3.5 is aangegeven hoe de opgaven zijn verdeeld over de domeinen en referentieniveaus in de taken van het onderdeel Lezen. In tabel 3.4 is de gewenste verdeling over de domeinen Begrijpend lezen, Opzoeken en Samenvatten aangegeven door deze (tussen haakjes) op te nemen achter de aantallen gerealiseerde opgaven. In alle gevallen is het aantal gerealiseerde opgaven exact gelijk aan het aantal gewenste opgaven. Aanvullende informatie over de inhoudsvaliditeit in termen van gewenste en gerealiseerde opgavencategorieën volgt hieronder bij de inhoudelijke bespreking.

Tabel 3.4 De drie domeinen met aantallen opgaven in de taken Lezen

	Basis	aantal opgaven	Module Verdieping	aantal opgaven	Totaal
Begrijpend lezen	taak 3	20 (20)	taak 10	20 (20)	40 (40)
Opzoeken van informatie	taak 7	10 (10)	taak 13	10 (10)	20 (20)
Samenvatten van teksten	taak 7	10 (10)	taak 13	10 (10)	20 (20)

In tabel 3.5 staan de percentages en aantallen opgaven op niveau 1F en 2F die zijn opgenomen in de taken Lezen. Daarbij zijn de aantallen opgaven uitgesplitst per domein, voor de basistaken en de verdiepende taken afzonderlijk. Vooraf is aangegeven dat voor beide referentieniveaus in de basismodule tenminste 15, maar bij voorkeur 20 opgaven zouden kunnen worden gekenschetst als behorend bij 1F, respectievelijk 2F. Wanneer alleen de basismodule wordt afgenomen, zien we dat aan die voorwaarde niet wordt voldaan voor referentieniveau 1F. Er zijn slechts tien 1F-opgaven, waar er minimaal 15 gewenst waren. Het is erg lastig gebleken om opgaven te construeren op 1F-niveau, vooral voor Begrijpend lezen en zeker ook voor Samenvatten: opgaven voor Samenvatten zijn per definitie te karakteriseren als 2F. Uiteindelijk is het niet zo'n probleem dat er zo weinig 1F-opgaven zijn. De vaardigheidsschalen voor 1F en 2F bleken samen te vallen: alle opgaven liggen op één schaal. Wanneer ook de module Verdieping is afgenomen, wordt ruimschoots aan de minimum- en voorkeurseisen voldaan: 23 opgaven zijn op niveau 1F, 57 op niveau 2F.

Tabel 3.5 De aantallen opgaven en percentages 1F en 2F van de referentieniveaus in de taken Lezen

	Gerealiseerde verdeling	
	1F	2F
Taken Lezen Basistoets	10 (25%)	30 (75%)
Begrijpend lezen	6	14
Opzoeken	4	6
Samenvatten	0	10
Taken Lezen Module Verdieping	13 (32%)	27 (68%)
Begrijpend lezen	11	9
Opzoeken	2	8
Samenvatten	0	10
Totaal aantal (% gemiddeld)	23 (29%)	57 (71%)

De inhoud van het domein Begrijpend lezen

De teksten

Tekstsoorten

Het belangrijkste onderscheid bij *tekstsoorten* is dat tussen zakelijke teksten enerzijds en literaire en fictionele teksten anderzijds. Bij de zakelijke teksten is verder onderscheid te maken tussen informatieve, instructieve en betogende teksten en bij literaire en fictionele teksten tussen verhalende en poëzieteksten:

- Informatieve teksten (zakelijke teksten): De schrijver geeft feitelijke informatie over de werkelijkheid. Soms voegt hij er ter verduidelijking ook zijn eigen opvattingen aan toe.
- Instructieve teksten (zakelijke teksten): De schrijver wil het gedrag, het handelen van de lezer richten en sturen. De schrijver zorgt ervoor dat de lezer de instructie goed kan uitvoeren.
- Betogende teksten (zakelijke teksten): De schrijver wil vooral het denken van de lezer beïnvloeden. Hij roept de lezer op om iets te doen of juist iets te laten via een brief, oproep of aankondiging.
- Verhalende teksten (fictionele teksten): De schrijver beschrijft een verbeelde werkelijkheid in verhalen. Hij bedenkt bijvoorbeeld een spannend of grappig verhaal.
- Poëzieteksten (fictionele teksten): De schrijver beschrijft een verbeelde werkelijkheid in poëzie.

De taken Begrijpend lezen in de Entreetoets bevatten zowel korte als langere teksten. De teksten zijn inhoudelijk en qua moeilijkheid afgestemd op de leerlingen in groep 7 van het basisonderwijs. Ze sluiten aan bij de leefwereld van de leerlingen. In de uiteindelijke versie van de toets zijn er geen zes teksten (zoals beoogd) opgenomen, maar slechts vijf. Daardoor werd maar één fictionele tekst gerealiseerd in plaats van twee. In de verdiepende module werd dit gecompenseerd door evenveel zakelijke als functionele teksten op te nemen, zodat uiteindelijk zeven zakelijke en vier fictionele teksten werden gerealiseerd.

In tabel 3.6 staat de verdeling van de teksten van het domein Begrijpend lezen naar tekstsoort.

Tabel 3.6 Percentages en aantallen zakelijke en fictionele teksten in domein Begrijpend lezen

Tekstsoort	Taak 3	Taak 10	Totaal
zakelijke teksten	80% (4)	50% (3)	64% (7)
fictionele teksten	25% (1)	50% (3)	36% (4)

Tekstgenres

Bij het onderscheid in tekstsoorten gaat het om functies of doelen van teksten. Teksten kunnen ook op grond van de vorm van de tekst ingedeeld worden: het tekstgenre. Tekstgenres waar basisschoolleerlingen zoal mee te maken krijgen zijn: nieuwsbericht, artikel, verslag, interview, e-mail, brief, verhaal, mop en gedicht.

Elke combinatie van tekstsoort en tekstgenre heeft specifieke kenmerken wat betreft opbouw, stijl, register, doel, publiek, taalgebruik, conventies, mate van formaliteit en presentatie.

Vaak hebben tekstgenre en tekstsoort een een-op-een-relatie: een nieuwsbericht informeert altijd; in een instructie wordt het handelen van de lezer gestuurd. Soms is de relatie echter complexer: in een brief kan geïnformeerd worden, maar er kan ook iets in betoogd worden en er kunnen allerlei mogelijke aanwijzingen of instructies in gegeven worden.

In tabel 3.7 staat welke tekstgenres in de taken Begrijpend lezen staan. Er werden vooraf geen harde streefaantallen genoemd ten aanzien van de verdeling over tekstgenres, wel werd zoveel mogelijk spreiding over de genres nagestreefd. Uiteindelijk komen alle genres tenminste éénmaal en in twee gevallen twee maal voor.

Tabel 3.7 De tekstgenres in de taken Begrijpend lezen

Tekstgenre	Percentage (aantal teksten)
nieuwsbericht	18% (2)
artikel	9% (1)
verslag	9% (1)
interview	9% (1)
e-mail	9% (1)
brief	9% (1)
verhaal	18% (2)
mop	9% (1)
gedicht	9% (1)

De opgaven

Opgaventypen

Bij het domein Begrijpend lezen onderscheiden we de volgende opgaven:

- *vragen over teksten*: traditioneel zijn de opgaven Begrijpend lezen vragen over teksten: op basis van de tekst of een deel van de tekst worden vragen gesteld. Dit type opgaven wordt ook in de leesmethoden veel gehanteerd.
- *openplaatsopgaven*: hierin gaat het om teksten waaruit zinnen of delen van zinnen ontbreken. De leerlingen moeten het stuk vóór en ná het ontbrekende stukje tekst lezen en geven aan welke zin of welk deel van de zin op de open plaats zou moeten staan.
- *voorspelopgaven*: opgaven bij teksten waarbij slechts de titel en het begin van de tekst worden gegeven; de rest van de tekst ontbreekt. De leerlingen beantwoorden op basis van het gegeven stukje tekst een vraag, waarbij ze als het ware een 'voorspelling' moeten doen over de inhoud of de bron van de tekst of de tekstsoort.

In tabel 3.8 zijn per opgaventype de gerealiseerde percentages en aantallen opgaven opgenomen. Tussen haakjes staan de aantallen opgaven die op voorhand werden nagestreefd. Er werd geprobeerd om 28 (70%) opgaven de vorm te geven van "vragen over teksten"; dat werden er uiteindelijk 27 in plaats van 28. In overeenstemming met de wensen ligt het accent dus sterk op vragen over de tekst. Er werden wat meer openplaatsopgaven gerealiseerd dan beoogd (elf in plaats van acht) en wat minder voorspelopgaven dan beoogd (twee in plaats van vier). De afwijkingen

kunnen vooral verklaard worden dat keuzes niet alleen op opgaveniveau, maar ook op het niveau van teksten worden gemaakt. Dit impliceert dat men bij een 'goede' tekst soms een openplaatsopgave in plaats van een voorspelopgave moet accepteren of omgekeerd. Voor de inhoudsvaliditeit van de toets is dit nauwelijks een probleem.

Tabel 3.8 Percentage en aantal verschillende opgaventypen in het domein Begrijpend lezen

Opgaventype	Taak 3	Taak 10	Totaal
vragen over teksten	70% 14 (14)	65% 13 (14)	68% 27 (28)
Openplaatsopgaven	25% 5 (4)	30% 6 (4)	27% 11 (8)
voorspelopgaven	5% 1 (2)	5% 1 (2)	5% 2 (4)

Verwerkingsprocessen

De opgaven Begrijpend lezen hebben betrekking op een woord(groep) of op een of enkele zinnen. Ze kunnen ook betrekking hebben op grotere tekstdelen of op de hele tekst.

We hebben de opgaven ingedeeld op basis van de verwerkingsprocessen 'begrijpen van geschreven teksten' en 'interpreteren van geschreven teksten' én op basis van de inhoud die in het begrijpend leesonderwijs aan bod komen. Deze inhoud hebben we vertaald in een aantal inhoudsaspecten die we gerelateerd hebben aan de verwerkingsprocessen. Op die manier zijn we gekomen tot een aantal verschillende opgaventypen. Elk opgaventype representeert een of meer inhoudsaspecten.

Onderstaande indelingen laten voor elk van de beide verwerkingsprocessen zien welke inhoudsaspecten aan welk verwerkingsproces gerelateerd zijn. Daarbij willen we benadrukken dat de verwerkingsprocessen in werkelijkheid niet altijd even duidelijk van elkaar te scheiden zijn.

De indeling in opgaven bij *begrijpen* van geschreven teksten is als volgt:

- Opgaven die vragen naar de betekenis van een woord, woordgroep of zin(nen) die expliciet in de tekst vermeld wordt (worden).
- Opgaven die vragen naar specifieke inhoudselementen die expliciet in de tekst aan de orde gesteld worden. Dit zijn bijvoorbeeld feiten en meningen, voorwerpen, aantallen, plaats van handeling of tijdsperiodes, (hoofd)personen.
- Opgaven die vragen naar eenvoudige expliciete verbanden op lokaal niveau. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen zoals signaalwoorden. Voorbeelden van verbanden zijn verwijzingen, vergelijkingen, tegenstellingen, generalisaties en voorbeelden, oorzaak en gevolg, vraag en antwoord, reden en verklaring, middel en doel, deel-/geheelrelaties, conclusie en argumenten.
- Opgaven die vragen naar complexe expliciete verbanden over grotere tekstdelen heen. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen. Voorbeelden van verbanden zijn vergelijkingen, tegenstellingen, volgordes, generalisaties en voorbeelden, oorzaak en gevolg, vraag en antwoord, reden en verklaring, middel en doel, deel-/geheelrelaties, conclusie en argumenten.

De indeling in opgaven bij *interpreteren* van geschreven teksten is als volgt:

- Opgaven die vragen naar het afleiden van de betekenis van een woord, woordgroep of zin(nen).
- Opgaven die vragen naar het afleiden van informatie uit de tekst op lokaal niveau. De leerling moet zijn voorkennis inzetten naast de informatie die de tekst geeft. Voorbeelden hiervan zijn opgaven over onderwerp, thema, hoofdlijnen, hoofdgedachte, hoofdpersoon, setting en doel en doelgroep van de tekst, het vertelperspectief en de tekststructuur.
- Opgaven die vragen naar het afleiden van informatie op het globale niveau van de tekst. Verbanden kunnen worden gelegd op basis van inhoudelijke en/of structurele elementen. Voorbeelden hiervan zijn opgaven over onderwerp, thema, hoofdlijnen, hoofdgedachte, hoofdpersoon, setting en doel en doelgroep van de tekst, het vertelperspectief en de tekststructuur.

- Opgaven die vragen naar het taalgebruik en de schrijfstijl van de schrijver en waar verbanden moeten worden gelegd tussen tekstuele informatie en kennis van het taalsysteem. Voorbeelden hiervan zijn vragen naar tekstsoort, naar het gebruik van aanhalingstekens of naar specifieke woordkeuze waarbij register, sociale en culturele conventies een rol spelen.

Bij het samenstellen van de taken Begrijpend lezen zijn we ervan uitgegaan dat de verwerkingsprocessen 'begrijpen van geschreven teksten' en 'interpreteren van gesproken teksten' beide een belangrijke rol in het leesproces vervullen. Echter, het interpreteren van geschreven teksten neemt in de bovenbouw van het basisonderwijs een steeds belangrijker plaats in. Dit was voor ons reden om een 30:70 verhouding in de verdeling van opgaven na te streven. In taak 10 is dit precies zo gerealiseerd, in taak 3 ligt het accent nog iets sterker dan bedoeld op de interpretatie-opgaven (17 opgaven met de focus op interpretatie, drie met de focus op begrijpen). De verdeling van de gerealiseerde opgaven binnen het domein Begrijpend lezen over de beide verwerkingsprocessen staat in tabel 3.9. Het merendeel van de opgaven doet dus, zoals de bedoeling was, een beroep op de interpretatie van de gelezen informatie. Het relatief grote aantal interpretatie-opgaven moet ook hier vooral begrepen worden vanuit het gegeven dat keuzes niet alleen op opgaveniveau, maar ook op het niveau van teksten moeten worden gemaakt. Daarnaast is het een weerspiegeling van het gegeven dat het erg lastig is om voor eind groep 7 opgaven te maken die uitsluitend een beroep doen op begrip (en niet op interpretatie).

Tabel 3.9 Percentages en aantal opgaven begrijpen en interpreteren in het domein Begrijpend lezen

Verwerkingsprocessen	Taak 3	Taak 10	Totaal
Begrijpen	15% 3 (6)	30% 6 (6)	23% 9 (12)
Interpreteren	85% 17 (14)	70% 14 (14)	77% 31 (28)

De inhoud van het domein Opzoeken

De informatiebronnen

Informatiebronnen kunnen worden onderverdeeld in papieren en digitale bronnen.

In de taken Lezen zijn voor het domein Opzoeken de volgende papieren informatiebronnen en ingangen tot papieren informatiebronnen opgenomen:

- (spreek)woordenboek;
- boekomslag of voorblad van een boek;
- inhoudsopgave.

We hebben voor deze papieren informatiebronnen gekozen, omdat leerlingen hiermee regelmatig in aanraking komen. Inhoudsopgaven en lemma's van woordenboeken komen de leerlingen bovendien ook tegen op het internet.

Door de toenemende digitalisering komen leerlingen steeds vaker in aanraking met informatiebronnen op het internet: zij zoeken informatie op via zoekmachines en op websites. Omdat het gebruik van digitale informatiebronnen toeneemt, ligt bij het domein Opzoeken daar de nadruk op.

De digitale bronnen die aan bod komen zijn:

- website;
- zoekmachine;
- zoekwoorden.

Ook het maken van een keuze tussen het gebruik van papieren en digitale informatiebronnen behoort tot de kennis waarover leerlingen dienen te beschikken.

In tabel 3.10 is de verdeling van de typen informatiebronnen in het domein Opzoeken te vinden. Het was de bedoeling om de verschillende informatiebronnen allemaal aan de orde te laten komen in de opgaven, in één of twee opgave(n) per taak. Dat is niet helemaal gelukt: in taak 13 zit geen opgave van type "(spreekwoorden)boek" en in taak 7 geen opgave van type "bron selecteren". Verder komen alle

informatiebronnen en opgaventypen in beide taken een à drie maal aan bod. Het type “website” komt zes maal voor (in plaats van de beoogde vier maal), het type “zoekmachine” twee maal (in plaats van de beoogde vier maal). De verhouding tussen de verschillende informatiebronnen komt aardig in de buurt van de gewenste verdeling.

Tabel 3.10 Type informatiebronnen bij het domein Opzoeken

Informatiebron	Type	Taak 7	Taak 13	Totaal
papieren bronnen	(spreekwoorden)boek	20% 2 (1)	----- - (1)	10% 2 (2)
	boekomslag, voorblad	20% 2 (1)	10% 1 (1)	15% 3 (2)
	inhoudsopgave	10% 1 (1)	20% 2 (1)	15% 3 (2)
digitale bronnen	website	30% 3 (2)	30% 3 (2)	30% 6 (4)
	zoekmachine	10% 1 (2)	10% 1 (2)	10% 2 (4)
	zoekwoorden	10% 1 (2)	20% 2 (2)	15% 3 (4)
papier en digitaal	bron selecteren	----- - (1)	10% 1 (1)	5% 1 (2)

De opgaven

De opgaven in het domein Opzoeken zijn onderverdeeld in twee categorieën: opgaven waarbij een gegeven informatiebron gebruikt moet worden en opgaven waarin het gaat om het inzicht in hoe bronnen in elkaar zitten.

In de opgaven uit de categorie ‘gebruik informatiebron’ wordt een informatiebron weergegeven, bijvoorbeeld een deel van een spreekwoordenboek of een boekomslag. Het antwoord op de vraag moet in de bron worden opgezocht. Het gaat bij deze opgaven om het juiste gebruik van de gegeven informatiebron.

Opgaven uit de categorie ‘inzicht in bronnen’ hebben betrekking op de aard of de structuur van de informatiebronnen. Er hoeft geen antwoord op de vraag gegeven te worden. Het gaat erom dat de leerlingen de volgende zaken kunnen aangeven:

- het soort informatiebron waarin het antwoord op de zoekvraag te vinden is;
- de plaats in een informatiebron waar het antwoord op de zoekvraag te vinden is; de manier waarop het antwoord op de zoekvraag te vinden is.

De inhoud van het domein Samenvatten

De teksten

Bij het domein Samenvatten staat het lezen van teksten die worden ingezet voor een praktisch doel centraal. Het gaat om teksten die de leerlingen kunnen gebruiken voor het maken van een werkstuk of het voorbereiden van een spreekbeurt. De teksten die in het domein Samenvatten worden ingezet zijn altijd informatieve teksten, waarin informatie over een bepaald onderwerp wordt overgedragen. Dit onderwerp is duidelijk afgebakend: in elke tekst staat namelijk één onderwerp centraal. De teksten hebben een heldere structuur die het mogelijk maakt de tekst samen te vatten.

De opgaven

De opgaven in het domein Samenvatten worden gesteld bij een tekst als geheel of bij een alinea. Bij opgaven over een alinea wordt vermeld op welke alinea deze betrekking heeft. Er wordt niet naar aparte zinnen verwezen.

De opgaven doen een beroep op verschillende aspecten die bij het samenvatten van een tekst belangrijk zijn, namelijk:

- het bepalen van de kern van een tekst(deel);
- het verkort weergeven van informatie uit de tekst;
- het vaststellen van het doel van een tekst.

Voor de aspecten onderscheiden we verschillende opgaventypen. Het eerste type opgaven zijn opgaven die vragen naar het bepalen van de kern van een tekst(deel) oftewel de hoofdzaak van een tekst. Daarbij gaat het om het benoemen van het onderwerp van een tekst(deel). Als het gaat om het verkort weergeven van informatie uit de tekst, kan dat zowel tekstueel als grafisch. Opgaven waarin gevraagd wordt naar de best passende samenvatting in woorden, naar de best passende tekstkop of naar een zin die de tekst of het tekstdeel het best samenvat zijn tekstueel.

Een tekst(deel) kan ook grafisch worden samengevat; in een afbeelding, tabel of schema. In opgaven met afbeeldingen vragen we welke foto of tekening de kern van een tekst(deel) het best weergeeft.

Een tabel is bedoeld om gegevens met elkaar te vergelijken en wordt daarom ingezet in opgaven bij teksten waarin verschillen worden benadrukt. Een tekst(deel) kan ook schematisch worden samengevat. Daarbij kunnen verschillende soorten schema's ingezet worden. Denk aan boomschema's waarin tekstblokken op een logische manier met elkaar worden verbonden en die passen bij de structuur van de tekst. Andere varianten zijn lineaire en circulaire schema's. In opgaven waarin schema's of tabellen verwerkt zijn, moet vaak een lege plek ingevuld worden. De leerlingen moeten daarvoor de structuur van het schema of van de tabel doorgronden en deze koppelen aan die van de tekst. Opgaven kunnen ook vragen naar het schema dat het best een tekst(deel) weergeeft.

Tot slot zijn er opgaven die vragen naar het tekstdoel. Dat hangt altijd samen met de kern van de tekst. Er wordt dan bijvoorbeeld gevraagd of een tekst bruikbaar is in een bepaalde situatie.

In tabel 3.11 is de verdeling over opgaventypen bij het domein Samenvatten te zien.

Tabel 3.11 Aantal vragen per opgaventype Samenvatten

Opgaventype	Taak 7	Taak 13	Totaal
benoemen onderwerp	----- - (1)	20% 2 (1)	10% 2 (2)
samenvatting in woorden	30% 3 (2)	10% 1 (1)	20% 4 (3)
samenvattende tekstkop	20% 2 (2)	10% 1 (1)	10% 3 (3)
samenvattende zin	----- - (1)	20% 2 (2)	10% 2 (3)
grafische samenvatting door een afbeelding	10% 1 (1)	10% 1 (1)	10% 2 (2)
grafische samenvatting door tabel	----- - (1)	10% 1 (1)	5% 1 (2 of 3)
grafische samenvatting door schema	20% 2 (1)	20% 2 (2)	20% 4 (2 of 3)
vraag naar tekstdoel	20% 2 (1)	----- - (1)	15% 2 (2)

In de tabel is te zien dat het de bedoeling was om opgaven op te nemen die gespreid waren over alle opgaventypen. In taak 7 is dat niet helemaal gelukt: in deze taak zitten geen opgaven over 'benoemen onderwerp', 'samenvattende zin' en 'grafische samenvatting door tabel'. In taak 13 is hiervoor gecompenseerd, zodat de basis- en verdiepende toets samen de bedoelde optimale spreiding van opgaven wél laten zien. De gerealiseerde aantallen per opgaventype wijken nergens meer dan één af van de bedoelde aantallen. Ten aanzien van de categorieën die betrekking hebben op grafische samenvatting zijn de aantallen opgaven met een tabel, dan wel schema wat onevenwichtig verdeeld (één respectievelijk vier opgave(n) waar twee à drie de bedoeling was).

3.2.1.3 Taalverzorging

Begripsomschrijving

Schriftelijke vaardigheden zijn van groot belang om goed te functioneren en te communiceren in de samenleving. Taalverzorging is een fase van het schrijfproces waarbij de schrijver alle taalvaardigheden en kennis inzet om tot een verzorgde schriftelijke taalproductie te komen. In het schrijfproces speelt taalverzorging met name een rol in de laatste fase, als de inhoud en de structuur van een tekst al ver gevorderd zijn en de tekst zijn voltooiing nadert.

Uit onderzoek is gebleken dat het produceren van een correct geschreven tekst een behoorlijke cognitieve krachtsinspanning vraagt van zowel beginnende als gevorderde schrijvers. Daarom is het belangrijk dat een schrijver onder andere beschikt over automatismen op het gebied van grammaticale vaardigheid, interpunctie en spelling. Hij kan zijn aandacht dan met name richten op de inhoud van de tekst. Goede taalverzoorgers maken gebruik van grammaticale regels, interpunctieregels en spellingregels en/of -strategieën om tot een verzorgd schrijfproduct te komen.

De verschillende aandachtsgebieden van het onderdeel Taalverzorging zijn verdeeld over de vier domeinen die in het taalonderwijs op de basisschool aan bod komen: Grammatica, Interpunctie, Spelling niet-werkwoorden en Spelling werkwoorden.

Grammatica

Grammatica wordt gerekend tot het domein van de taalbeschouwing en kan heel breed gezien worden als het leren reflecteren op het taalsysteem en op de betekenis en functies van taal. Vanwege het reflecteren op taal zijn in de begrippenlijst van het Referentiekader Taal en Rekenen (Ministerie van Onderwijs, Cultuur en Wetenschap, 2009) grammaticale begrippen opgenomen. In het onderwijs zien we naast deze reflecterende kant dat grammatica met name gericht is op redekundig en taalkundig ontleden.

Het grammaticaonderwijs richt zich op het verschaffen van kennis van grammaticale regels en relaties.

Van belang is het kunnen toepassen van kennis over grammaticale relaties binnen woorden, tussen woorden, binnen zinnen en tussen zinnen. Het toepassen van die kennis in de vorm van het correct formuleren van woorden, zinsdelen en zinnen is een voorwaarde om goed te kunnen schrijven.

Grammaticaal inzicht en kennis van de begrippen beschouwen we niet alleen als een aparte vaardigheid binnen Taalverzorging maar ook als een onderliggende vaardigheid. Bij het aanleren van de werkwoordspelling zijn kennis van en inzicht in de grammaticale begrippen van wezenlijk belang. Om een werkwoordsvorm goed te kunnen spellen, moet de leerling de functie ervan in de zin kunnen vaststellen. Zo is het noodzakelijk dat de leerling onderscheid kan maken tussen de persoonsvorm en het onderwerp om werkwoorden juist te kunnen spellen. Ook bij interpunctie is inzicht in zinsbouw en taalstructuur een vereiste; een schrijver moet zijn grammaticale kennis immers inzetten om leestekens in en tussen zinnen te plaatsen.

Interpunctie

Onder Interpunctie verstaan we het plaatsen van leestekens in geschreven taal, gebruikmakend van regels en conventies. Het primaire doel van het plaatsen van leestekens is het bevorderen van de leesbaarheid en het tekstbegrip voor de lezer. De lezer krijgt meer grip op een tekst omdat leestekens helpen bij het doorgronden van de structuur. Bij gesproken taal geven intonatie en pauzes aan waar de accenten liggen en ondersteunen zo de luisteraar bij het interpreteren. In geschreven taal ontbreekt deze ondersteuning en heeft de lezer een hulpmiddel nodig in de vorm van leestekens.

Leestekens kunnen geplaatst worden tussen zinnen om zinsgrenzen aan te geven. Denk hierbij aan punten, hoofdletters en vraagtekens. Ook zijn er leestekens die in een zin aangebracht worden om het verband tussen woordgroepen of delen van een zin aan te geven, bijvoorbeeld komma's en dubbele punten. Verder gebruiken schrijvers markeerders van citaten of bijzondere woorden, de zogenoemde aanhalingstekens.

Om leestekens correct te kunnen plaatsen en zo de lezer te kunnen ondersteunen is het van belang dat de schrijver inzicht in de opbouw van een zin en/of tekst heeft.

Spelling niet-werkwoorden

Spelling omvat de notatiesystematiek voor zowel niet-werkwoorden als werkwoorden. Bij het spellen gaat het erom dat de leerling van gesproken woorden 'schriftbeelden' (dus geschreven woorden) kan maken. Bij Spelling niet-werkwoorden maken we onderscheid tussen klankzuivere en niet-klankzuivere woorden. De eerste fase van het schrijfonderwijs richt zich op het correct leren schrijven van de klankzuivere woorden; de leerling schrijft op wat hij hoort (begin groep 3). Snel daarna komen de niet-klankzuivere woorden. Dat zijn de woorden waarbij er geen eenduidige relatie is tussen klank en letter. Het gaat dan om woorden zoals *bomen*, *trein*, *begin*. Om die goed te kunnen schrijven moet de leerling regels kunnen toepassen of weten dat hij de letters in een woord net zo moet schrijven als in een ander woord dat hij al kent (vanaf eind groep 3).

Naast de actieve kant van spelling (het zelf foutloos woorden schrijven) kent spelling ook een passieve kant, namelijk het herkennen en verbeteren van fouten in geschreven tekst: tekstrevisie. Het zelf kunnen schrijven van begrijpelijke en correct gespelde teksten is een belangrijke communicatieve vaardigheid. Een tekst zonder spelfouten maakt, op welk niveau dan ook, een heel andere indruk dan een tekst met spelfouten. Om die reden is ook het passief spellen van belang: als leerlingen een tekst geschreven hebben, staan daar meestal nog fouten in. Het is dan zaak te leren om die fouten op te sporen en te verbeteren. In de bovenbouw van het basisonderwijs worden leerlingen steeds meer geacht hun eigen schrijfwerk – en vaak ook dat van medeleerlingen – na te kijken.

Spelling werkwoorden

Bij werkwoordspelling gaat het om het spellen van veranderlijke woorden. Dat wil zeggen dat onder invloed van het zinsverband de schrijfwijze kan veranderen. Horen we het woord 'bied', dan kunnen we dit op twee manieren schrijven: *bied* als het onderwerp *ik* is en *biedt* als het onderwerp *hij* is. Dit betekent dat naast kennis van de algemene beginselen van de Nederlandse spelling, het voor werkwoordspelling noodzakelijk is om inzicht te hebben in grammaticale principes. Om een werkwoordsvorm correct te spellen, moeten leerlingen de relatie tussen onderwerp en de persoonsvorm herkennen. Verder is het belangrijk dat leerlingen de grondvorm van werkwoorden kunnen onderscheiden en dat ze de tijdsvorm verleden tijd herkennen. Werkwoorden met een voorvoegsel nemen daarbij een aparte positie in (*be-*, *ge-*, *ver-*, *her-* en *ont-*).

Referentieniveaus

Het Referentiekader Taal en Rekenen (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009a) beschrijft het taaldomein: *Begrippenlijst en taalverzorging*. De referentieniveaus voor het onderdeel Taalverzorging brengen in kaart wat leerlingen moeten kennen en kunnen op het gebied van taalverzorging. Voor leerlingen in de bovenbouw van het basisonderwijs zijn de niveaus 1F en 2F belangrijk.

Inhoudsverantwoording

De inhoud van de opgaven in de taken Taalverzorging is gebaseerd op verschillende bronnen, waaronder:

- de tussendoelen gevorderde geletterdheid voor de middenbouw (Aarnoutse & Verhoeven, 2003);
- de kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006);
- de referentieniveaus Taal voor het basisonderwijs (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009a);
- het document 'Over de drempels met Taal en Rekenen' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008b);
- het document 'Over de drempels met taal' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008a);
- het document 'Een nadere beschouwing' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009b);
- het document 'Leerstoflijnen begrippenlijst en taalverzorging beschreven' van SLO (van der Beek & Paus, 2011);

- de diverse spelling- en taalmethoden die in het basisonderwijs gebruikt worden (zie ook het methode-onderzoek in de LVS-toetsen Taalverzorging (Cito, 2015);
- het boek 'Spelling' van Huizenga (2010), een standaardwerk op het gebied van de spellingdidactiek;
- Taalunie (2009);
- digiwak.nl.

De opgaven in het onderdeel Taalverzorging weerspiegelen de leerstof van de meest gebruikte taalmethoden in het basisonderwijs, zowel qua inhoud als qua volgorde van aanbod. Bij het vaststellen van de te toetsen categorieën voor grammatica, interpunctie en spelling is gestreefd naar aansluiting bij het merendeel van de taalmethoden. De passieve kant van spelling wordt getoetst, zowel bij niet-werkwoorden als bij werkwoorden. Dit sluit aan bij het concept taalverzorging zoals dat ook in de LVS-toetsen Taalverzorging van groep 6 tot en met 8 wordt gemeten. Actieve spelling komt in de LVS-toetsen Spelling 3.0 aan bod.

In tabel 3.12 is te zien welke domeinen in de taken zijn opgenomen. De verdeling van de gerealiseerde aantallen opgaven over de vier domeinen komt overeen met de vooraf geformuleerde wensen omtrent de toetsmatrijs. In de toetsmatrijs is het aandeel van de opgaven Spelling (niet-werkwoorden en werkwoorden) groter dan het aandeel van Interpunctie en Grammatica. Dit heeft te maken met de positie van spelling als onderdeel van Taalverzorging, het vierde domein van taal, in het Referentiekader Taal en Rekenen. Taalverzorging is in het referentiekader met name gericht op spelling en in het verlengde daarvan op interpunctie (maar dan in mindere mate). Het domein omvat ook de *begrippenlijst* waarin begrippen en concepten staan die leerlingen en docenten nodig hebben om over taal van gedachten te wisselen. Een aantal van deze begrippen heeft betrekking op grammaticale kennis en is opgenomen in de taken grammatica. De begrippenlijst heeft echter niet dezelfde status die spelling heeft. Om recht te doen aan de indeling die in het Referentiekader gehanteerd is en aan het belang dat aan spelling gehecht wordt, hebben we ervoor gekozen om in de basistoets 25% opgaven spelling extra op te nemen.

Tabel 3.12 De vier domeinen met aantallen opgaven in de taken Taalverzorging

	Basis	aantal opgaven	Verdieping	aantal opgaven	Totaal
Grammatica	taak 1	10 (10)	taak 8	10 (10)	20 (20)
Interpunctie	taak 5	10 (10)	taak 11	10 (10)	20 (20)
Spelling niet-werkwoorden	taak 1	15 (15)	taak 8	10 (10)	25 (25)
Spelling werkwoorden	taak 5	15 (15)	taak 11	10 (10)	25 (25)

In tabel 3.13 staan de percentages en aantallen opgaven 1F en 2F die zijn opgenomen in de taken Taalverzorging. De toewijzing van opgaven aan referentieniveau is gebeurd op basis van de inhoudsomschrijvingen voor het domein Taalverzorging zoals deze zijn beschreven in de publicaties van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (zie de eerder beschreven bronnen).

Tabel 3.13 De aantallen opgaven en percentages 1F en 2F van de referentieniveaus in de taken
Taalverzorging

	1F	2F
Taken Taalverzorging Basistoets	26 (52%)	24 (48%)
Grammatica	5	5
Interpunctie	5	5
Spelling niet-werkwoorden	9	6
Spelling werkwoorden	7	8
Taken Taalverzorging Verdieping	21 (52%)	19 (48%)
Grammatica	6	4
Interpunctie	5	5
Spelling niet-werkwoorden	5	5
Spelling werkwoorden	5	5
Totaal aantal (% gemiddeld)	47 (52%)	43 (48%)

Daarbij zijn de aantallen opgaven uitgesplitst per domein, voor de taken in de basistoets en de taken in de module Verdieping afzonderlijk. Vooraf is als voorwaarde geformuleerd dat voor beide referentieniveaus in de basismodule tenminste 15, maar bij voorkeur 20 opgaven zouden kunnen worden gekenschetst als behorend bij 1F, respectievelijk 2F. Wanneer alleen de basistoets wordt afgenomen, zien we dat aan die voorwaarde wordt voldaan voor beide referentieniveaus. Er zijn 26 1F-opgaven en 24 2F-opgaven. Wanneer ook de verdiepende module wordt afgenomen zijn er in totaal 47 opgaven op 1F-niveau en 43 op 2F.

In het navolgende zullen we de inhoud van de vier domeinen bespreken. Voorafgaand daaraan geven we hier een aantal algemene principes en inhoudelijke criteria die zijn gehanteerd bij het selecteren van de opgaven voor de definitieve toetsversie. Ieder domein van Taalverzorging bevat een groot aantal categorieën (Grammatica 9, Interpunctie 10, Spelling niet-werkwoorden 27 en Spelling werkwoorden 12), zodat het nodig was om dergelijke criteria te formuleren.

Als in de taalmethoden in groep 7 een bepaalde taalverzorgingscategorie werd behandeld, dan wilden wij die categorie in de toets terug laten komen. Verder moest de verdeling van opgaven over categorieën en taken zo gelijkmatig mogelijk zijn. Een categorie moest minimaal één keer voorkomen in de toets, bij voorkeur twee of drie keer en niet meer dan vijf keer.

Het eerste criterium leverde geen probleem op. Het tweede criterium, een zo gelijkmatig mogelijke verdeling van het aantal opgaven per categorie, is waar mogelijk aangehouden. Factoren die van invloed waren op de verdeling zijn onder andere het verschil in aantal categorieën per domein en het verschil in moeilijkheid van bepaalde categorieën. Sommige categorieën vereisen immers een hoger vaardigheidsniveau dan andere.

Er is telkens opnieuw een afweging gemaakt op basis van inhoudelijke en psychometrische informatie, waarbij psychometrische argumenten het soms wonnen van inhoudelijke. In een paar gevallen is daarom de verdeling van opgaven niet helemaal gelijkmatig. De uiteindelijke verdeling van aantallen opgaven per categorie is een zo goed mogelijk compromis tussen eisen van inhoudelijke aard en psychometrische kwaliteit.

De inhoud van het domein Grammatica

De opgaven over grammaticale begrippen gaan zowel over redekundig als over taalkundig ontleden.

We hanteren twee hoofdcategorieën:

- woordbenoeming en benoeming werkwoorden;
- zinsvormen en zinsontleding.

We hebben geprobeerd een zo evenwichtig mogelijke spreiding over de categorieën te hanteren. In tabel 3.14 is aangegeven tot welke categorieën de opgaven Grammatica in de taken Taalverzorging behoren.

Tabel 3.14 *Verdeling van opgaven over grammaticacategorieën per taak in de basistoets en de module Verdieping*

Categorie	Taak 1	Taak 8
Woordbenoeming en benoeming werkwoorden	6 opgaven (6)	7 opgaven (7)
zelfstandig naamwoord	18	13
bijvoeglijk naamwoord	22	17
lidwoord	25	19
infinitief/hele werkwoord	19	14
voltooid deelwoord	23	18
tijd van het werkwoord	17	12 en 20
Zinsvormen en zinsontleding	4 opgaven (4)	3 opgaven (3)
onderwerp	20 en 24	15
persoonsvorm	16	11
(werkwoordelijk) gezegde	21	16

Voor de taken Grammatica gebruiken we verschillende typen meerkeuzeopgaven die we afwisselend aanbieden. We maken gebruik van verschillende opgaventypen om zo de flexibiliteit van de toepassing van de grammaticale kennis te toetsen. Het opgaventype hangt ook af van de te bevragen categorie. De ene categorie leent zich beter voor een bepaald opgaventype dan de andere. In de tabel is te zien dat we de gewenste verdeling van opgaven volledig hebben kunnen realiseren. De nummers in de tabel verwijzen naar de specifieke opgaven in de toetsen.

De inhoud van het domein Interpunctie

We onderscheiden bij Interpunctie twee hoofdcategorieën:

- zinseindetekens: dit zijn leestekens die zinsgrenzen markeren, zoals hoofdletter, punt, vraagteken en uitroepetekens;
- zinsgeleders en citaatmarkeerders: dit zijn leestekens die binnen een zin geplaatst worden als ondersteuning van het lezen en het begrijpen van de inhoud van de zin. Het gaat om komma, dubbele punt en aanhalingstekens.

In tabel 3.15 is te zien tot welke categorieën de opgaven Interpunctie in de taken Taalverzorging behoren.

Tabel 3.15 Verdeling van opgaven over interpunctiecategorieën per taak in de basistoets en de module Verdieping

Categorie	Opgavennummer in taak 5	Opgavennummer in taak 11
Zinseindetekens	5 opgaven (5)	5 opgaven (5)
Hoofdletter als markering zinsgrens	7	4
Punt als markering zinsgrens	6	6
Hoofdletter directe rede	1	9
Vraagtekens	5	5
Uitroeptekens	10	2
Zinsgeleders en markeerders van een citaat	5 opgaven (5)	5 opgaven (5)
Dubbele punt bij opsomming	--	10
Dubbele punt bij conclusie	3	7
Komma bij opsomming	2	--
Komma tussen twee persoonsvormen	8	1
Aanhalingstekens bij opening en sluiting van directe rede	4 en 9	3 en 8

Voor Interpunctie hebben we verschillende typen meerkeuzeopgaven gebruikt die we afwisselend aanbieden. Er is bewust voor gekozen om nergens leestekens of hoofdletters in de opgaven te plaatsen, omdat hiermee mogelijk het antwoord op andere opgaven kan worden weggegeven. De meeste leerlingen weten bijvoorbeeld dat na een punt een hoofdletter geplaatst moet worden. Als we een punt zouden plaatsen in opgaven die de hoofdletter bevragen, zouden de opgaven erg eenvoudig worden. Het was de bedoeling om zowel in de basistoets als in de module Verdieping steeds vijf opgaven op te nemen in de categorieën 'Zinseindetekens' en 'Zinsgeleders en markeerders van een citaat'. Tabel 3.15 laat zien dat we in deze opzet zijn geslaagd.

De inhoud van het domein Spelling niet-werkwoorden

Vanaf groep 6 ligt de nadruk bij het deelgebied spelling op het spellen van complexere en niet-klankzuivere woorden van twee of meer lettergrepen. In de Entreetoets groep 7 zijn vooral deze woorden opgenomen in de taken voor het domein Spelling niet-werkwoorden.

In tabel 3.16 staan de spellingcategorieën die gebruikt zijn in de opgaven Spelling niet-werkwoorden. We hebben dezelfde categorieënindeling aangehouden als in de LVS-toetsen Spelling 3.0 en Taalverzorging voor groep 6, 7 en 8 van het Cito Volgstelsel primair en speciaal onderwijs. Er is getracht een zo evenwichtig mogelijke spreiding over de categorieën heen te hanteren. In elke opgave zijn twee spellingcategorieën opgenomen. Vandaar dat in elke kolom elk opgavennummer tweemaal voorkomt. Achter de opgavennummers staat steeds het aantal gewenste opgaven vermeld. Uit de aantallen opgavennummers is het aantal gerealiseerd opgaven eenvoudig af te leiden. In de basistoets (taak 1) is vrijwel overal het gewenste aantal opgaven in een categorie (variërend tussen 0 en 3) gerealiseerd. Er is nergens méér dan een opgave afgeweken. In de gevallen dat er wel is afgeweken, is daarvoor gecompenseerd in de keuze van de opgave voor de verdiepingmodule (taak 8).

Tabel 3.16 Verdeling van opgaven over spellingcategorieën niet-werkwoorden per taak in de basistoets en de module Verdieping

Cat.	Omschrijving	Opgavennummer in taak 1	Opgavennummer in taak 8
14	woorden met (-)ei(-) of (-)ij(-)	1 (1)	5 (1)
15	woorden eindigend op -d	8 (1)	3 (1)
17	woorden met -au(-), -auw, -ou(-) of -ouw	4 (1)	-- (0)
20	woorden met open lettergreep	7 (1)	6 (1)
21	woorden met gesloten lettergreep	7, 9, 12 en 14 (3)	3 (1)
22	verandering van -f in -v- en -s in -z- bij vervoeging of meervoudsvorming	13 (1)	-- (0)
24	woorden op -ig(e) en -lijk(e)	-- (0)	5 (1)
25	woorden waarin /ie/ geschreven wordt als i	1 (1)	-- (0)
26	woorden waarin /s/ geschreven wordt als c	5 (1)	-- (0)
27	woorden waarin /k/ geschreven wordt als c	-- (0)	8 (1)
28	woorden beginnend met 's of eindigend op 's	-- (0)	7 (1)
30	woorden met -heid of -teit	15 (1)	2 (1)
32	leenwoorden waarin /sj/ geschreven wordt als ch (nieuw)	13 (1)	
33	woorden met -b(-)	2 (1)	
34	woorden met (-)y(-)		10 (1)
35	woorden met een trema	3 (1)	
37	samenstellingen met tussen -e(n)- en tussen -s-	6 (1)	8 en 10 (1)
38	woorden met of zonder een hoofdletter	11 (2)	7 en 9 (1)
39	Franse leenwoorden	5 (1)	6 (1)
40	Engelse leenwoorden	-- (2)	2, 4 en 4 (1)
41	woorden waarin /t/ geschreven wordt als th	2 en 14 (1)	-- (1)
42	woorden met -isch(e)	-- (0)	9 (1)
44	verkleinwoorden -aatje, -eetje, -ootje, -uutje en met de uitgang -nkje	8 en 11 (2)	1 (1)
45	woorden met assimilatieverschijnselen (nieuw)	3, 4 en 9 (2)	-- (1)
46	woorden op -iaal, -ieel, -ueel, -eaal	15 (1)	-- (0)
47	stoffelijke bijvoeglijke naamwoorden	10 en 12 (2)	1 (1)
49	meervouds-n bij zelfstandig gebruikte verwijzing	6 en 10 (2)	-- (1)

Voor de Entreetoets groep 7 hebben we dezelfde nieuwe opgavenform voor de spelling van de niet-werkwoorden gebruikt als voor de toetsen Taalverzorging. Aan de hand van meerkeuzevragen gaan leerlingen op zoek naar de zin waarin de twee dikgedrukte woorden *juist* gespeld zijn.

De opgaven zijn zo samengesteld dat de alternatieven van de meerkeuzevragen betrekking hebben op twee spellingcategorieën. Dat betekent dat in elke opgave over het algemeen meerdere spellingcategorieën zijn opgenomen. Op deze manier wordt een brede dekking van het domein gerealiseerd.

De inhoud van het domein Spelling werkwoorden

In de taken Spelling werkwoorden worden vijf spellingcategorieën onderscheiden, die zijn onderverdeeld in een aantal subcategorieën. In tabel 3.17 volgt een beschrijving van de spellingcategorieën werkwoorden die gebruikt zijn in de Entreetoets groep 7. We hebben dezelfde categorie-indeling aangehouden als in de toetsen Spelling 3.0 en Taalverzorging voor groep 6, 7 en 8 van het Cito Volgsysteem primair en speciaal onderwijs. Ook hier is getracht een zo evenwichtig mogelijke spreiding over de categorieën te hanteren. In elke opgave zijn twee spellingcategorieën opgenomen. Vandaar dat in elke kolom elk opgavennummer tweemaal voorkomt. Ook bij Spelling werkwoorden geldt dat de afwijkingen tussen gewenste en

gerealiseerde aantallen opgaven per spellingcategorie in de regel niet groot zijn. In categorie 1d waren twee opgaven gewenst, waar er geeneen gerealiseerd is; dat is gecompenseerd in de Verdiepende module. In de overige categorieën komen de gewenste en gerealiseerde aantallen opgaven overeen of wijken zij niet meer dan één opgave af. Wanneer er afwijkingen zijn worden deze gecompenseerd via de Verdiepende module.

Tabel 3.17 Verdeling van opgaven over spellingcategorieën werkwoorden per taak in de basistoets en de module Verdieping

Cat.	omschrijving	Vraagnummer in taak 5	Vraagnummer in taak 11
1.a	tijd van nu (ott): -t achter stam van zwak ww dat in o.v.t. de uitgang -de(n) krijgt	14, 17, 18 en 21 (3)	19 (2)
1.b	tijd van nu (ott): wel of geen -t achter een stam op -d	15, 17 en 24 (3)	13, 13 en 17 (3)
1.d	tijd van nu (ott): homofone gevallen: o.t.t.	-- (2)	17 en 19 (1)
2.b	tijd van toen (ovt): verdubbeling d of t bij zwak ww met stam op -d of -t	11, 11, 16 en 19 (3)	18 (2)
2.c	tijd van toen (ovt): geen -t bij sterk ww dat in 2e en 3e persoon eindigt op -d	16 en 23 (2)	14 (1)
2.d	tijd van toen (ovt): uitgang -sde(n) of -fde(n) bij zwak ww met stam op -z of -v	12, 23 en 25 (3)	14 en 20 (2)
2.e	tijd van toen (ovt): geen verdubbeling medeklinker bij sterk werkwoord	12 en 13 (3)	11, 15 en 20 (2)
3.a	voltooid deelwoord: keuze voor eind-d of eind-t bij zwakke werkwoorden met een stam die niet eindigt op -d of -t of -v en -z	13, 20 en 22 (3)	12, 15 en 16 (2)
3.b	voltooid deelwoord: homofone gevallen: volt.dw.	21, 22 en 25 (2)	-- (1)
3.c	zwakke werkwoorden met stam op v en z of d en t	20 (1)	11 (1)
4.a	wel of geen -n aan het eind; d of t; woorden met d en t onvoltooid deelwoord bijvoeglijk gebruikt	15, 18, 19 en 24 (3)	16 (2)
5.a	infinitief: homofone gevallen	14 (2)	12 en 18 (1)

We hebben voor het domein Spelling werkwoorden de nieuwe opgavenvorm gebruikt, waarbij leerlingen aan de hand van meerkeuzeopgaven op zoek gaan naar de zin met de twee *juist* gespelde dikgedrukte woorden. In elke opgave zijn over het algemeen meerdere spellingcategorieën opgenomen, waarmee een brede dekking van het deelgebied wordt gerealiseerd.

3.2.2 De module Verbreding

3.2.2.1 Luisteren

Begripsomschrijving

Luisteren is het proces waarbij de luisteraar betekenis toekent aan gesproken taal. Een luisteraar reconstrueert betekenis: hij zet reeksen klanken waarin de bedoeling van de spreker verpakt is, om in inhouden. Daarna probeert hij met die inhouden 'opnieuw' een betekenis samen te stellen.

Die samenstelling is geslaagd als de 'nieuw' gereconstrueerde betekenis overeenkomt met de betekenis die de spreker wil overbrengen.

Luisteren is dus een actief proces, maar is ook een interactief proces, waarbij interactie plaatsvindt tussen de luisteraar, de tekst en de context. Wanneer de luisteraar betekenis toekent aan gesproken taal gebeurt dat altijd in interactie met de tekst. Het gedrag van de luisteraar wordt bepaald door datgene wat de spreker ter sprake brengt, maar ook door de inbreng van zijn 'eigen' kennis en zijn eerdere (luister)ervaringen. Daarnaast is ook het doel dat de luisteraar voor ogen heeft bepalend voor zijn gedrag. De ontwikkeling van het luisteren kan gezien worden als een cyclisch, concentrisch proces waarin de luisteraar herhaaldelijk dezelfde ontwikkelings- en leerprocessen doorloopt, maar op een steeds hoger niveau.

In het toekennen van betekenis aan gesproken taal spelen zowel tekst- als kennisgestuurde verwerkingsprocessen een belangrijke rol. Bij tekstgestuurde verwerking staat de inhoud van de tekst centraal en verwerkt de luisteraar de informatie die de spreker expliciet ter sprake brengt. Dit duiden we aan als het begrijpen van gesproken teksten. Om tot begrip van de tekst te komen, maakt de luisteraar gebruik van de inhoud (de betekenis van woorden, woordgroepen, zinnen, langere tekstpassages en hun onderlinge betekenisrelaties), van expliciete relaties tussen elementen in een uiting of tekst (woord- en zinsvolgorde, verwijzingen en talige structuurmarkeerders) en van de expliciete structuur van een tekst. Kennisgestuurde verwerking gaat verder: om tot begrip van de tekst te komen, zet de luisteraar ook 'eigen' kennis in, waaronder zijn kennis van de wereld, zijn kennis over taal en zijn kennis over taalgebruikssituaties. De spreker veronderstelt dat bepaalde kennis bij de luisteraar bekend is en zal deze niet altijd expliciet verwoorden. Het is aan de luisteraar om deze kennis te activeren en aan te vullen met eigen kennis. Dit duiden we aan als het interpreteren van gesproken teksten. Het onderkennen en afleiden van impliciete informatie in een tekst en het maken van inferenties zijn hierbij belangrijk. Tussen tekst- en kennisgestuurde verwerking vindt een continue wisselwerking plaats. Pas wanneer ze in samenhang en gelijktijdig ingezet worden, is er sprake van werkelijk en diepgaand tekstbegrip.

Inhoudsverantwoording

De inhoud van de teksten en de opgaven in de taken Luisteren is gebaseerd op verschillende bronnen. Ze vormen de basis voor de verschillende inhoudsaspecten en opgaventypen die in de taken Luisteren zijn opgenomen. Het gaat om de volgende bronnen:

- De tussendoelen Mondelinge communicatie (Verhoeven, Biemond & Litjens, 2007)
- De kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006);
- De taalmethoden die in het basisonderwijs in groep 7 gebruikt worden;
- De referentieniveaus Taal voor het basisonderwijs (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009a)
- Het document 'Leerstoflijnen mondelinge taalvaardigheid beschreven. Uitwerking van het referentiekader Nederlandse taal voor het basisonderwijs in mondelinge taalvaardigheid op de basisschool (van der Beek, Hoogeveen & Prenger, 2015);
- Het document 'Over de drempels met Taal en Rekenen' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008b);
- Het document 'Over de drempels met taal' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008a);
- Het document 'Een nadere beschouwing' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009b);
- Balans van de luistervaardigheid in het basis- en speciaal basisonderwijs 4 (Krom, van Berkel, van der Schoot, Sijstra, Hemker & Marsman, 2011).

De inhoud van het onderdeel Luisteren

De teksten

Tekstsoorten

Bij het onderscheid in tekstsoorten gaat het om functies of doelen van teksten: met een betogende tekst bijvoorbeeld wil de spreker de luisteraar ergens van overtuigen, en met een verhalende tekst wil de spreker de luisteraar vermaken of boeien. In het onderwijs krijgen de leerlingen te maken met verschillende soorten luisterteksten. Er kan onderscheid worden gemaakt tussen zakelijke teksten en fictionele teksten, die weer zijn onderverdeeld in verschillende tekstsoorten. De volgende tekstsoorten komen voor:

- Informatieve teksten (zakelijke teksten): De spreker geeft feitelijke informatie over de werkelijkheid. Soms voegt hij er ter verduidelijking ook zijn eigen opvattingen aan toe.
- Instructieve teksten (zakelijke teksten): De spreker wil het gedrag, het handelen van de luisteraar richten en sturen. De spreker zorgt ervoor dat de luisteraar de instructie goed kan uitvoeren.
- Betogende teksten (zakelijke teksten): De spreker wil vooral het denken van de luisteraar beïnvloeden. Hij roept de luisteraar op om iets te doen of juist iets te laten via een oproep of aankondiging.
- Verhalende teksten (fictionele teksten): De spreker beschrijft een verbeelde werkelijkheid in verhalen. Hij bedenkt bijvoorbeeld een spannend of grappig verhaal.
- Poëzieteksten (fictionele teksten): De spreker beschrijft een verbeelde werkelijkheid in poëzie.

In tabel 3.18 staat de verdeling van de teksten van het onderdeel Luisteren naar tekstsoort. Tussen haakjes staat steeds vermeld welke aantallen teksten we vooraf wensten te realiseren. De gewenste en gerealiseerde aantallen teksten komen overeen. Ze weerspiegelen de situatie in de bovenbouw dat leerlingen van het basisonderwijs via radio, televisie en internet overwegend zakelijke teksten krijgen aangeboden. In overeenstemming met deze waarneming zijn drie zakelijke en twee fictionele teksten in het onderdeel Luisteren opgenomen.

Tabel 3.18 Percentages en aantallen tekstsoorten in de taken in het onderdeel Luisteren

Tekstsoort	Taak 14	Taak 15	Totaal
zakelijke teksten	50% 1 (1)	67% 2 (2)	60% 3 (3)
fictionele teksten	50% 1 (1)	33% 1 (1)	40% 2 (2)

De teksten zijn inhoudelijk en qua moeilijkheid afgestemd op de leerlingen in groep 7 van het basisonderwijs. Ze sluiten aan bij de leefwereld van de leerlingen en zijn relatief kort: de luisterduur per tekst bedraagt maximaal 5 minuten. De leerlingen luisteren naar audio, of, met andere woorden, naar luistermateriaal dat niet ondersteund wordt door beeld.

Tekstgenres

Teksten kunnen, naast een indeling in soorten, ook op grond van de vorm van de tekst ingedeeld worden. Het gaat dan om het tekstgenre. Tekstgenres waar basisschoolleerlingen zoal mee te maken krijgen zijn: gesprek, nieuwsbericht, interview, toespraak, verhaal, mop, gedicht en lied.

De percentages tekstgenres die in de taken aan bod komen staan vermeld in tabel 3.19; de gerealiseerde aantallen komen overal overeen met de vooraf gewenste aantallen.

Tabel 3.19 Percentages en aantallen tekstgenres in de taken in het onderdeel Luisteren

Tekstgenre	Taak 14	Taak 15	Totaal
Nieuwsbericht		33% 1 (1)	20% 1 (1)
Interview/gesprek	50% 1 (1)		20% 1 (1)
Verhaal	50% 1 (1)		20% 1 (1)
Toespraak/betoog		33% 1 (1)	20% 1 (1)
Lied		33% 1 (1)	20% 1 (1)

Elke combinatie van tekstsoort en tekstgenre heeft specifieke kenmerken wat betreft opbouw, stijl, register, doel, publiek, taalgebruik, conventies, mate van formaliteit en presentatie.

Vaak hebben tekstgenre en tekstsoort een een-op-een-relatie: een nieuwsbericht informeert altijd; in een instructie wordt het handelen van de luisteraar gestuurd. Soms is de relatie echter complexer: in een verhaal kan informatie gegeven worden, maar er kan ook iets in betoogd worden en er kunnen aanwijzingen in gegeven worden.

De inhoud van het onderdeel Luisteren

De opgaven in het onderdeel Luisteren zijn meerkeuzeopgaven. De vragen worden voorgelezen en de antwoordalternatieven staan in het opgavenboekje. De opgaven hebben betrekking op een woord(groep) of op een of enkele zinnen. Ze kunnen ook betrekking hebben op grotere tekstdelen of op de hele tekst.

We hebben de opgaven ingedeeld op basis van de verwerkingsprocessen 'begrijpen van gesproken teksten' en 'interpreteren van gesproken teksten' én op basis van de inhoud die in het luisteronderwijs aan bod komen. Deze inhoud hebben we vertaald in een aantal inhoudsaspecten die we gerelateerd hebben aan de verwerkingsprocessen. Op die manier zijn we gekomen tot een aantal verschillende opgaventypen. Elk opgaventype representeert een of meer inhoudsaspecten.

Onderstaande indelingen laten voor elk van de beide verwerkingsprocessen zien welke inhoudsaspecten aan welk verwerkingsproces gerelateerd zijn. Daarbij willen we benadrukken dat de verwerkingsprocessen in werkelijkheid niet altijd even duidelijk van elkaar te scheiden zijn.

De indeling in opgaven bij begrijpen van gesproken teksten is als volgt:

- Opgaven die vragen naar de betekenis van een woord of woordgroep die expliciet door de spreker vermeld wordt.
- Opgaven die vragen naar specifieke inhoudselementen die expliciet in de tekst aan de orde gesteld worden. Dit zijn bijvoorbeeld (hoofd)personen, feiten en meningen, voorwerpen, aantallen, een plaats van handeling of tijdsperioden.
- Opgaven die vragen naar eenvoudige expliciete verbanden op basis van inhoudelijke en structurele elementen. Voorbeelden daarvan zijn vergelijkingen, tegenstellingen, generalisaties en voorbeelden, vraag en antwoord. Ook verwijzingen en verbanden tussen kleine stukjes informatie die expliciet verwoord worden – terwijl het verband zelf niet geëxpliciteerd is – en expliciete verbanden die de spreker legt tussen gebeurtenissen, personen of plaatsen zijn hier voorbeelden van.
- Opgaven die vragen naar complexe expliciete verbanden op basis van inhoudelijke en structurele elementen over grotere tekstdelen heen. Dit zijn bijvoorbeeld: reden en verklaring, oorzaak en gevolg, middel en doel, deel-/geheelrelaties, conclusie en argumenten, generalisaties en voorbeelden of hoofd- en bijzaken. Maar ook opgaven die vragen naar de chronologie van gebeurtenissen of naar opeenvolgende stappen vallen hieronder.

De indeling in opgaven bij interpreteren van gesproken teksten is als volgt:

- Opgaven die vragen naar het afleiden van de betekenis van een woord of woordgroep.
- Opgaven die vragen naar het afleiden van informatie uit de tekst, waarbij de leerling zijn voorkennis moet inzetten, naast de informatie die de spreker geeft. Het gaat dan om opgaven waarbij ontbrekende informatie moet worden aangevuld, waarbij moet worden geanticipeerd of waarbij naar de bedoeling,

gevoelens of mening van de spreker gevraagd wordt. Maar ook opgaven die vragen naar de functionele betekenis van tekstdelen vallen hieronder.

- Opgaven die vragen naar de globale inhoud van de tekst of een tekstdeel waarbij expliciete en/of impliciete inhoudelijke en/of structurele elementen verspreid over de tekst of over grotere tekstdelen, moeten worden verbonden. Voorbeelden hiervan zijn opgaven over onderwerp, thema, hoofdlijnen, hoofdgedachte, hoofdpersoon en doel en publiek van de tekst. Ook opgaven waarbij de leerlingen informatie in de tekst moeten vergelijken en/of doorzien of waarbij ze de inhoud van de tekst of een tekstdeel moeten samenvatten of de opbouw van een tekst moeten doorzien, zijn hier voorbeelden van.
- Opgaven die vragen naar de manier van spreken, bijvoorbeeld naar: klemtoon, intonatie, volume, tempo, toon, accent, register, sociale en culturele conventies en waarbij een verband gelegd moet worden tussen tekstuele informatie en kennis van het taalsysteem.

Bij het samenstellen van de taken Luisteren zijn we ervan uitgegaan dat de verwerkingsprocessen 'begrijpen van gesproken teksten' (kortweg Begrijpen) en 'interpreteren van gesproken teksten' (kortweg Interpreteren) beide een belangrijke rol in het luisterproces vervullen. Echter, het interpreteren van gesproken teksten neemt in de bovenbouw van het basisonderwijs een steeds belangrijkere plaats in, waarbij het verwerken van expliciete informatie ('begrijpen') steeds meer op de achtergrond raakt. In tabel 3.20 is het percentage gerealiseerde opgaven 'Begrijpen' en het percentage opgaven 'Interpreteren' in de taken Luisteren af te lezen. De verhouding Begrijpen en Interpreteren die we in de hoogste groepen van het basisonderwijs nastreven is 25% Begrijpen - 75% Interpreteren voor leerlingen in groep 7 en 20% Begrijpen - 80% Interpreteren voor leerlingen in groep 8. Met de gerealiseerde verhouding van 20% Begrijpen - 80% Interpreteren is deze verhouding nog iets schever dan we vooraf hadden beargumenteerd en komt deze in de buurt van wat we opportuun achten voor het hoogste leerjaar.

Tabel 3.20 Percentages en aantallen opgaven Begrijpen en Interpreteren in de taken Luisteren

	Taak 14	Taak 15	Totaal
Begrijpen	33% 5 (5)	7% 1 (2)	20% 6 (7)
Interpreteren	67% 10 (10)	93% 14 (13)	80% 24 (23)

3.2.2.2 Schrijven

Begripsomschrijving

Schrijven is een complex proces waarbij de schrijver, in interactie met een ruwe schets van zijn schrijfplan of aandachtspunten, betekenis toekent aan zijn tekst. Hij construeert betekenis op basis van gegeven informatie en op basis van zijn eigen kennis. De schrijver organiseert zijn ideeën en bepaalt zijn schrijfdoel. Ook bedenkt hij vooraf voor wie hij schrijft. Deze activiteiten bepalen wat voor soort tekst de schrijver gaat schrijven. Het formuleren van de tekst en het herlezen daarvan tijdens het schrijven (monitoring), het evalueren en het reviseren van de geschreven tekst én het coördineren van al die activiteiten in het schrijfproces, maken schrijven tot een complexe vaardigheid.

Inhoudsverantwoording

De inhoud van de teksten en opgaven in de taken Schrijven is gebaseerd op verschillende bronnen, waaronder:

- De tussendoelen gevorderde geletterdheid voor de middenbouw. Leerlijnen voor groep 4 tot en met groep 8 (Aarnoutse & Verhoeven, 2003);

- De kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006);
- De methoden schrijven die in het basisonderwijs gebruikt worden;
- De referentieniveaus Taal voor het basisonderwijs (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009a);
- Het document 'Over drempels met Taal en Rekenen' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2008b);
- Het document 'Een nadere beschouwing' (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009b);
- Het document 'Leerstoflijnen schrijven beschreven' van SLO (Van Gelderen, 2010);
- Balans van de schrijfvaardigheid in het basis- en speciaal basisonderwijs 2 (Kuhlemeier, van Til, Hemker, de Klijn & Feenstra, 2013).

De inhoud van het onderdeel Schrijven

Het reviseren van geschreven teksten

Schrijven komt in de Entreeoets groep 7 niet zo breed aan de orde als hierboven in de begripsomschrijving is geschetst. Met de opgaven Schrijven meten we het schrijven indirect. De leerling schrijft niet zelf een tekst aan de hand van een opdracht, maar hij laat zien wat hij weet over het schrijven van teksten door opgaven te beantwoorden over die tekst. We richten ons binnen de indirecte meting alleen op de laatste fase van het schrijfproces: het reviseren van een tekst. De tekst die de leerling voor zich krijgt is een onvolledig of onaf product, een schrijfwerk dat zogenaamd door een leerling uit groep 7 is gemaakt. De voorgelegde tekst verkeert nog in een soort conceptfase. De leerling beantwoordt opgaven die te maken hebben met het reviseren van de tekst én vervolmaakt de tekst als het ware tot een volledig correcte tekst. Hij heeft dan in feite 'zijn eigen tekst' óf de tekst van een medeleerling gereviseerd.

Van werkelijke en diepgaande revisie van een tekst is pas sprake wanneer informatie uit de tekst verbonden wordt aan de 'eigen' kennis van de lezer. Tussen de lezer en de tekst vindt interactie plaats. Een lezer benadert een 'ruwe' versie van een schrijfproduct niet blanco, maar zet bij de verwerking van de tekst ook zijn eigen kennis in, waaronder zijn kennis van de wereld en kennis van taal, tekstgenres en tekstenkenmerken. Dit samenspel van verwerkingsprocessen is aan de orde bij het reviseren van teksten. Door interactie van de leerling met de tekst, komt, na beantwoording van de bijbehorende opgaven, de definitieve versie van de tekst tot stand. De teksten in het onderdeel Schrijven veranderen dus niet, maar de leerling heeft met zijn antwoorden aangegeven wat er volgens hem zou moeten worden aangepast om te komen tot een goed schrijfproduct.

In het onderdeel Schrijven gaat het er om dat de leerling de tekst leest en op fouten beoordeelt. De leerling moet in staat zijn om teksten te lezen als een kritische lezer, als een soort eindredacteur van een tekst voor een jeugdwebsite of nieuwsbrief. Hij gaat namelijk de tekst beoordelen op fouten ten behoeve van revisie. Het kunnen beoordelen van een tekst vraagt een bepaalde mate van reflectie op de aangeboden tekst. De leerling wikt en weegt wat hij ervan vindt. Vervolgens, of misschien wel tegelijkertijd, moet hij begrijpen wat samenhang in een tekst is en hoe samenhang kan worden aangebracht. De leerling moet woordgebruik en idioom begrijpen en kunnen zien wat de mogelijkheden voor variatie of verbetering zijn. Goed reviseren kan niet zonder het begrijpen en toepassen van de regels van spelling, interpunctie en grammatica in een tekst; ook die aspecten moet een leerling in ogenschouw nemen. Ook moet de leerling beslissingen nemen over het schrappen of toevoegen van informatie. Hij moet schrappen als hij vindt dat de schrijver te veel uitweidt of afdwaalt in zijn betoog. Aan de andere kant moet hij besluiten dat informatie toegevoegd moet worden aan de voorgelegde tekst, als er essentiële informatie ontbreekt (bijvoorbeeld een adres, of een uitleg). De leerling beoordeelt de inhoud én vaak ook de structuur van de tekst. Als het om passend taalgebruik gaat, moet een leerling kunnen begrijpen wat het verschil is in stijl: tussen formeel en informeel schrijven. Hij beoordeelt bijvoorbeeld of een deel van de tekst niet te populair of juist te deftig geschreven is en daarom niet past bij de rest van de tekst. Verder moet de leerling ook het register (het behoren tot een specifieke context) kunnen begrijpen en beoordelen op de vraag of het passend gebruikt is. Tot slot moet een leerling aandacht schenken aan de leesbaarheid van de aangeboden tekst en aan de publieks-

gerichtheid. De leerling beoordeelt of de lay-out visueel aantrekkelijk is en uitnodigt tot lezen, en of de tekst inhoudelijk en visueel op het juiste lezerspubliek is gericht. Op alle genoemde aspecten van het schrijven wordt een beroep gedaan in het onderdeel Schrijven.

De teksten

Tekstsoorten

In het onderwijs krijgen de leerlingen te maken met verschillende soorten teksten. Deze worden onderscheiden in zakelijke en fictionele teksten, die zijn onderverdeeld in een aantal tekstsoorten:

- Informatieve teksten (zakelijke teksten): De schrijver geeft feitelijke informatie over de werkelijkheid. Soms voegt hij er ter verduidelijking ook zijn eigen opvattingen aan toe.
- Instructieve teksten (zakelijke teksten): De schrijver wil het gedrag, het handelen van de lezer richten en sturen. De schrijver zorgt ervoor dat de lezer de instructie goed kan uitvoeren.
- Betogende teksten (zakelijke teksten): De schrijver wil vooral het denken van de lezer beïnvloeden. Hij roept de lezer op om iets te doen of juist iets te laten via een oproep, aankondiging, hartenkreet of muurkrant.
- Verhalende teksten (fictionele teksten): De schrijver beschrijft een verbeelde werkelijkheid in verhalen. Hij bedenkt bijvoorbeeld een spannend of grappig verhaal.
- Poëzieteksten (fictionele teksten): De schrijver beschrijft een verbeelde werkelijkheid in poëzie.

De tekstsoorten die in de taken Schrijven vertegenwoordigd zijn, zijn zakelijke teksten die zogenaamd door leerlingen in de bovenbouw van het basisonderwijs geschreven zijn en die aansluiten bij de belevingswereld van de leerlingen uit groep 7. Alle vier de teksten zijn zakelijk van karakter. Twee ervan zijn betogend, één tekst is informatief en één instructief/informatief. Het was gewenst om ook een fictionele tekst, een spannend verhaal bijvoorbeeld, op te nemen. In het basisonderwijs moeten leerlingen immers verschillende soorten teksten leren schrijven en moeten zij ook verschillende soorten teksten ter revisie aangeboden krijgen. In de opgaven voor de proefvoetsing bleek echter geen geschikt verhaal voorhanden te zijn. Deze beperking in de tekstsoorten heeft consequenties voor de tekstgenres die mogelijk zijn.

Tekstgenres

Bij het onderscheid in tekstsoorten gaat het om de functies of doelen van teksten. Schrijfproducten kunnen ook op grond van de vorm van de tekst ingedeeld worden; het gaat dan om het tekstgenre. Een tekst kan bijvoorbeeld een brief, maar ook een verhaal of een muurkrant zijn. Tekstgenres waar basisschoolleerlingen zoal mee te maken krijgen, zijn:

- persoonlijke of zakelijke correspondentie, zoals een kaartje, e-mail of brief, een uitnodiging of oproep;
- formulieren, berichten, advertenties en aantekeningen, zoals een dagboek- of logboekfragment, een poster of muurkrant;
- verslagen, werkstukken, samenvattingen, instructies en artikelen;
- verhalen of gedichten waarbij het gaat om vrij schrijven en creatief schrijven.

De teksten in de taken Schrijven behoren daarom ook tot verschillende tekstgenres. In tabel 3.21 wordt weergegeven welke tekstgenres aan bod komen in de taken Schrijven. Tussen haakjes is aangegeven welke en hoeveel teksten van een bepaald genre gewenst waren.

Tabel 3.21 Percentages en aantallen teksten naar tekstgenre in de taken in het onderdeel Schrijven

Tekstgenre	Taak 16	Taak 17	Totaal
Uitleg	50% 1 (1)	----	25% 1 (1)
Brief	50% 1 (1)	----	25% 1 (1)
Verslag of werkstuk	---- (0)	100% (2)	50% 2 (2)
Verhaal of gedicht	---- (1)	---- (0)	---- (1)

Vaak hebben tekstgenre en tekstsoort een een-op-een relatie: een verslag informeert altijd; in een instructie wordt het handelen van de lezer gestuurd. Soms is de relatie echter complexer. In een brief kan geïnformeerd worden, maar er kan ook iets in betoogd worden en er kunnen allerlei mogelijke aanwijzingen of instructies in gegeven worden. Het was de bedoeling om ook een verhaal of gedicht in de taken op te nemen, maar door het ontbreken van een geschikte fictionele tekst was dat niet mogelijk. Wél kon er een verslag worden opgenomen van een grappige persoonlijke gebeurtenis. Daarnaast was het opportuun geweest om ook een tekst in de vorm van een e-mail op te nemen, dit vanwege de hoge frequentie van voorkomen van e-mails als correspondentievorm in vergelijking met brieven. Een geschikte tekst in de vorm van een e-mail was echter niet voorhanden bij de proefvoetsing, terwijl een (zakelijke) brief wél geschikt bleek. Al met al zijn de vier teksten in het onderdeel Schrijven wat eenzijdiger qua tekstsoort en -genre dan de bedoeling was. Dit komt deels ook doordat uiteindelijk niet vijf maar vier teksten werden opgenomen in de taken.

De opgaven

Het onderdeel Schrijven bevat teksten waarin fouten staan die leerlingen in groep 7 tijdens het schrijven daadwerkelijk maken. Bij het maken van de opgaven verbeteren de leerlingen als het ware de fouten in de teksten en beoordelen ze de passendheid van tekstdelen. Er zijn verschillende soorten fouten die leerlingen kunnen maken. Deze betreffen aspecten van het schrijven, namelijk:

- *Samenhang*: Hieronder vallen fouten rondom voegwoorden, verwijzwoorden, informatie ordenen, alineagebruik en titels.
- *Woordgebruik en woordenschat*: Hieronder vallen fouten met betrekking tot het dubbel laten voorkomen van woorden, tegen het plaatsen van het juiste woord op de juiste plek in de zin; fouten tegen variatie in woordkeus (het ontbreken van synoniemen), fouten tegen het correcte gebruik van spreekwoorden, gezegden en beeldspraak, fouten met betrekking tot overbodige uitweidingen in een betoog, fouten tegen het compact schrijven en fouten die te maken hebben met het ontbreken van woorden.
- *Spelling, interpunctie en grammatica*: Hieronder vallen fouten met betrekking tot zinsbouw, werkwoordcongruentie, interpunctie en voorzetsels.
- *Inhoud*: Hieronder vallen fouten die te maken hebben met het toevoegen van relevante informatie voor beter begrip en het weglaten van overbodige informatie. Het toevoegen van absoluut noodzakelijke informatie om de communicatie te doen slagen valt ook onder het aspect inhoud.
- *Passendheid*: Hieronder vallen fouten die betrekking hebben op stijl en register. Bij dit aspect is steeds de vraag of het taalgebruik past bij de situatie en bij de beoogde lezer.
- *Publieksgerichtheid/Leesbaarheid*: Hieronder vallen fouten waarbij het gaat om het afstemmen op of het overtuigen van de lezer. Fouten die te maken hebben met het kennen van en inzicht hebben in (brief)conventies, de logische opbouw en structuur van een tekst (inleiding, middenstuk en conclusie), de visuele structuur van een tekst (lay-out, tekst op papier of op beeldscherm). Ook het herkennen van fouten tegen het leggen van nadruk door opvallende herhaling of door woordplaatsing aan het begin of juist het einde van de zin, behoort tot dit aspect.

Alle opgaven in de taken Schrijven zijn meerkeuzeopgaven: een vraag die gesteld wordt naar aanleiding van een gefingeerde fout in een tekst of in een deel van een tekst (een of meer zinnen, een of meer alinea's, enzovoort). Zo'n fout kan ook het ontbreken van informatie in de tekst betreffen.

In tabel 3.22 wordt per aspect aangegeven welke opgaven in de taken een beroep doen op het genoemde aspect.

Tabel 3.22 *Verdeling van opgaven over aspecten van het schrijven in de taken van het onderdeel Schrijven*

Aspecten van het schrijven	Taak 16	Taak 17	Aantal opgaven	
			Gewenst	Gerealiseerd
Samenhang	15	2, 8	5	3
Woordgebruik, woordenschat		4, 9, 10, 13	5	4
Spelling, interpunctie, grammatica	1, 2, 4, 6, 12, 14, 16	1, 3, 11, 12	7	11
Inhoud	3, 5, 9, 13	6, 7	4	6
Passendheid	7		3	1
Publieksgerichtheid, leesbaarheid	8, 10, 11	5, 14	6	5

Bij toetsen waarbij niet alleen keuzes moeten worden gemaakt op het niveau van de opgaven, maar ook op het niveau van de teksten, is het dikwijls erg lastig om de gewenste verdeling van de opgaven over aspecten van het schrijven te realiseren bij de gegeven combinatie van geschikte teksten en (ook psychometrisch) geschikte opgaven. Ook al zijn er (veel) meer teksten en opgaven beschikbaar bij de proeftoetsing dan er uiteindelijk nodig zijn, toch is het moeilijk om een bepaalde gewenste verdeling exact te realiseren. Uit recente ontwikkelingen in het schrijfonderwijs en uit het onderzoek naar schrijven blijkt dat het kunnen aanbrengen van samenhang een belangrijk kwaliteitsaspect is van een goede schrijver. Het is niet geheel gelukt om de gewenste weging van samenhang binnen de toetsen aan te brengen (3 in plaats van 5 opgaven representeren dit aspect). Ook het gebruik van de juiste woorden en variatie in woordenschat, een ander kenmerk van een goede schrijver, is niet helemaal tot zijn recht gekomen (4 in plaats van 5). Omgekeerd hadden de aspecten spelling, interpunctie en grammatica wat minder aandacht kunnen krijgen (7 in plaats van 11), temeer, omdat die aspecten ook al bij Taalverzorging aan bod komen. Ook het aspect inhoud zou wat minder aandacht hebben kunnen krijgen, omdat het belang van inhoudelijke juistheid of het inhoudelijk genereren van ideeën in een toets gericht op revisie, minder groot is. De aspecten passendheid en publieksgerichtheid of leesbaarheid hadden wat meer aandacht verdiend dan zij nu in de toetsen krijgen. Niettemin kunnen we constateren dat alle aspecten van het schrijven door de opgaven worden gedekt in een frequentie die een aardige benadering vormt van wat we ons vooraf hadden gewenst.

3.2.2.3 Woordenschat

Begripsomschrijving

Onder woordenschat verstaan we de verzameling labels waarover taalgebruikers beschikken voor het begrijpen en gebruiken van taal. Elk woord heeft een woordvorm en een woordbetekenis. De woordvorm is waarneembaar, in tegenstelling tot de woordbetekenis, die is opgeslagen in het hoofd van de taalgebruiker. Woorden en hun betekenis(sen) zijn ingebed in een begrippennetwerk, het zogeheten mentale lexicon. Als de woordenschat zich ontwikkelt, vindt een uitbreiding van dit begrippennetwerk plaats: er komen nieuwe begrippen bij én er worden steeds meer relaties gelegd tussen al aanwezige begrippen.

Woorden en hun betekenis(sen) worden stapsgewijs begrepen, ingepast en toegepast, totdat ze geautomatiseerd zijn. Hoe vaker kinderen ermee in aanraking komen, hoe duidelijker en preciezer ze de betekenis ervan kunnen vaststellen. Ook kunnen ze gemakkelijker nieuwe kennis verwerven, omdat ze deze kunnen koppelen aan bekende woorden en aan al aanwezige kennis. Zowel de breedte als de diepte van de woordenschat spelen daarbij een belangrijke rol. Bij een brede woordenschat staat de omvang van de woordenschat centraal ofwel het aantal woorden dat wordt beheerst. Het gaat dan om kwantitatieve aspecten van de woordenschat. Bij een diepe woordenschat spelen de kwalitatieve aspecten van de

woordenschat een rol en gaat het om de vraag hoe goed woorden en hun betekenisaspecten in relatie tot andere woorden worden gekend.

De breedte en de diepte van de woordenschat overlappen elkaar voor een groot deel: leerlingen die meer woorden kennen, kennen deze woorden vaak gedetailleerder en dieper, ze hebben immers meer woorden tot hun beschikking.

Inhoudsverantwoording

De inhouden die in de taken Woordenschat aan bod komen, zijn gebaseerd op verschillende bronnen, waaronder:

- De kerndoelen primair onderwijs van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006);
- De woordfrequentielijst *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen.* (Schrooten & Vermeer, 1994);
- Het boek *Taal en didactiek. Woordenschat.* Huizenga (2005), een standaardwerk op het gebied van de didactiek van woordenschat.
- Diverse taalmethoden die in de bovenbouw van het basisonderwijs gebruikt worden.

De inhoud van het onderdeel Woordenschat

Kwantitatieve en kwalitatieve aspecten van de woordenschat

In de taken Woordenschat komen inhouden aan bod die refereren aan de kwantitatieve aspecten van de woordenschat. Het toekennen van betekenis is daarin belangrijk. Het gaat daarbij om woorden met eenzelfde betekenis, om definities en beschrijvingen én om belangrijke betekeniskenmerken van woorden. In de taken Woordenschat komen ook inhouden aan bod die refereren aan de kwalitatieve aspecten van de woordenschat. Het gaat dan met name om betekenisrelaties: tegenstellingen, betekenisvelden van woorden, gezamenlijke woordkenmerken en vergelijkingen tussen woorden.

Met uitzondering van de belangrijke betekeniskenmerken van woorden zijn alle genoemde inhouden in de taken Woordenschat opgenomen. We hebben opgaven waarbij het gaat om betekenis-toekenning ondergebracht in de categorie Betekenis en opgaven die vragen naar de relaties tussen woorden in de categorie Betekenisrelaties. Deze indeling sluit aan bij het onderscheid naar de kwantitatieve en kwalitatieve aspecten van de woordenschat, zoals we dat hebben aangetroffen in de literatuur en in woordenschatoefeningen in verschillende taalmethoden voor het basisonderwijs.

Tabel 3.23 Percentages en aantallen opgaven Betekenis en Betekenisrelaties in de taken in het onderdeel Woordenschat

Aard van de opgave	Taak 18	Taak 19	Totaal
Betekenis	55% 11 (10)	65% 13 (10)	60% 24 (20)
Betekenisrelaties	45% 9 (10)	35% 7 (10)	40% 16 (20)

In tabel 3.23 staan de percentages en aantallen opgaven Betekenis en Betekenisrelaties die in de taken zijn opgenomen. Tussen haakjes staan de vooraf gewenste aantallen aangegeven. Het streven was om evenveel opgaven te realiseren met het accent op betekenis als opgaven met het accent op betekenisrelaties. Psychometrisch gezien is zo'n exact gelijke verdeling niet noodzakelijk (doorslaggevend is de schaalbaarheid van een opgave), maar er is ook geen reden om een andere verhouding na te streven: beide aspecten dienen gerepresenteerd te worden om recht te doen aan een volledige omschrijving van het begrip woordenschat, ze nemen een (even) belangrijke plaats in en bovendien zijn taken die uit meerdere soorten opgaven bestaan voor leerlingen aantrekkelijker en motiverender om te maken. Uiteindelijk wijkt de keuze uit de beschikbare opgaven bij de proeftoetsing (i.e. de verzameling psychometrisch geschikte

opgaven) enigszins af van de genoemde 50:50 verhouding. Dit is mede veroorzaakt door het feit dat in de beide taken verschillende subtypen van opgaven (zie verderop) geclusterd moesten worden en we ernaar streefden in beide taken dezelfde gemiddelde moeilijkheidsgraad te realiseren. Uiteindelijk valt 60% van de opgaven in de categorie Betekenis en 40% in de categorie Betekenisrelaties.

De stimuluswoorden in de taken Woordenschat

Tijdens de opgavenconstructie stond voorop dat de opgaven in de taken Woordenschat stimuluswoorden zouden moeten bevatten waarmee de leerlingen in groep 7 van het basisonderwijs in aanraking komen. Daartoe hebben we gebruikgemaakt van de woordfrequentielijst *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen* van Schrooten en Vermeer (1994). De lijst geeft een beeld van het mondelinge en schriftelijke taalaanbod van vier- tot twaalfjarige leerlingen op de basisschool en is gebaseerd op het mondelinge taalaanbod van leerkrachten, op tekstmateriaal uit prenten- en leesboeken en op teksten uit lesmethoden (taalmethoden Nederlands en zaakvakkenmethoden).

Omdat in de bovenbouw van het basisonderwijs steeds meer abstracte woorden voorkomen en het 'vakjargon' van het onderwijs een steeds grotere plaats inneemt, bevatten de taken Woordenschat vooral abstracte woorden én woorden die vrijwel alleen in schoolse contexten geleerd worden en als het ware over thema's heen aan bod komen.

Verder kan er onderscheid worden gemaakt tussen inhoudswoorden en functiewoorden. De taken Woordenschat bevatten overwegend inhoudswoorden. Dit zijn woorden met een duidelijk omschreven betekenis, zoals zelfstandig naamwoorden, werkwoorden en bijvoeglijk naamwoorden, maar ook woorden die een vaste verbinding vormen. Al deze woorden verwijzen naar voorwerpen, ideeën, activiteiten, plaatsen of eigenschappen. Daarnaast zijn er functiewoorden, de zogenaamde 'kleine' woordjes in de taal, die zonder context weinig betekenis hebben en meestal in combinatie met inhoudswoorden voorkomen. Lidwoorden, telwoorden, voornaamwoorden, voorzetsels, voegwoorden en bijwoorden zijn voorbeelden van functiewoorden.

Tabel 3.24 laat zien welke woordsoorten in de taken Woordenschat voorkomen en wat hun frequentie is. De verdeling van de verschillende woordsoorten over de beide taken benadert de verdeling in woordsoorten in de Nederlandse taal: er komen relatief weinig functiewoorden voor.

Tabel 3.24 Percentages en aantallen stimuluswoorden in de taken in het onderdeel Woordenschat

Type stimuluswoord	Taak 18	Taak 19	Totaal
Inhoudswoorden			
Zelfstandige naamwoorden	35% 7	35% 7	35% 14
Werkwoorden	20% 4	15% 3	35% 7
Bijvoeglijke naamwoorden	25% 5	30% 6	27,5% 11
Woorden die een vaste verbinding vormen	5% 1	5% 1	5% 2
Functiewoorden			
Bijwoorden	15% 3	15% 3	15% 6

De verzameling woorden en woordcombinaties in de taken is niet alleen representatief naar woordsoort, maar ook evenwichtig gespreid naar de mate waarin zij vaker voorkomen volgens deze lijst. Ook bij de keuze van de antwoordalternatieven bij de opgaven werd de lijst van Schrooten en Vermeer gebruikt. Dit om uit te sluiten dat de woorden in de antwoordalternatieven moeilijker zouden zijn dan het stimuluswoord zelf.

De opgaven in de taken Woordenschat doen een beroep op de receptieve woordenschat van de leerlingen. Het kunnen begrijpen van mondelinge en schriftelijke informatie neemt immers een belangrijke plaats in het

onderwijs in. Het gaat daarbij om het kunnen identificeren en herkennen van woorden. Alle opgaven zijn meerkeuzeopgaven. Ze bestaan uit een stimuluswoord met vier antwoordalternatieven; één van de vier alternatieven definieert de betekenis of betekenisrelatie. De stimuluswoorden zijn vetgedrukt en maken deel uit van een korte zin, van een vraag of van een rijtje woorden dat moet worden afgemaakt. De alternatieven bestaan uit woorden of uit een nog ontbrekend laatste deel van een zin. In alle gevallen vormt de gehele opgave – vraag, zin of woordrij plus antwoordalternatieven – de context.

In tabel 3.25 wordt de indeling van de opgaven naar de aspecten betekenis en betekenisrelaties en stimuluswoorden weergegeven. Er is geen specifieke (dezelfde) verdeling nagestreefd van de stimuluswoorden over de categorieën Betekenis en Betekenisrelaties. De gerealiseerde verdeling van de stimuluswoorden kan nogal verschillen. Zo zijn er relatief veel zelfstandige naamwoorden opgenomen als stimuluswoord bij opgaven in de categorie Betekenis (11) in vergelijking met de opgaven in de categorie Betekenisrelaties (3). Dit is min of meer toevallig.

Tabel 3.25 Verdeling van opgaven in de taken Woordenschat naar Betekenis en betekenisrelaties en stimuluswoord

Opgaventype	Opgavenummers in taak 18	Opgavenummers in taak 19	Aantal opgaven
Betekenis			
<i>Inhoudswoorden</i>			
- Zelfstandige naamwoorden	3,6,7,8,9	2,5,6,9,10,11	11
- Werkwoorden	18	1	2
- Bijvoeglijke naamwoorden	2,17	4, 8,16	5
- Woorden die een vaste verbinding vormen	4	---	1
<i>Functiewoorden</i>			
- Bijwoorden	1,5	3,7,15	5
Betekenisrelaties			
<i>Inhoudswoorden</i>			
- Zelfstandige naamwoorden	14,19	14	3
- Werkwoorden	15,16,20	12,13	5
- Bijvoeglijke naamwoorden	10,11,12	17,18,19	6
- Woorden die een vaste verbinding vormen	---	20	1
<i>Functiewoorden</i>			
- Bijwoorden	13		1

3.2.2.4 Wereldoriëntatie

Begripsomschrijving

Wereldoriëntatie is een verzamelnaam voor onderwijs gericht op kennis van de wereld. Het uitgangspunt bij dit onderwijs is steeds een concreet verschijnsel in de wereld zoals leerlingen dat leren kennen uit eigen ervaring of via allerlei media. In het basisonderwijs blijft onderwijs in de theoretische achtergronden bij deze verschijnselen beperkt.

In het wereldoriëntatie-onderwijs worden de verschijnselen vanuit verschillende perspectieven verkend: het fysisch-biologische, het politieke, economische of sociaal-culturele perspectief. Soms is dit onderwijs georganiseerd in projecten, zoals 'afval' of 'onze stad', maar de meeste scholen gebruiken een methode voor de kennisgebieden aardrijkskunde, geschiedenis en natuur en techniek.

Inhoudsverantwoording

De inhoud die in de taken Wereldoriëntatie aan bod komen, zijn gebaseerd op verschillende bronnen, waaronder:

- De kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006). De leerdoelen voor wereldoriëntatie in het primair onderwijs zijn beschreven in de Kerndoelen Oriëntatie op je zelf en de wereld. Ze bevatten beschrijvingen van de gewenste kennis en vaardigheden. Er zijn kerndoelen bij de thema's Mens en samenleving (kerndoel 34 tot en met 39), Natuur en Techniek (kerndoel 40 tot en met 46), Ruimte (kerndoel 47 tot en met 50) en Tijd (kerndoel 52 tot en met 53).
- Toetswijzer bij de centrale eindtoets primair onderwijs (van der Schee, van Graft, Bakker, van Herwaarden, Keissen, Notté & Wagenaar, 2012).

De inhoud van het onderdeel Wereldoriëntatie

Domeinen

In het basisonderwijs gebruiken we de indeling naar de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek.

Bij Aardrijkskunde is er aandacht voor de bestudering van gebieden vanuit verschillende perspectieven.

- Aarde, landschap en milieu
- Bevolking, wonen en welzijn
- Grondstoffen, bedrijven en transport
- Topografie

Bij Geschiedenis worden de tien tijdvakken uitgewerkt en is er aparte aandacht voor maatschappelijke verhoudingen waaronder staatsinrichting en geestelijke stromingen.

- Tijd van jagers en boeren
- Tijd van Grieken en Romeinen
- Tijd van monniken en ridders
- Tijd van steden en staten
- Tijd van ontdekkers en hervormers
- Tijd van regenten en vorsten
- Tijd van pruiken en revoluties
- Tijd van burgers en stoommachines
- Tijd van de wereldoorlogen en holocaust
- Tijd van televisie en computer
- Maatschappelijke verhoudingen waaronder staatsinrichting en geestelijke stromingen.

De toetsdoelen voor natuur en techniek bestrijken vier deeldomeinen.

- Biologie, met de onderwerpen biologische eenheid, instandhouding, gedrag en interactie, voortplanting, groei en ontwikkeling;
- Natuurkunde met de onderwerpen krachten, energie, licht & kleur, elektriciteit, geluid en eigenschappen van stoffen;
- Techniek met de onderwerpen eigenschappen van materiaal, constructies & verbindingen, overbrengingen en geautomatiseerde systemen;
- Weer en heelal.

Elk domein wordt in de Entreetoets in een aparte taak van 20 opgaven aangeboden. De taken 20 tot en met 22 bevatten opgaven van respectievelijk de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek.

Studievaardigheden

Leerlingen maken op verschillende manieren kennis met de verschijnselen bij de hierboven genoemde onderwerpen: door eigen waarneming, door te luisteren naar verhalen, door teksten te lezen of beelden te bekijken. Het opdoen en verwerken van deze informatie is het domein van studievaardigheden. Er zijn twee categorieën studievaardigheden: vaardigheden bij talige bronnen en bij grafische informatiebronnen. Bij talige informatieverwerking gaat het om het lezen van studieteksten en het gericht opzoeken van (digitale) informatie. Bij grafische informatieverwerking gaat het om lezen van schema's, tabellen, diagrammen en grafieken en het kaartlezen. Kaartlezen is een speciaal onderwerp bij de kerndoelen over Ruimte en in de Toetswijzer wereldoriëntatie. Voor de andere studievaardigheden legt het onderwijs in taal en rekenen de basis.

Ook bij wereldoriëntatie wordt de informatie geselecteerd, gelezen, geanalyseerd en geïnterpreteerd, maar er wordt een extra vaardigheid verlangd. Leerlingen moeten nadat ze de informatiebron begrepen hebben een relatie leggen met kennis die al eerder bij wereldoriëntatie is opgedaan. Leren op deze manier is een zichzelf versterkend proces: een goede analyse en interpretatie van informatie leiden tot meer kennis, die vervolgens nieuwe kennisverwerving vergemakkelijkt.

Hieronder worden de studievaardigheden kort nader toegelicht.

- Lezen van studieteksten

Bij *Lezen van studieteksten* staat het kunnen structureren en interpreteren van een tekst centraal. Leerlingen moeten de hoofd- en bijzaken onderscheiden en de tekst of delen van de tekst verkort weergeven. Ze gebruiken deze kennis om efficiënt informatie uit een tekst te halen.

- Opzoeken van informatie

Hier gaat het om algemene vaardigheden voor het *opzoeken van informatie*, zoals het selecteren van een geschikte informatiebron, het doorzien van de structuur van een informatiebron en het gebruiken van ingangen tot de bron.

- Kaartlezen

Bij *Kaartlezen* laten leerlingen zien dat ze zich een voorstelling kunnen maken bij een gebied op de kaart en deze informatie kunnen gebruiken om een probleem op te lossen. Zij passen deze vaardigheden toe op kaarten die variëren in schaal en in mate van dichtheid en abstractie van de informatie op de kaart.

- Lezen van schema's, tabellen, diagrammen en grafieken

De vaardigheid Lezen van schema's, tabellen, diagrammen en grafieken, gaat om het selecteren, lezen, analyseren en interpreteren van informatie uit grafische weergaven.

De opgaven

De taken en opgaven van het onderdeel Wereldoriëntatie zijn ingedeeld volgens de eerder genoemde kennisgebieden Aardrijkskunde, Geschiedenis en Natuur en techniek. De kennis en vaardigheden die worden getoetst zijn in overeenstemming met de leerdoelen zoals beschreven in de Kerndoelen Oriëntatie op je zelf en de wereld (Natuur en Techniek, Ruimte en Tijd) voor het primair onderwijs en de Toetswijzer bij de centrale eindtoets PO wereldoriëntatie.

In de opgaven worden leerlingen geconfronteerd met voor hen onbekende contexten. Deze informatie wordt gepresenteerd in een bronnenboekje bij de opgaven. Leerlingen worden uitgedaagd hun kennis van wereldoriëntatie toe te passen.

De informatiebronnen hebben betrekking op allerlei concrete verschijnselen in twee voor de meeste Nederlandse leerlingen onbekende contexten: de stad Gent en het eiland Tenerife. Om de opgaven tot een goed einde te brengen is geen kennis vereist van deze gebieden. Ze zijn als context gekozen vanwege hun rijkdom aan verschijnselen, waardoor het mogelijk is uiteenlopende domeinen van wereldoriëntatie toe te passen. Tenerife is een toeristengebied met een grote variatie aan landschappen en natuur. Gent heeft veel overeenkomsten met Nederlandse steden en een duidelijke relatie met de geschiedenis van Nederland.

Deze toetsopzet maakt het mogelijk om voor wereldoriëntatie belangrijke kennis en vaardigheden in een samenhangende, voor leerlingen betekenisvolle context te toetsen. Daarmee wordt niet het gehele domein van wereldoriëntatie bestreken. Andere belangrijke doelen van het onderwijs in wereldoriëntatie, zoals het

doen van onderzoek of het beargumenteren van een standpunt, zijn niet in een schriftelijke toets te operationaliseren.

Er zijn andere nadelen verbonden aan deze toetsopzet. Het accent ligt duidelijk op toepassing van kennis, waardoor de reproductie van kennis onderbelicht blijft. Ook door het ontbreken van Nederlandse contexten, kunnen sommige toetsdoelen, bijvoorbeeld over Nederlandse landschappen of de staatinrichting in Nederland niet in de taken gerepresenteerd worden.

Tabel 3.26 geeft een overzicht van de opgaven in de taken Wereldoriëntatie, verdeeld over de kennisgebieden en de studievaardigheden. Voor de toets zijn 60 opgaven gereserveerd. Hierdoor blijft een verslag van de prestaties van de leerling beperkt tot scores over de drie onderscheiden kennisdomeinen en de beide vaardigheidsdomeinen (Talige vaardigheden en Grafische vaardigheden).

Tabel 3.26 Percentages en aantallen opgaven van de verschillende studievaardigheden in de taken Wereldoriëntatie

Studievaardigheden	Kennisgebieden			
	Aardrijkkunde Taak 20	Geschiedenis Taak 21	Natuur & Techniek Taak 22	Totaal
<i>Talige vaardigheden</i>	30% 6	70% 14	50% 10	50% 30 (30)
Lezen	10% 2	50% 10	30% 6	30% 18 (15)
Opzoeken	20% 4	20% 4	20% 4	20% 12 (15)
<i>Grafische vaardigheden</i>	70% 14	30% 6	50% 10	50% 30 (30)
Kaartlezen	50% 10	15% 3	10% 2	25% 15 (15)
Schema's, tabellen, diagrammen en grafieken	20% 4	15% 3	40% 8	25% 15 (15)
<i>Totaal</i>	100% 20 (20)	100% 20 (20)	100% 20 (20)	100% 60 (60)

In de tabel zijn steeds tussen haakjes de gewenste aantallen opgaven aangegeven. Deze aantallen zijn niet gespecificeerd voor de cellen in de matrix. Vooraf werd vastgelegd dat de opgaven gelijk verdeeld moesten zijn over de drie kennisgebieden (ieder 20 opgaven). Ook de verdeling over de twee hoofdvvaardigheden (talige en grafische vaardigheden; ieder 30 opgaven) zou gelijk moeten zijn. Er is geen duidelijk aanwijsbare reden om niet uit te gaan van deze gelijke verdeling. Bovendien moest betrouwbare rapportage van de prestaties op deze kennis- en vaardigheidsdimensies mogelijk zijn, wat eisen stelt aan de minimum aantallen opgaven. Daarnaast was een evenwichtige verdeling van de opgaven over de vier deelvaardigheden (ieder 15 opgaven) eveneens het uitgangspunt bij de opgavenconstructie- en selectie. Daarbij gold de aanname dat sommige vaardigheden beter bij bepaalde kennisgebieden passen dan bij andere. De verwachting is dat de verdeling van vaardigheden over specifieke kennisgebieden min of meer organisch zou voortkomen uit hun specifieke aard.

Te constateren valt dat de aantallen gerealiseerde opgaven per kennisgebied in overeenstemming zijn met de gewenste aantallen. Dat geldt ook voor hoofdvvaardigheden in de studievaardigheden. Voor de grafische vaardigheden geldt dat er steeds overeenkomstig de bedoelingen 15 opgaven zijn gerealiseerd voor 'Kaartlezen' en 'Schema's, tabellen, diagrammen en grafieken'. Voor de overige vaardigheden geldt dat er 18 opgaven zijn geselecteerd voor Lezen en maar 12 voor Opzoeken. Bij de proefvoetsing bleek dat voor Opzoeken te weinig psychometrisch geschikte opgaven voorhanden waren. Dit werd gecompenseerd door de nog ontbrekende opgaven te zoeken binnen het andere talige vaardigheidendomein (Lezen).

Zoals verwacht passen sommige vaardigheden beter bij bepaalde kennisgebieden dan bij andere. Goede voorbeelden zijn het hoge aantal opgaven voor Kaartlezen bij Aardrijkkunde (10 opgaven), het hoge aantal

opgaven voor talige vaardigheden bij Geschiedenis (14), met name de vaardigheid Lezen (10 opgaven) en het hoge aantal opgaven voor Schema's, tabellen, diagrammen en grafieken voor het kennisgebied Natuur en techniek (8).

In tabel 3.27 zijn de opgavenummers ingedeeld naar vaardigheden. Met behulp van de opgavenummers is te achterhalen welke specifieke vaardigheid in elke opgaven wordt getoetst. De opgaven zijn genummerd per kennisgebied.

Tabel 3.27 Opgavennummers per vaardigheid en per kennisgebied van Wereldoriëntatie

Studievaardigheden	Taak 20 Aardrijkskunde	Taak 21 Geschiedenis	Taak 22 Natuur & techniek
Talige vaardigheden			
• Lezen	1, 18	2, 3, 6, 7, 9, 10, 11, 13, 14, 20	3, 4, 9, 13, 14, 19
• Opzoeken	2, 7, 8, 13	4, 12, 17, 18	7, 12, 16, 17
Grafische vaardigheden			
• Kaartlezen	5, 6, 9, 10, 11, 12, 14, 16, 19, 20	1, 15, 16	2, 10
• Schema's, tabellen, diagrammen en grafieken	3, 4, 15, 17	5, 8, 19	1, 5, 6, 8, 11, 15, 18, 20

3.3 Kenmerken van de toets: beschrijvende gegevens

Itemkenmerken: moeilijkheidsgraad en interne consistentie

Wat de moeilijkheid van de opgaven betreft: voor de opgavenselectie geldt het uitgangspunt dat de p-waarden per onderdeel bij voorkeur tussen .40 - .90 moeten liggen. Bovendien zouden ze (eveneens per onderdeel) een gemiddelde P-waarde moeten hebben van rond .70. In tabel 3.28 rapporteren we de range van P-waarden en de gemiddelde p-waarde van de opgaven voor de Entreetoets groep 7.

De waarden zijn berekend op de afnamedata van de toets in 2015.

Omdat de Entreetoets groep 7 in zijn huidige vorm is opgesplitst in verschillende modules, moeten we de kengetallen voor de verschillende onderdelen in verschillende tabellen opnemen. Tabel 3.28a heeft betrekking op de module Basis. In deze module is sprake van een beperkt aantal opgaven per onderdeel; er wordt over deze onderdelen alleen globaal gerapporteerd (dat wil zeggen over Rekenen, Lezen en Taalverzorging). Niettemin geven we informatie over de opgaven op het niveau van de domeinen (dat wil zeggen bijvoorbeeld voor Rekenen op de Verhoudingen, Getallen, Verbanden en Meten). Op die manier zijn de resultaten vergelijkbaar met de resultaten voor de domeinen die in de Verdiepende module zijn opgenomen. Deze resultaten staan in tabel 3.28b. De gegevens over de opgaven van de module Verbreding ten slotte staan in tabel 3.28c.

Tabel 3.28 Range en gemiddelde van P- en R_{it}-waarden per onderdeel en domein

Tabel 3.28a De onderdelen van de module Basis²

Onderdeel/domein	P-waarden		R _{it} -waarden		Aantal items
	Range	Gemiddeld	Range	Gemiddeld	
Rekenen	.42 - .91	.68	.23 - .58	.39	90
Verhoudingen	.42 - .84	.65	.32 - .51	.44	20
Getallen	.42 - .91	.70	.23 - .58	.38	30
Verbanden	.50 - .90	.70	.30 - .51	.40	15
Meten	.47 - .80	.65	.26 - .48	.36	25
Lezen	.47 - .86	.68	.23 - .50	.37	40
Begrijpend lezen	.50 - .86	.69	.23 - .50	.37	20
Samenvatten	.48 - .82	.67	.33 - .46	.38	10
Opzoeken	.47 - .86	.68	.29 - .44	.37	10
Taalverzorging	.24 - .93	.68	.18 - .44	.32	50
Spelling niet-werkwoorden	.44 - .89	.73	.18 - .44	.31	15
Grammatica	.52 - .85	.72	.31 - .44	.38	10
Interpunctie	.43 - .93	.72	.23 - .43	.32	10
Spelling werkwoorden	.24 - .76	.57	.23 - .44	.30	15

Tabel 3.28b De onderdelen en domeinen van de modules Basis en Verdieping

Onderdeel/domein	P-waarden		R _{it} -waarden		Aantal items
	Range	Gemiddeld	Range	Gemiddeld	
Rekenen	.29 - .92	.69	.17 - .58	.38	140
Verhoudingen	.43 - .86	.68	.28 - .55	.43	30
Getallen	.43 - .92	.70	.22 - .58	.37	50
Verbanden	.50 - .90	.70	.17 - .50	.38	25
Meten	.29 - .86	.67	.19 - .48	.34	35
Lezen	.44 - .89	.69	.18 - .48	.35	80
Begrijpend lezen	.44 - .87	.69	.18 - .48	.34	40
Samenvatten	.48 - .88	.69	.28 - .43	.36	20
Opzoeken	.48 - .89	.69	.25 - .44	.36	20
Taalverzorging	.24 - .94	.70	.13 - .44	.31	90
Spelling niet-werkwoorden	.44 - .89	.75	.13 - .43	.29	25
Grammatica	.54 - .90	.74	.29 - .44	.36	20
Interpunctie	.44 - .94	.76	.21 - .44	.32	20
Spelling werkwoorden	.24 - .84	.58	.13 - .43	.28	25

² In de tabel zijn voor de module Basis ook itemkenmerken voor de domeinen opgenomen. Deze domeinen worden echter uitsluitend afzonderlijk gerapporteerd wanneer ook de verdiepende module is afgenomen.

Tabel 3.28c De onderdelen en domeinen van de module *Verbreiding*

Onderdeel/domein	P-waarden		R _{it} -waarden		Aantal items
	Range	Gemiddeld	Range	Gemiddeld	
Luisteren	.37 - .95	.71	.21 - .52	.35	30
Schrijven	.46 - .84	.69	.22 - .54	.39	30
Woordenschat	.39 - .95	.66	.21 - .45	.32	40
Wereldoriëntatie³	.26 - .92	.63	.10 - .48	.34	60
Aardrijkskunde	.38 - .91	.64	.31 - .52	.40	20
Geschiedenis	.27 - .84	.59	.20 - .50	.37	20
Natuur en Techniek	.40 - .84	.63	.32 - .49	.40	20

Wanneer we eerst de domeinen van de module Basis bezien, valt op dat de gemiddelde P-waarden van de domeinen van de onderdelen Rekenen en Lezen in de buurt liggen van de nagestreefde .70, maar nooit hoger zijn dan .70. Voor de domeinen van het onderdeel Taalverzorging geldt dat de gemiddelde P-waarden juist iets boven .70 liggen, met uitzondering van de gemiddelde P-waarde voor Spelling werkwoorden (.57). De 15 opgaven voor dit domein zijn dus aan de moeilijke kant. In de module Basis weegt dit op tegen de wat hogere gemiddelden voor de andere domeinen van Taalverzorging, maar bij specifieke rapportage over Spelling werkwoorden op basis van de verdiepende module is de score dus gebaseerd op een relatief moeilijke toets (met een gemiddelde P-waarde van .58). Het is ook het enige onderdeel in de module Basis met een P-waarde onder de nagestreefde .40. Er zijn twee van dergelijke opgaven met P-waarden van .24 en .39. Te gemakkelijke opgaven (met P-waarden van .91 of hoger) komen in de module Basis sporadisch voor, namelijk één opgave bij Rekenen (Getallen) en één opgave bij Taalverzorging (Interpunctie). De gemiddelde R_{it}-waarden zijn in de module Basis hoog: ze variëren tussen .30 en .44. Het COTAN-beoordelings-systeem geeft voor R_{it}-waarden als kwaliteitscriterium aan, dat waarden boven de .30 als goed beoordeeld worden. De gemiddelde R_{it}-waarden voldoen dus ruim aan deze norm en pleiten voor de zorgvuldigheid van de itemselectie. R_{it}-waarden lager dan .20 worden volgens dezelfde criteria als onvoldoende beoordeeld. In de module Basis komt maar één opgave (van de 180) met zo'n relatief lage waarde voor (bij Spelling niet-werkwoorden). Overigens vormen R_{it}-waarden een overschatting en zijn R_{ir}-waarden eigenlijk te prefereren omdat zij een realistischer beeld geven van de correlatie met de schaa score, maar helaas zijn ons geen normgegevens bekend voor R_{ir}. Voor R_{it}-waarden kent het COTAN-beoordelingssysteem (Evers, Lucassen, Meijer & Sijtsma, 2010) wél kwaliteitscriteria. Voor de basistoets geldt dat met name wordt gerapporteerd over de onderdelen Rekenen. Lezen en Taalverzorging als geheel (scores voor de domeinen verschijnen uitsluitend op het Leerlingprofiel indien de module Verdieping is afgenomen). Daarom zijn ook steeds de samenvattende gegevens voor de onderdelen (in de tabel vetgedrukt) opgenomen. Deze gegevens behoeven geen nadere toelichting in relatie tot wat hierboven al is gezegd over de onderdelen.

Als we aan de hand van tabel 3.28b conclusies trekken over de module Verdieping zijn deze vergelijkbaar met het bovenstaande. Deels komt dit uiteraard omdat de betreffende uitkomsten niet onafhankelijk van elkaar zijn. In tabel 3.28b zijn immers ook de opgaven van de module Basis meegenomen. De bevindingen met betrekking tot (gemiddelde) P-waarden zijn voor de basistoets en

³ Bij het berekenen van deze waarden is Wereldoriëntatie geschaald alsof het één toetsonderdeel vormt, bestaande uit 60 opgaven. De domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek zijn ieder gekalibreerd op hun eigen vaardigheidsschaal.

de basistoets met module Verdieping vergelijkbaar. Voor Rekenen (Meten) is er sprake van een moeilijke opgave (P-waarde .29) extra. De gemiddelde R_{it} -waarden voor de verdiepende module variëren tussen .29 en .43 en zijn dus ongeveer even hoog als voor de basistoets. Het aantal 'onvoldoende' R_{it} -waarden ligt voor de modules Basis en Verdieping samen op 5 (van in totaal 310 items). De betreffende waarden variëren van .13 tot .19 en zijn te vinden bij Rekenen (Verbanden en Meten), Begrijpend lezen, Spelling niet-werkwoorden en Spelling werkwoorden. De basistoets en de module Verdieping overziende, kunnen we concluderen dat de itemkwaliteit van beide vergelijkbaar hoog is. Ook hier hebben we de samenvattende gegevens voor Rekenen, Lezen en Taalverzorging (vergedrukt) aan de tabel toegevoegd, al is het bij de module Verdieping vooral interessant om de gegevens voor de (verlengde) domeinen in ogenschouw te nemen. Juist de toetsverlenging maakt het mogelijk om betrouwbaar op dit gedifferentieerde niveau te rapporteren, zodat de leerkracht ook maximaal kan differentiëren naar zwakke en sterke kanten van de leerlingen.

Voor de module Verbreding (tabel 3.28c) zijn over de itemparameters soortgelijke conclusies te trekken als hierboven. De gemiddelde P-waarden liggen tussen .63 en .71. Ze zijn over het algemeen iets lager dan de beoogde waarden van omstreeks .70. Er is één te moeilijke opgave (.27) bij het domein Geschiedenis. Er zijn 8 opgaven met een P-waarde van .91 of hoger (bij Luisteren drie, bij Woordenschat vier en bij Aardrijkskunde één). Ondanks de hoge P-waarde is het onderscheidend vermogen van de betreffende opgaven aanzienlijk. Geen enkele opgave in de module Verbreding heeft een R_{it} -waarde lager dan .20. De gemiddelde R_{it} -waarden liggen tussen .32 en .40.

Verdeling van de ruwe scores

In tabel 3.29 zijn de verdelingskarakteristieken gegeven van de ruwe scores op de onderscheiden toetsonderdelen, zowel voor de module Basis (tabel 3.29a) als voor de modules (Basis en) Verdieping (tabel 3.29b) en Verbreding (tabel 3.29c). De cijfers zijn gebaseerd op de afnamegegevens van 2015.

Tabel 3.29 Verdelingskenmerken van alle onderdelen van de Entreetoets groep 7

Tabel 3.29a De onderdelen en domeinen van de module Basis

Onderdeel/domein	Aantal opgaven	Min	Max	M	SD	Skewness	Kurtosis	N
Totaal	180	33	179	122.3	26.14	-.368	-.505	56196
Rekenen	90	7	90	6.9	15.97	-.380	-.659	56225
Verhoudingen	20	0	20	13.0	4.42	-.416	-.683	56224
Getallen	30	1	30	21.0	5.51	-.522	-.391	56225
Verbanden	15	0	15	1.5	3.10	-.578	-.373	56225
Meten	25	1	25	16.3	4.58	-.244	-.626	56225
Lezen	40	4	40	27.6	6.57	-.508	-.345	56354
Begrijpend lezen	20	1	20	13.9	3.58	-.585	-.218	56353
Samenvatten	10	0	10	6.8	2.14	-.487	-.389	56348
Opzoeken	10	0	10	6.9	2.02	-.499	-.205	56345
Taalverzorging	50	6	50	33.8	7.07	-.327	-.251	56337
Spelling niet-werkwoorden	15	0	15	1.9	2.46	-.591	.058	56333
Grammatica	10	0	10	7.2	2.19	-.670	-.228	56335
Interpunctie	10	0	10	7.2	1.94	-.506	-.312	56332
Spelling werkwoorden	15	0	15	8.5	2.69	.032	-.383	56329

Tabel 3.29b De onderdelen en domeinen van de modulen Basis en Verdieping

Onderdeel/domein	Aantal opgaven	Min	Max	M	SD	Skewness	Kurtosis	N
Rekenen	140	13	140	96,0	23,88	-0,434	-0,575	35688
Verhoudingen	30	1	30	20,3	6,27	-0,538	-0,544	35687
Getallen	50	2	50	35,0	8,77	-0,506	-0,418	35688
Verbanden	25	1	25	17,5	4,75	-0,523	-0,445	35689
Meten	35	3	35	23,3	5,98	-0,326	-0,507	35689
Lezen	80	11	80	55,4	12,37	-0,539	-0,323	35637
Begrijpend lezen	40	3	40	27,8	6,35	-0,585	-0,188	35638
Samenvatten	20	0	20	13,8	3,71	-0,522	-0,321	35638
Opzoeken	20	1	20	13,8	3,66	-0,508	-0,299	35637
Taalverzorging	90	16	90	62,6	12,00	-0,426	-0,108	35636
Spelling niet-werkwoorden	25	2	25	18,6	3,69	-0,671	0,219	35635
Grammatica	20	0	20	14,6	3,91	-0,672	-0,203	35636
Interpunctie	20	1	20	15,1	3,35	-0,644	-0,099	35636
Spelling werkwoorden	25	1	25	14,3	4,09	0,063	-0,318	35635

Tabel 3.29c De onderdelen en domeinen van de module Verbreding

Onderdeel/domein	Aantal opgaven	Min	Max	M	SD	Skewness	Kurtosis	N
Luisteren	30	1	30	21,1	4,73	-0,778	0,558	8771
Schrijven	30	1	30	20,6	5,25	-0,661	-0,100	11146
Woordenschat	40	4	40	26,5	5,64	-0,361	-0,128	13323
Wereldoriëntatie	60	4	60	37,3	9,47	-0,286	-0,531	10413
Aardrijkskunde	20	1	20	12,8	3,68	-0,323	-0,528	10415
Geschiedenis	20	0	20	11,9	3,44	-0,238	-0,429	10404
Natuur en Techniek	20	1	20	12,7	3,67	-0,279	-0,548	10369

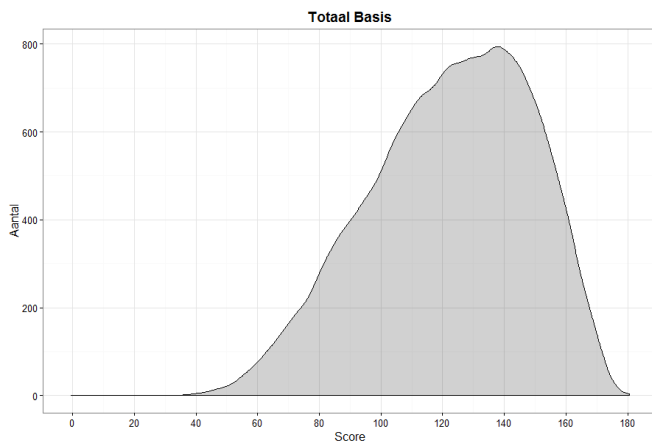
Bezien we allereerst de psychometrische kenmerken voor de onderdelen van de module Basis. Het betreft de 56.354 leerlingen die uitsluitend de Basisoets hebben gemaakt én alle leerlingen bij wie daarnaast ook opgaven uit de module Verdieping zijn afgenomen. Soms wijken de aantallen licht af vanwege ontbrekende gegevens. De gemiddelden komen uiteraard overeen met wat men bij een gegeven aantal items mag verwachten bij de gekozen (gemiddelde) moeilijkheidsgraad (variërend tussen .57 en .73; zie tabel 3.28a). Alle verdelingen zijn linksscheef (vergelijk de negatieve waarden in de kolom 'skewness'), de ene wat meer dan de andere, met het praktisch symmetrisch verdeelde domein Spelling werkwoorden als enige uitzondering. Spelling werkwoorden is voor groep 7 leerlingen verreweg het moeilijkste onderdeel omdat het over het algemeen pas betrekkelijk kort tevoren in het onderwijs is geïntroduceerd. Bij de LVS-toetsen Taalverzorging voor de groep 6, 7 en 8 is een soortgelijk beeld vastgesteld (vergelijk Roumans & Engelen, in voorbereiding). Qua scheefheid ontlopen de verdelingen elkaar niet veel, de parameters variëren tussen -.244 en -.670. Alle verdelingen zijn ééntoppig en lijken onderling vrij sterk op elkaar. Ook in dit opzicht is het een onderdeel van Taalverzorging (Spelling niet-werkwoorden) dat een uitzondering vormt met een

positieve kurtosis-waarde. Bij alle domeinen is de maximum-score gelijk aan de maximaal haalbare score. Aan de andere kant zijn er enkele domeinen waarbij geen enkele leerling alle opgaven fout had. Voor dezelfde onderdelen en domeinen, maar dan inclusief de opgaven die in het kader van de module Verdieping aanvullend zijn opgenomen (N = 35.689), staan de psychometrische kenmerken in tabel 3.29b. De taalonderdelen/domeinen van deze module zijn in enkele groepen niet afgenomen (waar Rekenen wél volledig werd gemaakt). De uitkomsten voor de onderdelen van Verdieping sluiten exact aan bij de uitkomsten voor de module Basis (zie hierboven), met inbegrip van de uitzonderingen bij de spellingdomeinen. Dat is op zichzelf goed te begrijpen omdat de resultaten op de opgaven van de basistoets in de scores zijn verwerkt. Aan de andere kant laten deze resultaten zien dat we erin geslaagd zijn de onderdelen en domeinen van de modules Basis en Verdieping op vergelijkbare wijze vorm te geven. De psychometrische kenmerken van de domeinen van de module Verbreding staan in tabel 3.29c. Niet alle onderdelen van deze module zijn even vaak afgenomen. Het onderdeel Luisteren is het minst vaak gemaakt (N = 8771). Mogelijk is dit op te vatten als een indicatie dat bij deze mondelinge taalvaardigheid wat minder prioriteit wordt gelegd in het basisonderwijs. Schrijven (N = 11.146) en Woordenschat (N = 13.323) worden wat vaker afgenomen, de onderdelen van Wereldoriëntatie weer wat minder (N = omstreeks 10.400). Ook hier is (in overeenstemming met de gekozen moeilijkheidsgraad) sprake van linksscheve verdelingen, waarbij de domeinen Wereldoriëntatie wat moeilijker zijn dan de taaldomeinen in de module. De (negatieve) kurtosis is over het algemeen vergelijkbaar met de kurtosis van de domeinen van de basistoets en de module Verdieping. De (positieve) kurtosis bij Luisteren vormt hierop een uitzondering.

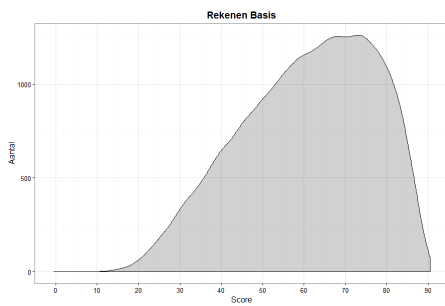
Voor de belangrijkste ruwe scores (Totaalscore, score op Rekenen, Lezen en Taalverzorging) zijn de verdelingen ook grafisch weergegeven in figuur 3.1 (a tot en met d). De verdelingen van de verschillende domeinen op 'lagere' aggregatieniveaus zijn niet afgebeeld, maar vertonen een sterke gelijkenis met de wél afgebeelde verdelingen.

Zoals al af te lezen was uit de cijfers in tabel 3.29 zijn alle verdelingen linksscheef, waarbij de verdeling voor Taalverzorging nog het meest symmetrisch is. Alle verdelingen zijn ééntoppig, waarbij de ruwe scores in frequentie geleidelijk oplopen.

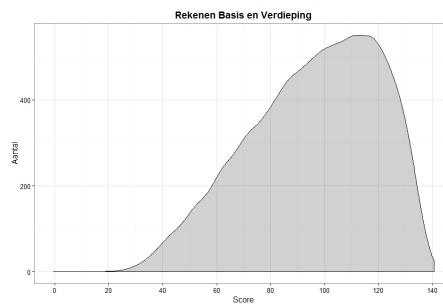
Figuur 3.1 Verdeling van de ruwe scores op de Entreetoets groep 7 – afname 2015 – voor de Totaalscore (a) en voor de scores op Rekenen (b), Lezen (c) en Taalverzorging (d)



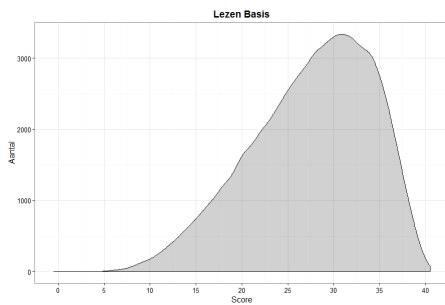
a Totaalscore (module Basis)



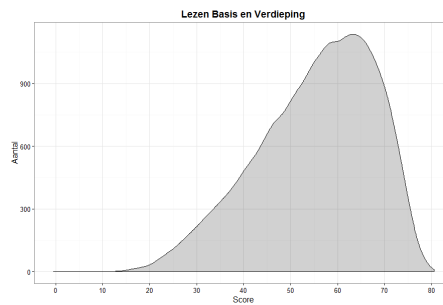
b Rekenen (module Basis)



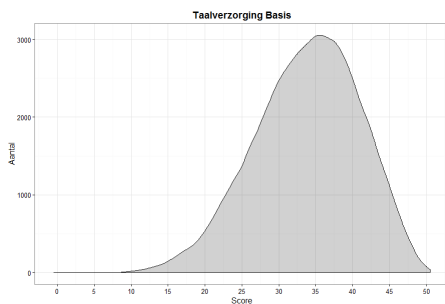
Rekenen (module Verdieping)



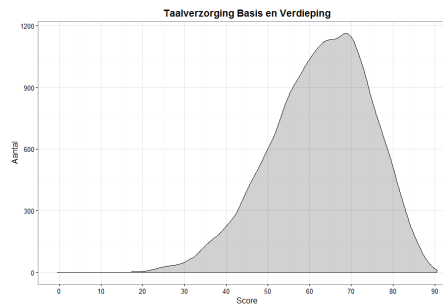
c Lezen (module Basis)



Lezen (module Verdieping)



d Taalverzorging (module Basis)



Taalverzorging (module Verdieping)

4 Kalibratie en normering

In de vorige hoofdstukken is de belangrijkste functie van de Entreetoets beschreven, namelijk het op objectieve wijze antwoord geven op de vraag hoe elke leerling afzonderlijk ervoor staat, waar hij of zij goed in is en waar misschien ondersteuning nodig is. Daarnaast zijn de resultaten op de Entreetoets te gebruiken als informatie voor het formuleren van een advies voor het brugklatype dat het beste past bij het niveau van leervorderingen van elke leerling (zie over de functies van de Entreetoets groep 7 hoofdstuk 2). Om de informatie die de toets geeft over wat een leerling heeft geleerd goed te kunnen duiden, is een adequate normering cruciaal. Aangezien kalibratie in de toetsconstructie nauw verwant is aan normeren, is het belangrijk dat kan worden onderbouwd dat de kalibratieprocedures het gewenste resultaat hebben opgeleverd.

In paragraaf 4.1 bespreken we de werkwijze van het kalibreren en geven we tevens antwoord op de vraag of de kalibratie heeft geleid tot de gewenste resultaten. Een positief antwoord op deze vraag zien we als noodzakelijke voorwaarde voor inhoudsvaliditeit. Op begripsvaliditeit komen we in hoofdstuk 6 terug. Aan de hand van itemparameters worden opgaven toegelaten tot de itembank voor een specifieke vaardigheid en vervolgens worden op grond van deze itemparameters toetsen samengesteld. Uiteindelijk bestaat de normering van deze toetsen uit het vaststellen van de populatieparameters voor de onderliggende vaardigheidsdimensies.

Wie de verschillende rapportages die we op het eind van het vorige hoofdstuk hebben gepresenteerd goed bestudeert, kan tot de conclusie komen dat alle vormen van normering die de COTAN (Evers et al., 2010) onderscheidt aan de orde zijn.

Op de eerste plaats is er sprake van relatieve normering. De rapportage geeft informatie over de toetsresultaten van de leerling in vergelijking met de toetsresultaten van andere leerlingen. Dit gebeurt op het Leerlingprofiel in de vorm van percentielscores, op basis waarvan niveauscores (I tot en met V) worden bepaald. Voor relatieve vergelijking is de representativiteit van de normgroep van cruciaal belang. De vaststelling van de relatieve normering en de evaluatie van de representativiteit van de normgroep vormen het onderwerp van paragraaf 4.2.

Op de tweede plaats is er sprake van vergelijking van het toetsresultaat met prestatiecriteria die door deskundigen zijn vastgelegd. De COTAN spreekt hier van “domeingerichte interpretatie”. De bedoelde prestatiecriteria zijn vastgelegd in de vorm van zogeheten referentieniveaus ((bijvoorbeeld 1F, 1S, 2F). Of een referentieniveau al dan niet is behaald (voor Rekenen, Lezen en Taalverzorging) staat vermeld op het Leerlingprofiel. De rapportage op referentieniveaus en de adequaatheid waarmee dat gebeurt zijn het onderwerp van paragraaf 4.3.

Ten slotte is er sprake van wat de COTAN noemt criteriumgeoriënteerde interpretatie. Deze vorm van normering verwijst naar de vraag welk brugklatype in het vervolgonderwijs (i.e. het criterium) het beste past bij een bepaald toetsresultaat (of bij een bepaalde range van toetsresultaten). Met het oog hierop worden totaalscores berekend (over de module Basis), die gekoppeld worden aan brugklatypes. In paragraaf 4.4 geven we een overzicht van deze normering. Het toelatings- en doorstroomonderzoek dat inzicht geeft in de adequaatheid van deze criteriumgerichte normering, bespreken we in hoofdstuk 6 onder de noemer criteriumvaliditeit.

In aanvulling op de leerlingrapportages zijn er ook rapportages op groepsniveau en op schoolniveau mogelijk. Daarnaast kan men ook de prestatie van de school vergelijken met die van andere scholen, waarbij men al dan niet corrigeert voor leerlinggewicht. De aan deze rapportages ten grondslag liggende normering op schoolniveau is rechtstreeks afgeleid van de normering op individueel niveau (zie hierboven en paragraaf 4.4).

4.1 Kalibratie en het welslagen van de kalibraties op de vaardigheidsschalen

In hoofdstuk 2 zijn in algemene zin de procedures beschreven die hebben geleid tot (1) gekalibreerde opgavenbanken per getoetste vaardigheid en (2) tot de inhoud van de Entreetoets. Tevens is daar ingegaan op het meetmodel dat ten grondslag ligt aan deze procedures. In deze paragraaf gaan we nog wat nauwkeuriger in op de stappen die in de kalibratie zijn genomen (4.1.1) en geven we resultaten van analyses die duidelijk maken dat de kalibratie geslaagd genoemd kan worden (4.1.2).

4.1.1 Het kalibratieproces: de stappen in de kalibratie

Kalibreren geschiedt in eerste instantie aan de hand van de gegevens die zijn verzameld in de proef-toetsen. In een proeftoets wordt nagegaan welke opgaven niet passen in de opgavenbank (en de toets) en wordt geschat wat de moeilijkheidsgraad van de opgaven is. Opgaven die niet voldoen aan de modeleisen of te gemakkelijk of te moeilijk zijn, vallen af. Welke opgaven in de toets worden opgenomen hangt niet alleen af van hun psychometrische kenmerken, maar ook van inhoudelijke eisen (zie hoofdstuk 3). Soms is afweging nodig. Dan moet omwille van de inhoud een psychometrisch goede opgave vervallen.

Het omgekeerde is ook mogelijk: een psychometrisch gezien minder goede opgave (die wel voldoet aan de model-eisen) wordt in de toets opgenomen omdat een bepaald inhoudelijk facet of aspect bij het niet opnemen van de opgave anders niet voldoende vertegenwoordigd zou zijn.

Kalibratie geschiedt niet alleen in de fase van toetsconstructie, maar ook op de data die worden verzameld wanneer de definitieve toets bij alle leerlingen (of een grote representatieve steekproef daaruit) wordt afgenomen. Deze kalibratie heeft niet de bedoeling om het constructie- en keuzep proces te ondersteunen, maar dient vooral om inzicht geven in de kwaliteit van de definitieve toets. Daarbij gaan we voor alle opgaven en voor een bepaald onderdeel als geheel na, hoe het gesteld is met de model-fit. Met deze analyses gaan we na in hoeverre we de kalibratie geslaagd mogen noemen. Kalibratiegegevens rapporteren we uitsluitend op data, zoals die (in 2016) zijn verzameld met de definitieve versie van de toets. De kalibratieresultaten van de proeftoetsing laten we hier onbesproken omdat ze van ondergeschikt belang zijn.

Met kalibratie wordt bedoeld dat we kengetallen zoeken bij de items die de antwoorden van de leerlingen goed representeren. Hoe de kengetallen gezocht worden ligt deels vast door het gekozen model (zie paragraaf 2.4.2.2) en hoe succesvol deze operatie is kan statistisch getoetst worden. Eenvoudig gezegd, schatten we in OPLM met de CML-methode de itemparameters en controleren we of deze de data goed voorspellen. Voor een exacte beschrijving van de statistische toetsen die in OPLM gebruikt worden, hun eigenschappen en feitelijke implementatie in OPLM verwijzen we naar Verhelst (1993). Hier beperken we ons tot een korte beschrijving van de principes van de statistische toetsen die gebruikt zijn in de kalibratie-procedure. Daarbij geldt dat de kalibratieprocedure in principe per onderscheiden vaardigheid (schaal, opgavenbank) wordt toegepast.

De statistische toetsen in OPLM hebben goede statistische en asymptotische eigenschappen daar OPLM behoort tot de exponentiële familie, met de gewogen somscore,

$$s = \sum_{i=1}^k a_i x_i, \quad (4.1)$$

als een 'afdoende statistiek' voor de vaardigheid θ . Dit betekent dat alle informatie in de data met betrekking tot de vaardigheid in deze statistiek aanwezig is. Hiervan wordt gebruik gemaakt bij de statistische toetsen in OPLM. Het basisprincipe van de statistische toetsen in OPLM is dat op grond van de afdoende statistiek s de personen in de data kunnen worden gegroepeerd.

En binnen deze groepen kan de verwachte proportie goede antwoorden op een item onder het model, $p(+|s)$, vergeleken worden met de feitelijk geobserveerde proportie goede antwoorden, $prop(+|s)$. Via de basisvergelijking van OPLM kunnen we eenvoudig de conditionele kans op het goed beantwoorden van de items afleiden en daarmee kunnen we $p(+|s)$ evalueren, $prop(+|s)$ volgt uit de data. Discrepancies tussen $p(+|s)$ en $prop(+|s)$ duiden op schendingen van het model. Deze discrepanties vormen de basis

voor de diverse statistische toetsen in OPLM. De toetsingsgrootheid voor de veronderstelde discriminatie-indices is gegeven door

$$M = f_{s \in H}(p(+|s) - prop(+|s)) + f_{s \in L}(prop(+|s) - p(+|s)). \quad (4.2)$$

Deze zogeheten M-toetsen verdelen de scoregroepen in een laag deel (L) en een hoog deel (H) en f is een monotone functie (indeling in een groter aantal groepen is ook mogelijk). M-toetsen hebben een duidelijke interpretatie: is M significant positief dan is de veronderstelde steilheid van de ICC (*item characteristic curve*) overschat in het model, is M daarentegen erg laag dan is de index te klein. Verhelst laat zien voor welke functie, f , $M \approx N(0,1)$. In OPLM zijn drie verschillende M-toetsen geïmplementeerd die verschillen in de definitie van de hoge en lage scoregroepen.

Naast deze M-toetsen is er een algemene itemtoets die de volgende vorm heeft

$$S = f(p(+|s) - prop(+|s)).$$

Deze zogenaamde S-toets heeft een χ^2 verdeling onder het model. Analoog hieraan is er ook een toets om vormen van vraagonzuiverheid (in het Engels 'itembias' of 'differential item functioning', afgekort DIF) op te sporen:

$$S = h(p_I(+|s) - prop_I(+|s), (p_{II}(+|s) - prop_{II}(+|s)),$$

waarbij I en II de twee niveaus van de variabele indiceren waarvoor we de bias onderzoeken.

Als globale model toets is de R1c-toets (Glas, 1988) geschikt. Ook de distributie van alle afzonderlijke S-toetsen komt hiervoor in aanmerking. Als we deze S-toetsen opvatten als onafhankelijk, wat ze strikt genomen niet zijn, dan zouden de overschrijdingskansen uniform verdeeld moeten zijn op het (0,1) interval. Kortom, als we verder afzien van de formeel-statistische achtergrond van de gehanteerde toetsen, kan de kalibratieprocedure als volgt worden samengevat:

- Met behulp van het programma OPCAT stellen we de discriminatie-indices in OPLM in en hercoderen we indien noodzakelijk de antwoordcategorieën in de data.
- Vervolgens schatten we de itemparameters met behulp van de CML-methode.
- Met behulp van de M-toetsen controleren we of de discriminatie-indices goed zijn ingesteld.
- Een volgende controle betreft de overschrijdingskansen van de S-toetsen en een grafische modelcontrole door middel van het programma OPDRAW (grafische inspectie van de ICC's).
- Vervolgens vindt een globale modelcontrole plaats in de vorm van een R1c-toets en de verdeling van de overschrijdingskansen van de S-toetsen.
- Daarna toetsen we nadrukkelijk op vormen van itembias.

Bovenstaande stappen worden een aantal malen doorlopen tot het resultaat bevredigend is. Ook inhoudelijke overwegingen (zie hiervoor hoofdstuk 3 over de achtergronden van de toetsinhoud) spelen een rol in dit beslissingsproces.

In de Entreetoets gaat het om de volgende onderdelen en bijbehorende vaardigheidsschalen:

Module Basis en module Verdieping

- Rekenen
- Lezen
- Taalverzorging

Module Verbreding

Met betrekking tot taal (semantisch):

- Luisteren
- Schrijven
- Woordenschat

Met betrekking tot Wereldoriëntatie:

- Aardrijkskunde
- Geschiedenis
- Natuur en techniek

De genoemde negen kalibraties worden in paragraaf 4.1.2 geëvalueerd. De onderdelen van de Basismodule (Rekenen, Lezen en Taalverzorging) die ieder bestaan uit meer dan één domein worden gekalibreerd alsof ze één onderdeel vormen waarvan alle opgaven één schaal representeren. Voor de basistoets en de basistoets met demodule Verdieping van deze onderdelen wordt een aparte kalibratie uitgevoerd.

4.1.2 Het welslagen van de kalibraties: passing van de opgaven; model-fit

In deze paragraaf bespreken we de vraag of de kalibratie van de toetsonderdelen in de Entreetoets groep 7 een bevredigend beeld laat zien. Dit doen we zowel voor afzonderlijke opgaven (binnen een onderdeel), als voor toetsonderdelen als geheel.

Alle gerapporteerde waarden vormen het resultaat van analyses op een representatieve steekproef van 5400 leerlingen uit de afnamedata. Er is gekozen voor kalibratiebestanden van steeds $N = 5400$ omdat deze aan de ene kant groot genoeg zijn om nauwkeurige kalibratieresultaten te verkrijgen en aan de andere kant in de praktijk voldoende handzaam. Daarnaast heeft een grote N voor veel power en veel significante toetsen, die daardoor weinig informatief zijn.

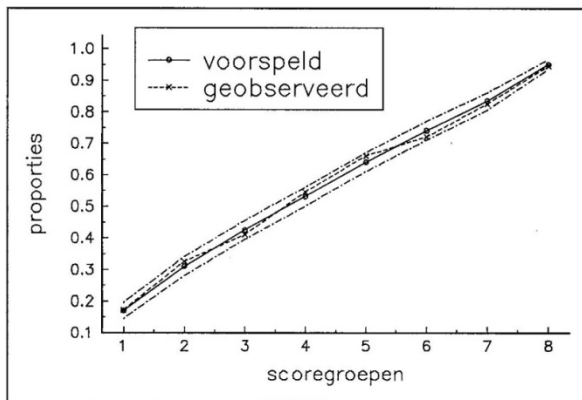
Passing van de afzonderlijke opgaven

Kalibratie van individuele opgaven vormt de ruggengraat van het psychometrische constructieproces. Het belangrijkste hulpmiddel daarbij is de itemkarakteristieke curve als grafische representatie van de itemkenmerken en de zogeheten S-toetsen.

Het is niet eenvoudig om de kwaliteit van de kalibratie aan te tonen. Het lastige is, dat de toetsing voor een groot deel visueel gebeurt. Dit kunnen we illustreren aan de hand van figuur 4.1 (ontleend aan Staphorsius, 1994, blz. 239). Figuur 4.1 beeldt voor een opgave de gegevens af waarop de betreffende S-toets gebaseerd is (zie handleiding OPLM: Verhelst, 1992). Voor deze toetsing wordt de totale groep van leerlingen die een verzameling opgaven gemaakt heeft, ingedeeld in een aantal scoregroepen (meestal acht; zie horizontale as). Elke groep bestaat uit leerlingen met een ongeveer even hoge score.

De geobserveerde proporties juiste antwoorden van deze groepen (telkens gesymboliseerd door een x) zijn door de middelste stippellijn verbonden. De volle lijn daarentegen verbindt de proporties die op grond van de parameterschattingen voorspeld kunnen worden. De twee buitenste lijnen geven het 95%-betrouwbaarheidsinterval aan. De breedte van dit interval is afhankelijk van het aantal leerlingen dat de opgave heeft beantwoord. Uit de figuur blijkt heel duidelijk dat de geobserveerde proporties, zoals bedoeld, binnen het 95%-betrouwbaarheidsinterval van de (geschatte) voorspelde proporties liggen, en dit komt in grote lijnen overeen met een niet-significante S-toetsingsgrootte (Verhelst, et al., 1995). Overigens zijn dit soort toetsingen op significantie bij zeer grote steekproeven weinig zinvol. Vanwege de grote statistische power zijn S-waarden in zo'n geval ook bij geringe, weinig zeggende afwijkingen al gauw significant.

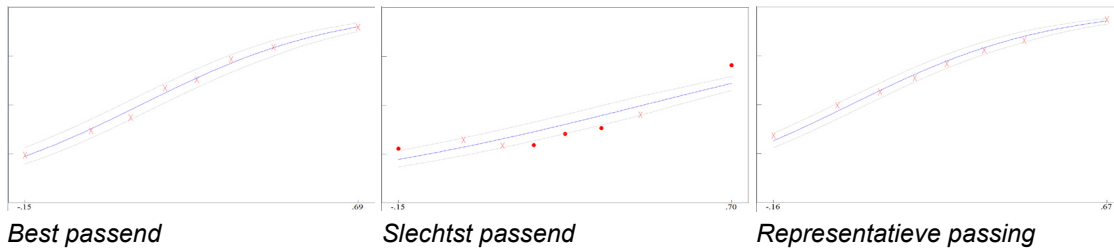
Figuur 4.1 Grafische voorstelling van een S-toets



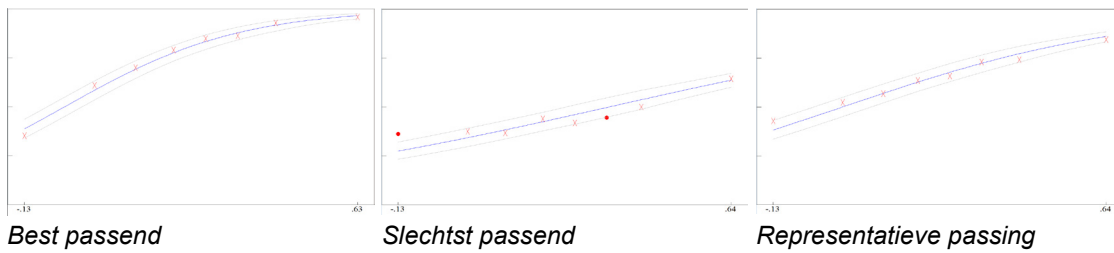
In deze verantwoording beperken we ons steeds per toetsonderdeel tot de opgave met de slechtste en de beste S-passing, aangevuld met een qua S-toetsingsresultaat representatieve passing. Met representatief bedoelen we hier een opgave die een goed voorbeeld is van de passing in het model zoals deze bij de meeste opgaven voorkomt. De voorbeelden in figuur 4.2 (a, b en c) illustreren dat zelfs bij de slechtst passende opgaven sprake is van een zeer aanvaardbaar beeld. Er wordt in deze gevallen voor een deel (van de onderscheiden scoregroepen) niet beantwoord aan de eis dat de geobserveerde proportie binnen het 95%- betrouwbaarheidsinterval van de geschatte proporties ligt. Dit beeld doet zich per toetsonderdeel slechts bij enkele opgaven voor die dan ook een uitzondering vormen. De overige opgaven voldoen voor alle scoregroepen wel aan die eis. Dit is een sterke aanwijzing dat het meetinstrument in combinatie met het meetmodel adequaat zijn om het gedrag van de leerlingen te verklaren. Bovendien blijkt dat gemeten verschillen in gedrag tussen de leerlingen per onderscheiden vaardigheid te verklaren zijn door één unidimensioneel concept.

Figuur 4.2a Voorbeelden van S-toetsen voor de Entreetoets groep 7 (afname 2016) met per onderdeel de best passende, de slechtst passende en een qua passing representatieve opgave: Module Basis en Module (Basis en) Verdieping

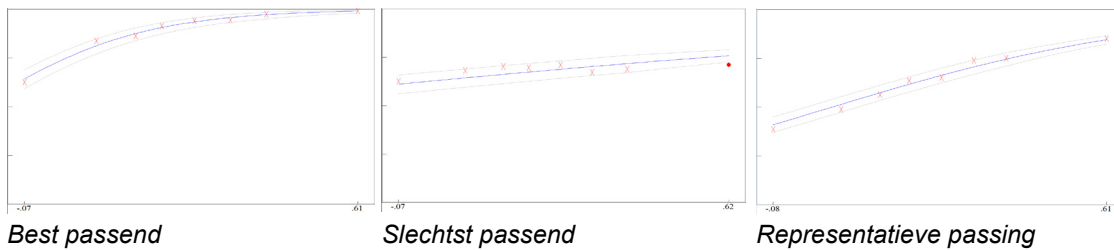
Module Basis en (Basis en) Verdieping: Rekenen



Module Basis en (Basis en) Verdieping: Lezen

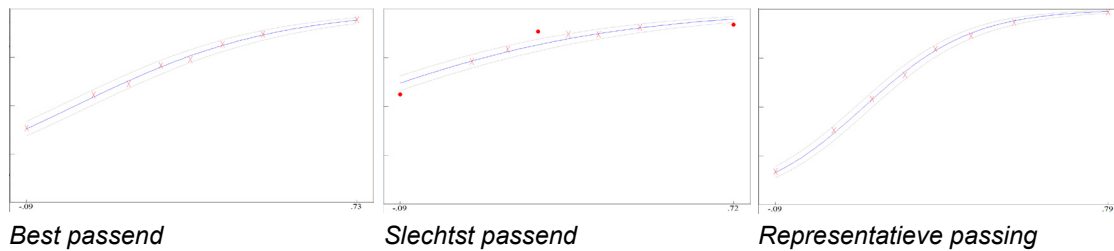


Module Basis en (Basis en) Verdieping: Taalverzorging

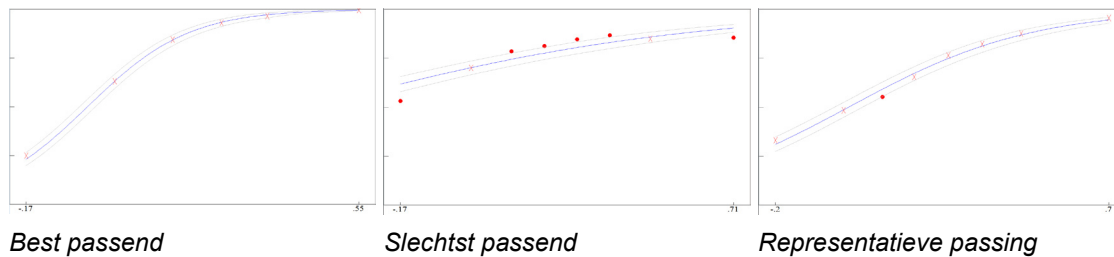


Figuur 4.2b Voorbeelden van S-toetsen voor de Entreetoets groep 7 (afname 2016) met per onderdeel de best passende, de slechtst passende en een qua passing representatieve opgave: Module Verbreding onderdelen taal

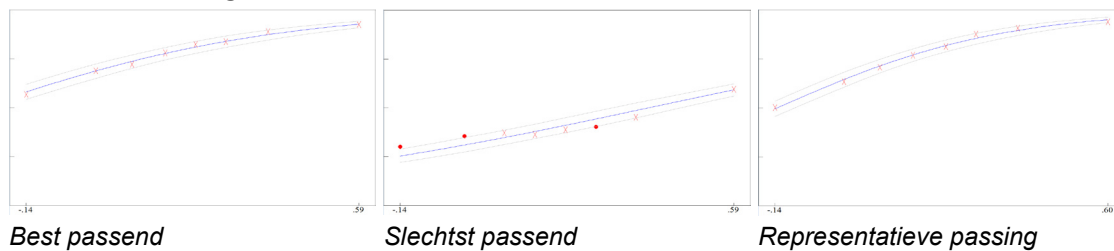
Module Verbreding taal: Luisteren



Module Verbreding taal: Schrijven



Module Verbreding taal: Woordenschat

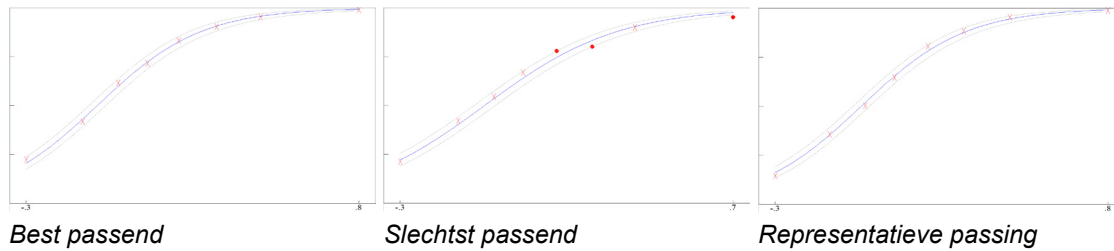


Een andere methode om op het niveau van de individuele opgaven de modelpassing te verantwoorden, is de aanpak die wordt besproken in het COTAN Beoordelingssysteem (Evers et al., 2010). Het betreft hier een poging om de nauwkeurigheid van de itemparameterschattingen te beoordelen op basis van een constante (in het COTAN Beoordelingssysteem met 'c' aangeduid) die weergeeft hoe de relatie is tussen de standaardfout van de moeilijkheidsparameter van een opgave en de standaarddeviatie van de vaardigheidsverdeling van de kalibratiepopulatie. Het beoordelingssysteem geeft ook richtlijnen voor het beoordelen van de grootte van deze 'c'. Deze dient te worden beoordeeld als goed als de waarde lager is dan of gelijk aan .20. Waarden tussen .30 en .40 gelden als voldoende.

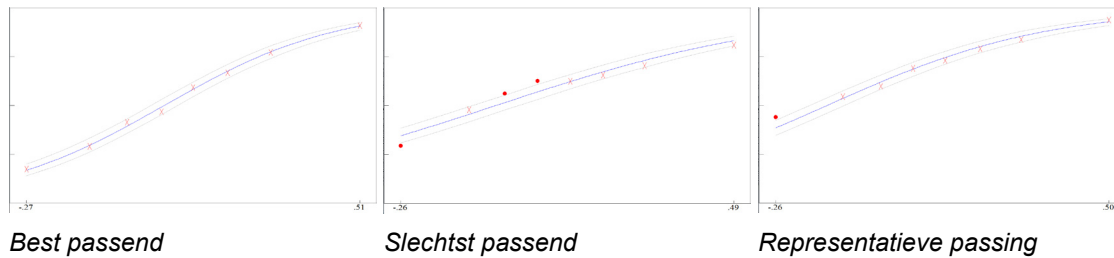
De waarden voor deze constante zijn weergegeven in tabel 4.1 voor elk van de geanalyseerde onderdelen. De opgaven in de module Basis staan gepresenteerd in het linkerdeel van de tabel. De opgaven in de module Verdieping staan gepresenteerd in het rechterdeel van de tabel. Daarnaast zijn ook de onderdelen van de module Verbreding in de tabel opgenomen.

Figuur 4.2c Voorbeelden van S-toetsen voor de Entreetoets groep 7 (afname 2016) met per onderdeel de best passende, de slechtst passende en een qua passing representatieve opgave: Module Verbreding onderdelen Wereldoriëntatie

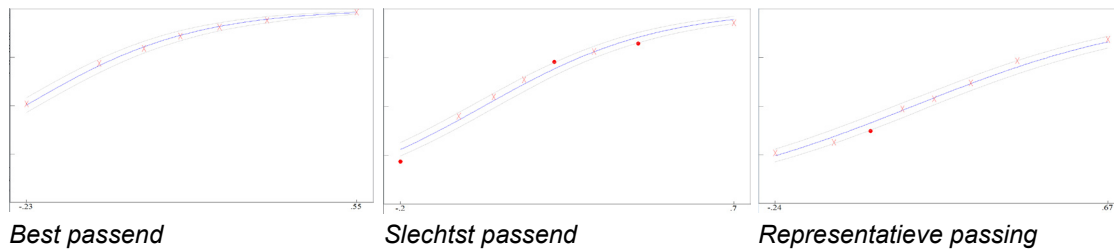
Module Verbreding Wereldoriëntatie-Aardrijkskunde



Module Verbreding Wereldoriëntatie-Geschiedenis



Module Verbreding Wereldoriëntatie-Natuur en techniek



Tabel 4.1 Nauwkeurigheid van de itemparameterschattingen (constante 'c') voor de gekalibreerde onderdelen van de Entreetoets groep 7

Toets(onderdeel)	Constante 'c'			
	Module Basis		Module (Basis en) verdieping	
	Range	Gemiddelde	Range	Gemiddelde
Basis/Verdieping				
- Rekenen	.011 - .032	.016	.011 - .050	.019
- Lezen	.012 - .033	.018	.012 - .045	.023
- Taalverzorging	.014 - .056	.021	.014 - .098	.027
Verbreding				
<i>Talige onderdelen</i>				
- Luisteren	.019 - .098	.040		
- Schrijven	.016 - .077	.028		
- Woordenschat	.019 - .043	.031		
<i>Wereldoriëntatie</i>				
- Aardrijkskunde	.016 - .040	.028		
- Geschiedenis	.021 - .099	.037		
- Natuur en techniek	.017 - .042	.029		

De gemiddelde waarden van de constante zijn uitstekend te noemen. Ze zijn alle duidelijk lager dan .10 en variëren tussen .016 (Rekenen Basis) en .045 (Wereldoriëntatie). Voor geen enkele opgave is c groter dan .20; de hoogste aangetroffen waarde voor een afzonderlijke opgave is .161. De hoogste waarden hebben betrekking op Wereldoriëntatie wanneer deze als volledige schaal wordt geschaald. In werkelijkheid zijn de domeinen van Wereldoriëntatie echter afzonderlijk gekalibreerd. Wanneer we uitgaan van de afzonderlijke subschalen voor Aardrijkskunde, Geschiedenis en Natuur en techniek (en dus Wereldoriëntatie als volledige schaal buiten beschouwing laten), is de hoogste gemiddelde 'c'-waarde .04 (voor Luisteren) en de hoogste waarde voor een afzonderlijke opgave .099 (voor Geschiedenis). Alle waarden voldoen ruimschoots aan de door de COTAN gestelde minimumeisen.

Model-fit voor de afzonderlijke toetsonderdelen

We beperken ons niet tot het evalueren van de afzonderlijke opgaven, maar kijken ook naar de verzameling opgaven die een toetsonderdeel vormen. In dit kader kijken we naar kwantitatieve gegevens over bijvoorbeeld de passing van de opgaven en de toets als geheel (model-fit).

In het navolgende zullen we deze kwantitatieve informatie presenteren en bespreken. In dat kader is het volgende van belang:

- Er is gekalibreerd aan de hand van willekeurige steekproeven ($N=5400$) uit het totaalbestand van afnamegegevens van het jaar 2016. Deze steekproeven werden bij elk van de kalibraties die we hier onderscheiden opnieuw getrokken.
- De omvang van de steekproeven is meer dan voldoende om adequaat te kunnen kalibreren. Wanneer het echter gaat om het statistische toetsen (bijvoorbeeld de al genoemde S-toetsen), dan zijn de steekproeven in feite te groot. Immers, zelfs zeer kleine, minimale afwijkingen van de gewenste situatie leiden dan al tot significante toetsingsresultaten. Zelfs als de lijn die de geobserveerde proporties representeert (zie figuur 4.1) geheel binnen het 95%-betrouwbaarheidsinterval ligt, kunnen de toetsingsresultaten significant zijn. In dit geval heeft het weinig zin om conclusies aan deze resultaten te verbinden.
- Om toch zinvolle resultaten te kunnen presenteren op (onder meer) de S-toetsen is deze steekproef verdeeld in 10 sub-datasets (van $N = 540$) uit het betreffende kalibratiebestand en heeft er tien keer een vergelijkbare toetsing plaatsgevonden. Alle leerlingen in het bestand zijn in (één van) de deelsteekproeven opgenomen.

Wat betreft de S-toetsen die steeds per toetsonderdeel en in elk van de tien (deel-)steekproeven zijn uitgevoerd, zijn we vooral geïnteresseerd in de distributie van de overschrijdingskansen van deze verzameling toetsingsresultaten. Als we de S-toetsen opvatten als onafhankelijk, wat ze strikt genomen niet zijn, dan zouden de overschrijdingskansen uniform verdeeld moeten zijn binnen het $(0,1)$ interval, met zo weinig mogelijk significante resultaten. Deze laatste mogen uiteraard voorkomen (dat is nu eenmaal het beginsel van statistische toetsing), maar niet heel veel vaker dan op basis van toeval te verwachten is. Zijn er relatief veel significante toetsingsresultaten, dan duidt dit op de aanwezigheid van (te) veel slecht passende opgaven.

Tabellen voor alle kalibraties afzonderlijk (en voor Rekenen, Lezen en Taalverzorging dubbel) zijn opgenomen in bijlage 1. Elke regel in elke tabel heeft betrekking op een van de tien deelsteekproeven (zie hierboven). We hebben geprobeerd de uitkomsten van alle S-toetsingen samen te vatten in tabel 4.2. Daarbij hebben we per onderdeel steeds de twee rijen uit de bijlage-tabellen vermeld die horen bij de twee mediane R1c-waarden (over R1c zo dadelijk meer). De lezer kan de resultaten van *alle* toetsingen raadplegen in de bijlage.

Tabel 4.2 *Entreetoets groep 7: verdeling van overschrijdingskansen voor S-toetsen bij tien aselecte deelsteekproeven uit de 2016-afnamedata (representatief bestand met N=5400) voor negen toetsonderdelen waarvan drie dubbel (Basis en Verdieping)*

Toetsonderdeel / vaardigheid	Aantal opgaven	0.-/---/----.1-----.2-----.3-----.4-----.5-----.6-----.7-----.8-----.9----1.											
		[0,1] interval van p-waarden											
Module Basis		180											
Rekenen	90	2 1	4 2	4 5	8 9	7 11	9 11	6 10	10 8	10 12	14 6	10 8	6 7
Lezen	40	1 0	3 1	0 1	4 6	3 3	6 6	3 4	4 2	3 5	4 4	3 4	6 4
Taalverzorging	50	1 1	0 2	1 4	2 3	8 6	4 2	3 6	4 7	6 3	9 8	4 5	8 3
Module Verdieping		310											
Rekenen	140	0 0	7 5	6 7	11 13	15 21	6 9	17 17	14 15	21 17	15 11	16 14	12 11
Lezen	80	0 0	2 2	2 8	8 8	9 4	8 9	12 6	7 10	6 11	13 8	7 8	6 6
Taalverzorging	90	1 1	2 3	4 4	6 4	16 11	4 6	7 13	9 11	11 6	6 8	9 14	15 9
Module Verbreding		160											
<i>Taalonderdelen</i>		100											
Luisteren	30	1 1	0 1	1 0	2 3	4 6	1 3	3 2	0 1	4 2	4 2	5 3	5 6
Schrijven	30	1 0	0 1	2 1	3 4	2 2	3 2	7 1	1 4	4 4	4 6	1 3	2 2
Woordenschat	40	0 0	0 2	1 4	7 7	6 1	5 2	3 2	5 3	1 1	3 5	6 7	3 6
<i>Wereldoriëntatie</i>		60											
Aardrijkskunde	20	0 0	0 2	0 0	0 1	3 1	4 4	2 1	3 2	2 4	1 2	2 2	3 1
Geschiedenis	20	0 0	2 0	0 0	2 4	2 3	4 2	0 0	2 0	0 3	0 4	7 2	1 2
Natuur en techniek	20	0 0	2 0	2 1	3 2	2 3	0 1	0 1	3 3	1 2	1 3	1 2	5 2

Per onderdeel zijn steeds de twee rijen opgenomen die horen bij de beide mediane R1c-waarden (toelichting in tekst)

We bespreken eerst de resultaten voor Rekenen, omdat de situatie daar relatief eenvoudig is: bij het onderdeel Rekenen zijn alle opgaven, ongeacht het domein (dat wil zeggen Getallen, Meten en meetkunde, Verbanden en Verhoudingen) op één en dezelfde vaardigheidsschaal gekalibreerd. Bijlage 1 waarin het (0,1) interval steeds is opgedeeld in tien gelijke secties, geeft – in termen van aantallen p-waarden per sectie – een beeld van de uitkomsten van de S-toetsen voor alle opgaven van de deeltoets per steekproef (90 opgaven voor de basistoets, 140 opgaven voor de basis en module Verdieping samen). Daarnaast is aangegeven in hoeveel gevallen de overschrijdingskans kleiner was dan .01 (donker gearceerde kolom), respectievelijk kleiner dan .05 (licht en donker gearceerde kolom samen). De totale aantallen toetsingen per rij (dus per steekproef) zijn in principe gelijk (i.e. gelijk aan het aantal opgaven), maar af en toe was toetsing

voor een enkele opgave niet mogelijk. Het is duidelijk dat in alle steekproeven voor Rekenen de verdeling redelijk gelijkmatig is over het gehele interval van overschrijdingskansen. Deze resultaten geven bovendien een bevestiging van het eerder geschetste beeld, dat met uitzondering van enkele opgaven, sprake is van niet-significante S-toetsen. Bij Rekenen zijn 134 van de 2300 toetsingen significant op het 5%-niveau (dat is 5.82% van de toetsingen, waarvan 24 (1.02%) significant op het 1%-niveau). Dit betekent dat de rechteroverschrijdingskansen van de S-toetsen een beeld vertonen dat overeenkomt met wat we op basis van het model mogen verwachten. Deze resultaten vormen een kwantitatieve ondersteuning van de conclusie dat de opgaven van het onderdeel Rekenen, in een unidimensionaal construct representeren.

Op soortgelijke wijze zijn de resultaten van de S-toetsen weergegeven voor respectievelijk Lezen en Taalverzorging, waarbij de opgaven van domeinen (Begrijpend lezen, Opzoeken en Samenvatten, respectievelijk Spelling niet-werkwoorden, werkwoorden, Interpunctie en Grammatica) telkens op één schaal zijn gekalibreerd. Voor de module Verbreding geldt dat per domein is gekalibreerd. Dat wil zeggen voor Luisteren, Schrijven, Woordenschat, Aardrijkskunde, Geschiedenis en Natuur en techniek. Omdat de aantallen opgaven, respectievelijk S-toetsen, per domeinen kalibratie bij enkele domeinen betrekkelijk klein zijn (bijvoorbeeld steeds slechts 20 opgaven voor Aardrijkskunde, Geschiedenis en Natuur en techniek), is het soms lastig om de gelijkmatigheid van verdelingen goed in beeld te brengen. Dit neemt niet weg dat de conclusies vergelijkbaar zijn met die voor Rekenen: de aantallen p-waarden zijn vrij gelijkmatig verdeeld over het hele interval en de aantallen significante S-toetsen zijn gering en ongeveer in overeenstemming met wat men op basis van het model zou mogen verwachten. Per domein ligt het aantal significante toetsingen (op het 1%-niveau) tussen 0,75% en 1,5%; voor het 5%-niveau varieert dat percentage tussen 3,75% en 6,0% met het onderdeel Geschiedenis als enige uitzondering (9% significante toetsingen).

Ook de samenvattende resultaten van deze analyses in tabel 4.2 vertonen een bevredigend beeld. In de tabel zijn de resultaten van in totaal 1300 S-toetsen afgebeeld. In de laagste van de tien secties waarin het [0,1]-interval is verdeeld, vallen 112 p-waarden (i.e. waarden $< .10$), Dat is 8,6% van de toetsingen. De andere toetsingsresultaten zijn mooi verdeeld over de overige secties van het interval; de percentages variëren tussen 9.2% en 11.5% per sectie. Er zijn 54 toetsingen significant op het 5%-niveau (4.2% van de gevallen) en elf (0,8%) op het 1%-niveau. Dat is minder dan men op basis van het model zou kunnen verwachten.

Hoe het is gesteld met de modelpassing voor de onderdelen en de domeinen van alle modules, is te zien in tabel 4.3. In deze tabel zijn de R1c-waarden weergegeven voor dezelfde steekproeven waarvoor in tabel 4.2 de resultaten van de S-toetsen zijn weergegeven. R1c is een statistiek die zicht geeft op de modelpassing van een toets (of een toetsonderdeel) als geheel. Ook voor deze statistiek was het nodig om ons te baseren op kleinere steekproeven (met adequate statistische power) om zinnige resultaten te verkrijgen. Gegevens over de R1c-toetsingen voor alle tien de steekproeven per onderdeel zijn opgenomen in bijlage 1. Voor een acceptabele model-fit geldt als vuistregel dat R1c bij voorkeur niet significant zou moeten zijn en niet groter dan ongeveer anderhalf maal het aantal vrijheidsgraden.

Het is duidelijk dat de modelpassing voor alle onderdelen prima is. In alle tien steekproeven per onderdeel/domein (ook de steekproeven met de ongunstigste uitkomsten voor R1c) is R1c voor alle onderzochte toetsonderdelen kleiner dan anderhalf maal het aantal vrijheidsgraden, dit met één uitzondering. Bij het onderdeel Geschiedenis is er een R1c-waarde die met een grootte van 1.8 maal het aantal vrijheidsgraden net niet aan de vuistregel voldoet. Bij de onderdelen van de modulen Basis en (Basis en) verdieping is daarnaast geen enkel toetsingsresultaat significant. Alleen bij de onderdelen van de module Verbreding vinden we een significante R1c-toets bij Geschiedenis (in het hierboven al genoemde geval ($p < .01$) en een significante R1c-toets bij Woordenschat ($p < .05$)). Er zijn in totaal 120 statistische toetsen uitgevoerd. Bij dit grote aantal toetsingen zijn dergelijke (geringe) aantallen significante resultaten te verwachten; we besteden er dan ook verder geen aandacht aan.

Wanneer we onze conclusies baseren op de beide mediane R1c-waarden die per onderdeel in tabel 4.3 zijn vermeld, zijn deze nog eenduidiger: geen enkele R1c-waarde is hoger dan anderhalf maal het aantal vrijheidsgraden en geen enkele R1c-waarde is significant.

Eerder, bij de beschrijving van de kalibratieprocedure, gaven we aan dat in de laatste stap van het kalibratieproces DIF-analyses worden uitgevoerd. De resultaten hiervan bespreken we in hoofdstuk 6 onder de noemer begripsvaliditeit.

Tabel 4.3 Entreetoets groep 7: Model-fit (R1c) bij tien aselechte deelsteekproeven uit de 2016-data (representatief bestand met N=5400) voor negen onderdelen/domeinen waarvan drie dubbel (Basis en Verdieping)

Onderdeel/domein	Aantal opgaven	R1c	df	p
Module Basis	180			
Rekenen	90	260.995 266.214	267 267	.5947 .5050
Lezen	40	103.932 98.068	117 117	.8015 .8961
Taalverzorging	50	144.375 145.053	147 147	.5496 .5338
Module Verdieping	310			
Rekenen	140	781.651 782.946	773 773	.4082 .3956
Lezen	80	357.536 344.931	393 393	.8992 .9602
Taalverzorging	90	441.611 435.344	463 463	.7564 .8176
Module Verbreding	160			
<i>Taalonderdelen</i>	100			
Luisteren	30	77.907 87.028	87 87	.7466 .4790
Schrijven	30	88.809 86.314	87 87	.4259 .5006
Woordenschat	40	115.687 116.065	117 117	.5213 .5114
<i>Domeinen</i>	60			
<i>Wereldoriëntatie</i>				
Aardrijkskunde	20	49.590 48.481	57 57	.7465 .7818
Geschiedenis	20	54.438 60.813	57 57	.5718 .3403
Natuur en techniek	20	54.357 49.645	57 57	.5748 .7447

Per onderdeel/domein zijn steeds de twee rijen opgenomen die horen bij de beide mediane R1c-waarden (toelichting in tekst)

Conclusie

Op grond van de uitgevoerde analyses kan geconcludeerd worden dat alle hier gedocumenteerde kalibraties zowel op het niveau van de individuele opgaven als op het niveau van de onderdelen/domeinen geslaagd genoemd kunnen worden. Dit impliceert dat voor alle negen onderzochte onderdelen/domeinen

geldt dat zij per onderdeel/domein één onderliggende latente vaardigheid representeren. Voor de onderdelen Rekenen, Lezen en Taalverzorging kan men deze conclusie trekken voor zowel basistoets als voor de basistoets met de Module Verdieping. Genoemde onderdelen bestaan uit meerdere domeinen. Voor de vier domeinen van Rekenen (Getallen, Meten en meetkunde, Verhoudingen en Verbanden) zijn geen afzonderlijke kalibraties uitgevoerd. De correlaties tussen de betreffende subscores zijn dermate hoog (in groep 7 en 8 in de regel $> .90$) dat het niet zinnig is op afzonderlijke schalen te kalibreren (ook al bevat het Leerlingprofiel aparte scores voor deze domeinen als de module Verdieping is afgenomen). Ook voor Lezen en Taalverzorging worden de domeinen (Begrijpend lezen, Opzoeken en Samenvatten; respectievelijk Spelling niet-werkwoorden, Spelling werkwoorden, Interpunctie en Grammatica) apart gescoord en gerapporteerd op het Leerlingprofiel als naast de basistoets de module Verdieping is afgenomen. In dit geval zijn deze domeinen afzonderlijk geschaald. We hebben gemeend ervan te kunnen afzien om al deze kalibraties hier te presenteren omdat de kalibraties al op het niveau van Lezen en Taalverzorging succesvol zijn gebleken. De kalibratie op domeinen laat enkel nog betere resultaten zien dan de hier gepresenteerde (wat logisch is omdat deelvaardigheden leiden tot homogener deeltoetsen). Wie geïnteresseerd is in de itemkarakteristieken van de domeinen, verwijzen we naar tabel 3.28b. Geslaagde kalibraties beschouwen we – bij toepassing van een IRT model bij de operationalisatie van een construct (vaardigheid) – als een noodzakelijke voorwaarde voor begripsvaliditeit. De kalibratie-analyse, als puur formeel proces, kan daarmee nog geen uitspraken doen over de inhoudsvaliditeit of over de constructvaliditeit. In hoofdstuk 6 komen we hier nog uitgebreid op terug.

4.2 Relatieve normering: toekenning van percentielscores

Om een adequate relatieve normering te realiseren, is het nodig om een representatieve steekproef van voldoende omvang te trekken uit de doel- of referentiepopulatie. Daarbij heeft elk lid van de populatie idealiter dezelfde kans om in die steekproef terecht te komen. Op basis van deze steekproef kunnen uitspraken worden gedaan over de vraag hoe een bepaald toetsresultaat zich verhoudt tot de resultaten van leerlingen in de referentiepopulatie.

De populatie kan in dit geval omschreven worden als de verzameling leerlingen van leerjaar 7 bij het afsluiten van dat leerjaar, of anders gezegd, bij intrede in leerjaar 8. In het geval van de Entreetoets groep 7 is de eerste normering gebaseerd op de afname van de toets in 2016, dus bij leerlingen die in het schooljaar 2015-2016 deel uitmaakten van groep 7. In dat schooljaar werd de module Basis van de Entreetoets afgenomen bij 56.392 leerlingen. Dat is een groot deel van de totale populatie van kinderen in groep 7 van het reguliere basisonderwijs. Hoeveel dat er exact zijn, is niet te zeggen omdat noch het CBS, noch DUO gegevens verschaft over aantallen leerlingen per leerjaar. Voor een schatting zou men kunnen uitgaan van het totale aantal leerlingen in een leeftijdscohort, bijvoorbeeld alle elfjarigen. CBS gegevens (verworven via de site van DUO) leren ons dat op de peildatum van 1 oktober 2015, 185.658 elfjarigen in het primair onderwijs verbleven, van wie 172.341 in het reguliere basisonderwijs. Afhankelijk van het gekozen uitgangsaantal en aannemende dat de omvang van een leerjaarcohort ongeveer gelijk is aan de omvang van een leeftijdscohort, zou men ruwweg kunnen stellen dat bijna een derde deel van de landelijke leerlingenpopulatie van groep 7-leerlingen aan de Entreetoets deelneemt. Het heeft weinig zin om exactere aantallen te schatten; het gaat om een globale schatting.

De groep van Entreetoetsgebruikers is te beschouwen als een, zeer grote, steekproef uit de totale doelpopulatie. Het is zeker dat dit in ieder geval geen aselechte steekproef is. Voor alle kinderen uit de doelgroep geldt immers dat zij niet per definitie dezelfde kans hadden om in deze steekproef terecht te komen. Alleen kinderen die op scholen zitten die deelnemen aan de Entreetoets hebben een kans, waarbij de scholen zichzelf hebben geselecteerd als gebruiker. De vraag is dan of in deze zelfselectie factoren een rol spelen die ervoor zorgen dat de groep van Entreetoetsgebruikers op een systematisch andere wijze is samengesteld als de totale doelpopulatie. Daarom is het interessant en belangrijk om inzicht te hebben in de verdelingskenmerken van de groep Cito Entreetoetsgebruikers en die van de totale doelgroep. Men kan dit nagaan op het niveau van de scholen en op het niveau van de leerlingen. In de volgende paragraaf zullen we daarom proberen antwoord te geven op de vraag of de groep kinderen bij wie de Entreetoets

is afgenomen in het schooljaar 2015-2016 op een aantal belangrijke kenmerken op dezelfde wijze is verdeeld als de gehele doelgroepopulatie (van kinderen in leerjaar 7 van het reguliere basisonderwijs in 2016).

4.2.1 Evaluatie van de representativiteit van het normeringsbestand

In deze paragraaf besteden we aandacht aan de kenmerken die normaliter relevant worden geacht om de relatie tussen normeringssteekproef en populatie in kaart te brengen. Het COTAN-beoordelingssysteem (Evers et al., 2010) geeft in dit opzicht aan dat de normeringssteekproef in ieder geval moet worden beschreven in termen van leeftijd, sekse, etniciteit en regio. Van de variabelen die daaraan volgens het beoordelingssysteem zouden kunnen worden toegevoegd, lijken urbanisatiegraad, sociaal-economische status en opleidingsniveau (in dit geval van de ouders) het meest in aanmerking te komen. We bespreken hieronder voor welke variabelen (of benaderingen daarvan) we gegevens ter beschikking hebben. De verdeling naar *regio* gaat uit van vier regioniveaus die overeenkomen met de volgende indeling naar provincies (CBS-indeling naar landsdeel):

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

De verdeling naar urbanisatiegraad of *mate van verstedelijking* is terug te voeren op de bij het CBS gebruikelijke indeling naar vijf niveaus, namelijk zeer sterk, sterk, matig, weinig en niet verstedelijkt. Verder kan er in dit verband ook gekeken worden naar schoolgrootte. Daarbij wordt onderscheid gemaakt tussen grote scholen (≥ 200 leerlingen) en kleine scholen (< 200 leerlingen).

Wat betreft *leeftijd* gaat het hier om een betrekkelijk homogene groep leerlingen die op een bepaald moment in groep 7 collectief (dat wil zeggen, per klas) de Entreetoets maakt. Daarbij is er sprake van de gebruikelijke spreiding van leeftijden over een schooljaar, waarbij sommige leerlingen voorlijk genoemd kunnen worden (bijvoorbeeld omdat ze ooit een klas hebben overgeslagen) en andere vertraagd (omdat ze bijvoorbeeld een of meerdere malen hebben gedoubleerd). Deze verdeling naar leeftijd zal worden beschreven zonder dat vergelijking met de totale doelgroepopulatie mogelijk is (omdat geregistreerde gegevens daarover niet beschikbaar zijn).

Wat betreft *etniciteit* is het voor de vergelijking een probleem dat geen eenduidige vergelijkingsgegevens voor steekproef, noch populatie bekend zijn. Voorheen kon er gewerkt worden met een bij Cito gebruikelijke indeling van scholen in groepen of strata op basis van een weging van de formatiegewichten van hun leerlingen op schoolniveau. Het formatiegewicht van de leerlingen werd gebruikt voor de bepaling van de formatieomvang van een school. Daarbij werden de leerlingen gecategoriseerd in vijf formatiegewichten die een combinatie vormden van opleidingsniveau, sociaal-economische status en etnische herkomst van de ouders. Inmiddels is echter de categorisering naar leerlinggewichten vervangen door een andere regeling waarbij nog slechts drie niveaus worden onderscheiden en waarbij etniciteit geen rol meer speelt. Voor een definitie van en toelichting op de gewichten en categorieën verwijzen we naar paragraaf 2.5.2.

Dit betekent dat we nog wel gebruik kunnen maken van gegevens over de *achterstandssituatie* van leerlingen in termen van het opleidingsniveau van de ouders (als indicator voor sociaal-economische status), maar dat een verwijzing naar etnische herkomst ontbreekt. Dit impliceert dat de analyses zich moeten beperken tot de achtergrondvariabele percentage achterstandsléerlingen op basis van leerlinggewicht (op schoolniveau). Op grond van peilingsonderzoek (PPON) is bekend dat dit kenmerk, samen met gegevens over regio en mate van verstedelijking, een vrij aardige benadering vormt van de achtergrondvariabele etnische herkomst (vergelijk Hemker, Kordes & van Weerden, 2011). We nemen daarom aan dat de normeringssteekproef voldoende representatief is naar etnische herkomst als de verdeling van de scholen naar mate van verstedelijking, regio en percentage gewichtenleerlingen overeenkomt met de verdeling in de landelijke populatie.

In het navolgende worden regio, mate van verstedelijking en schoolgrootte op schoolniveau besproken. Daarnaast worden de scholen gekarakteriseerd en vergeleken in termen van het percentage

achterstandsleerlingen (dat wil zeggen, het percentage leerlingen met een leerlinggewicht). De kenmerken sekse, leeftijd en achterstand (leerlinggewicht) worden op leerlingniveau aan de orde gesteld.

Schoolkenmerken

In tabel 4.4 is de verdeling gegeven van de Nederlandse populatie van basisscholen. De gegevens zijn gebaseerd op cijfers van CFI DUO (Dienst Uitvoering Onderwijs van het ministerie van OCW). Voor een aantal scholen ontbraken CFI-gegevens en/of gegevens over provincie en mate van verstedelijking. Deze scholen zijn verwijderd waarna de percentages voor de populatieverdeling konden worden bepaald. Er namen 2085 scholen deel aan de Entreetoets afname 2016.

Tabel 4.4 Verdeling van basisscholen naar regio, mate van verstedelijking, schoolgrootte en percentage achterstandsleerlingen voor de Nederlandse populatie en voor de gebruikers van de Entreetoets (afname 2016)

Regio (scholen)				Schoolgrootte (scholen)			
	Populatie (%)	Steekproef aantal	Steekproef (%)		Populatie (%)	Steekproef aantal	Steekproef (%)
Noord	14.6	129	6.2	klein (<=200)	51.0	994	48.7
Oost	24.8	617	29.6	groot (>200)	49.0	1048	51.3
West	42.1	851	40.8	Totaal	100	2042	100
Zuid	18.5	488	23.4				
Totaal	100	2085	100				
	chi²	df	p		chi²	df	p
chi ²	149.63	3	< .001	chi ²	4.35	1	0.04
phi	.268			phi	.046		

Mate van verstedelijking (scholen)				Percentage achterstandsleerlingen (scholen)			
	Populatie (%)	Steekproef aantal	Steekproef (%)		Populatie (%)	Steekproef aantal	Steekproef (%)
niet stedelijk	13.7	218	10.5	0 - 10%	66.6	1309	64.1
matig stedelijk	18.1	404	19.4	10% - 25%	22.9	483	23.7
weinig stedelijk	26.8	539	25.9	25% - 50%	9.3	217	10.6
sterk stedelijk	25.7	570	27.3	50% - 100%	1.2	33	1.6
zeer sterk stedelijk	15.6	354	17.0	Totaal	100	2042	100
Totaal	100	2085	100				
	chi²	df	p		chi²	df	p
chi ²	23.26	4	< .001	chi ²	8.77	3	0.03
phi	.106			phi	.066		

In de tabel zijn de verdelingen gegeven naar regio, schoolgrootte, mate van verstedelijking en percentage achterstandsleerlingen, voor zowel de populatie van basisscholen als voor de scholen die in 2016 aan de Entreetoets (aangeduid als ET7) deelnamen. Buitenlandse scholen en enkele scholen voor speciaal onderwijs werden uit het bestand verwijderd.

Het is duidelijk dat de scholen die in 2016 de Entreetoets groep 7 hebben afgenomen bij hun leerlingen een vrij nauwkeurige afspiegeling vormt van de totale populatie van basisscholen in Nederland. Dat geldt voor alle besproken achtergrondkenmerken. Er is weliswaar sprake van lichte afwijkingen van de

populatieverdelingen, die bij toetsing met behulp van de Chi-kwadraattoets significant zijn, maar significantie zegt bij deze grote aantallen niet veel. Wanneer we ons baseren op effectgroottes in termen van Phi, dan blijken deze nergens groter dan .106 (wat geldt als een klein en verwaarloosbaar effect), met uitzondering van regio. Voor dit kenmerk zijn de afwijkingen substantiëler: met name de scholen in de regio Noord zijn ondervertegenwoordigd ($\phi = .268$). Ook al lijkt regio op zichzelf niet een kenmerk dat impact heeft op de scores (en daarmee op de normering), zijn we voor alle zekerheid nagegaan of dit het geval is. Daarbij hebben we de effecten voor de onderscheiden groepen steeds per tweetal bepaald. De effecten bleken klein. Voor de regio's die het meest van elkaar verschilden varieerden de effecten tussen .13 (voor Rekenen basistoets & module Verdieping) en .22 (voor Taalverzorging basistoets & module Verdieping). Dit impliceert dat de effecten te kwalificeren zijn als verwaarloosbaar tot klein, zeker als men bedenkt dat alle andere vergelijkingen tussen regio's per score kleinere effecten laten zien dan de genoemde. Bovendien zijn per score niet steeds dezelfde regio's in het geding waar het de hoogste effecten betreft. Voor Rekenen en Lezen betreft het grootste effect de regio's Noord en Zuid. Voor Taalverzorging is dat het contrast tussen de regio's Noord en West. Wél laat de regio Noord steeds de laagste gemiddelde score zien, maar het aandeel scholen van deze regio in de populatie is het kleinst (in de populatie 14.6; in de steekproef 6.2). Ten slotte is het ook in theorie niet goed interpreteerbaar of het om systematische effecten die echt door regio zelf worden bepaald. Waarschijnlijk is het eerder het gecombineerde effect van andere variabelen (urbanisatiegraad; % leerlingen met een afwijkend leerlinggewicht op school) dat de verschillen bepaalt. Omdat het gaat om verwaarloosbare tot kleine effecten, die niet systematisch zijn en theoretisch niet duidelijk te interpreteren, vonden we statistische terugweging niet noodzakelijk.

Leerlingkenmerken

In aanvulling op de evaluatie van de representativiteit van de steekproef op schoolniveau doen we dat voor sekse en leeftijd ook op leerlingniveau. De populatieverdeling werd zo goed mogelijk beschreven of geschat op basis van beschikbare gegevens van CBS Statline en CFI DUO. Voor de schatting van de populatieverdeling werden steeds de referentiegegevens gebruikt die het meest van toepassing waren. Voor leeftijd zoals hier geoperationaliseerd (dat wil zeggen, naar de mate waarin de leerling als 'vroeg' dan wel 'vertraagd' kan worden gekarakteriseerd) is de verdeling in de populatie niet door genoemde instanties geregistreerd.

Sekse

Volgens tellingen voor het schooljaar 2011-2012 bestaat 49.7% van de populatie 11-jarigen in het reguliere basisonderwijs uit jongens en 50.3% uit meisjes (CvTE, in voorbereiding). Voor de groep deelnemers aan de Entreetoets 7 (die niet samenvalt met de groep van elfjarigen) treffen we een soortgelijke verdeling aan: 64170 (49.8%) jongens en 64627 (50.2%) meisjes. Bij toetsing blijkt Chi-kwadraat (1.95) niet significant met $\Phi = .006$ (zie tabel 4.5).

Leeftijd/leertijd

Op basis van de leeftijd van de kinderen is nagegaan welke kinderen op de peildatum van 1 oktober 2015 11 jaar of ouder waren. Van deze kinderen mag worden aangenomen dat zij langer dan op basis van hun leeftijd noodzakelijk was in de kleutergroep zijn gehouden en/of dat zij in het verdere verloop van hun schoolloopbaan zijn blijven zitten. Daarnaast hebben we leerlingen als 'voorlijk' gekwalificeerd volgens de criteria die ook bij de afname van de Centrale Eindtoets 2015 zijn gehanteerd (vergelijk de Terugblik en de Wetenschappelijke verantwoording, CvTE, 2016a, 2016b).

Tabel 4.5 Verdeling van leerlingen naar sekse en leeftijd (leertijd) voor de Nederlandse populatie en voor de deelnemers aan de Entreetoets (afname 2016)

Sekse (leerlingen)				
	Populatie		Steekproef	
	(%)	aantal	(%)	
jongen	49.7	28015	50.0	
meisje	50.3	28021	50.0	
Totaal	100	56036	100	
	chi²	df	p	
chi ²	1.95	1	ns	
phi	.006			

Leeftijd (leerlingen)				
	Populatie		Steekproef	
	(%)	aantal	(%)	
'voorlijk'	10.7	5956	10.7	
regulier	75.7	42310	75.7	
vertraagd	13.6	7642	13.7	
Totaal	100	55908	100	
	chi²	df	p	
chi ²	.313	2	ns	
phi	.002			

Achterstand (formatiegewicht; leerlingen)				
	Populatie		Steekproef	
	(%)	aantal	(%)	
0.0 (regulier)	89.9	44329	89.9	
0.30	5.4	2511	5.1	
1.20	4.7	2471	5.0	
Totaal	100	49315	100	
	chi²	df	p	
chi ²	17.60	2	< .001	
phi	.019			

De gegevens met betrekking tot de ET-afname in 2016 laten zien dat 7.642 leerlingen vertraging hebben opgelopen (13.6%) en dat 5.956 leerlingen (10.7%) 'vroeg' genoemd kunnen worden. Dat komt zowel overeen met schattingen die voor groep 8 worden gegeven (zie CvTE, 2016b), als met de werkelijke cijfers die bij de CET-afname werden vastgesteld. Laatstgenoemde hebben we als schatting voor de Entreetoets-populatie gehanteerd. De populatieschattingen en de gerealiseerde verdeling voor de ET7-afname ontlopen elkaar nauwelijks. Bij toetsing blijkt Chi-kwadraat (.313) niet significant met Phi = .002 (zie tabel 4.5).

Achterstand (formatiegewicht)

Voor de meeste leerlingen konden we beschikken over informatie met betrekking tot het leerlinggewicht. Voor ongeveer 12% van de leerlingen in het bestand ontbrak deze informatie. Op schoolniveau hebben we dit leerlingkenmerk al gedefinieerd in termen van het percentage achterstandskinderen op een school. Net als in de populatie heeft 89.9% van de leerlingen geen leerlinggewicht. Ongeveer 10% heeft dus een leerlinggewicht, waarbij in de steekproef het percentage 1.20-leerlingen wat groter is dan in de populatie, het percentage 0.30-leerlingen daarentegen wat kleiner. De afwijkingen zijn weliswaar significant (Chi-kwadraat = 17.60; $p < .001$), maar in termen van Phi (.019) verwaarloosbaar.

We kunnen dus concluderen dat onze steekproef ook op leerlingniveau een vrij nauwkeurige afspiegeling vormt van de populatie.

Andere overwegingen met betrekking tot representativiteit

In het bovenstaande is de representativiteit besproken in relatie tot een aantal (potentieel) relevante achtergrondvariabelen. Het feit dat de steekproefverdelingen in hoge mate de populatieverdelingen benaderen, sluit echter nog niet uit dat bepaalde subgroepen leerlingen systematisch over- of ondervertegenwoordigd zijn. In dit opzicht is het onderstaande relevant.

Voor speciale leerlingen zijn weliswaar geen afzonderlijke normen vastgesteld, maar de toetsresultaten van deze leerlingen worden geïnterpreteerd met behulp van de gemiddelde vaardigheidsscores voor leerlingen uit het reguliere onderwijs. Voor deze leerlingen gelden namelijk dezelfde kerndoelen als voor leerlingen in het basisonderwijs, met dien verstande dat leerlingen in het speciaal (basis)onderwijs meer tijd krijgen om de kerndoelen te bereiken. Deze leerlingen kunnen én moeten dus langs dezelfde meetlat gehouden worden als de 'reguliere' leerlingen. De leerlingen in het regulier basisonderwijs waarop de normering gebaseerd is, vormen daarmee ook voor de leerlingen in het speciaal (basis)onderwijs een correcte referentiegroep.

Daarnaast is bij de Entreetoets de mogelijkheid gecreëerd voor het afnemen van gemakkelijkere versies van de toets bij leerlingen die zo zwak zijn dat leerkrachten ze bij ontbreken van deze mogelijkheid waarschijnlijk niet zouden hebben laten deelnemen aan de Entreetoets 7 of aan specifieke onderdelen daarvan. Bij wijze van alternatief kunnen zij leerlingen met een grote leerachterstand (van naar schatting een jaar of meer) de Entreetoets voor groep 6 laten maken. Omdat beide Entreetoetsen (voor groep 6 en groep 7) bestaan uit onderdelen en opgaven die op één schaal zijn gebracht, kunnen de behaalde scores op de Entreetoets groep 6 worden omgezet in scores die de leerling zou hebben behaald wanneer deze de Entreetoets 7 zou hebben gemaakt (zie paragraaf 2.2). Deze getransformeerde scores zijn in het normeringsbestand meegenomen. De ervaring leert dat ongeveer .7% van de leerlingen de Entreetoets in zijn geheel op een lager niveau maakt.

Ten slotte merken we in dit verband nogmaals op dat er voor enkele groepen leerlingen met beperkingen voor wie de afname van de Entreetoets problemen zou kunnen opleveren, speciale versies van de toets zijn uitgebracht. Voor leerlingen met een visuele beperking of dyslexie zijn aangepaste versies beschikbaar in braille, zwart-wit, op vergroot A3-formaat en met auditieve ondersteuning op cd of daisy-cd.

Conclusie

Op basis van de hierboven gepresenteerde gegevens is de conclusie gerechtvaardigd dat de totale groep leerlingen die in 2016 heeft deelgenomen aan de Entreetoets groep 7 sterk lijkt op de totale populatie leerlingen die leerjaar 7 van het reguliere basisonderwijs bevolkt. Dat is af te leiden uit gegevens op schoolniveau (regio, mate van verstedelijking, schoolgrootte en percentage achterstandsleerlingen) en gegevens op leerlingniveau (sekse, leeftijd-leertijd en leerlinggewicht). Dit maakt de kans klein, dat zelfselectie van scholen de populatie van Entreetoetsgebruikers heeft gemaakt tot een specifieke subpopulatie die op relevante factoren afwijkt. Het feit dat de deelpopulatie van Entreetoetsgebruikers zo'n grote omvang heeft in verhouding tot de gehele populatie (een derde deel) draagt aan deze conclusie bij. Anders gezegd: de groep Entreetoetsgebruikers kan worden beschouwd als een (zeer grote) representatieve steekproef uit de totale populatie van leerlingen in groep 7 van het reguliere basisonderwijs.

5 Betrouwbaarheid en meetnauwkeurigheid

5.1 Betrouwbaarheid

Voor het schatten van de betrouwbaarheid van de Entreetoets en de onderdelen daarvan is gebruik gemaakt van methoden uit zowel de klassieke testtheorie als de item respons theorie. We rapporteren betrouwbaarheden met betrekking tot de toetsonderdelen die worden gerapporteerd op het leerlingrapport (zie figuur 4.3 en figuur 4.5).

Het betreft de volgende onderdelen, respectievelijk scores:

Voor de basistoets worden ruwe scores en percentielscores gerapporteerd voor de totaalscore (180 opgaven) en voor de onderdelen Rekenen (90 opgaven), Lezen (40 opgaven) en Taalverzorging (50 opgaven). Voor de domeinen worden geen deelscores gepresenteerd wanneer uitsluitend de basistoets is afgenomen. De aantallen opgaven zijn hiervoor te gering (en de betrouwbaarheden als gevolg hiervan te laag). De totaalscore wordt niet alleen gebruikt voor beschrijvende doeleinden, maar ook als basis voor een voorspelling omtrent het best passende brugklatype. We beoordelen dat als belangrijke informatie waarop de COTAN-criteria voor 'belangrijke, dan wel irreversibele beslissingen' van toepassing zijn. Voor de totaalscore geldt dat deze pas als goed beoordeeld kan worden als deze .90 of hoger (.80 voor voldoende). Voor alle overige scores geldt het criterium .80 voor een beoordeling als goed en .70 voor voldoende (zie Evers et al., 2010).

Wanneer naast de basistoets ook de module Verdieping is afgenomen wordt er geen totaalscore berekend over alle onderdelen en opgaven (wél de totaalscore voor de basistoets). Daarnaast worden ruwe scores en percentielscores berekend voor de onderdelen Rekenen (140 opgaven), Lezen (80 opgaven) en Taalverzorging (90 opgaven), evenals voor alle domeinen (met aantallen opgaven variërend van 20 tot 50). We hebben betrouwbaarheidsgegevens berekend voor al deze onderdelen en domeinen.

Voor de module Verbreding geldt dat scores worden berekend voor de onderdelen Luisteren (30 opgaven), Schrijven (30 opgaven) en Woordenschat (40 opgaven), alsmede voor de drie domeinen van Wereldoriëntatie (60 opgaven), te weten Aardrijkskunde, Geschiedenis en Natuur en techniek (elk 20 opgaven). Voor alle onderdelen en domeinen van Verbreding worden betrouwbaarheidsgegevens verstrekt.

In tabel 5.1 geven we informatie over de aantallen opgaven in de betreffende toetsonderdelen alsmede de Greatest Lower Bound (GLB). Deze kan beschouwd worden als een zuiverder schatter van de betrouwbaarheid dan Cronbachs alfa (die een lichte onderschatting van de betrouwbaarheid geeft). Er heeft geen test-hertest onderzoek plaatsgevonden. De afnamecontext van de Entreetoets Groep 7 leent zich daar niet goed voor. Het feit dat alle items echter OPLM-gekalibreerd zijn, maakt het mogelijk een hertest te simuleren. We hebben een dubbele afname gesimuleerd van 1.000.000 leerlingen. Daarbij hebben we enerzijds de vaardigheidsverdeling van alle leerlingen, anderzijds alle itemparameters als uitgangspunt genomen. Steeds is een bepaalde vaardigheid aselekt uit de verdeling genomen en zijn twee bij deze vaardigheid horende afnames gesimuleerd. Uiteindelijk is de correlatie tussen deze 1.000.000 dubbele (virtuele) afnames berekend. Men kan deze simulatie beschouwen als een test-hertest onderzoek onder ideale condities. De tweede toetsafname is immers volledig onafhankelijk van de eerste toetsafname en wordt niet beïnvloed door de kennis die de leerling mogelijk verworven heeft in de eerste toetsafname. Daarnaast is er geen sprake van invloed van een test-hertest interval: beide afnames worden gesimuleerd alsof zij op hetzelfde moment plaats zouden vinden. We hebben ons in de presentatie beperkt tot de totaalscore op de basistoets en de onderdelen van deze module, Rekenen, Lezen en Taalverzorging. In tabel 5.1 is verder ook informatie opgenomen over de standaardmeetfout. Deze is gebaseerd op de GLB. De standaardmeetfout is via de geëigende formules omgezet in betrouwbaarheidsintervallen. We presenteren gegevens over de breedte van zowel het 90%- als het 95%-betrouwbaarheidsinterval.

Tabel 5.1 *Betrouwbaarheden, standaardmeetfout (SE) en betrouwbaarheidsintervallen (90%- en 95%-) voor de Entreetoets groep 7*

Onderdeel/domein	Aantal opgaven	GLB	Test-Hertest**)	SE	BI 90%	BI 95%
Basis						
Totaal	180	.97	.95	7.04	±11.58	±13.79
Rekenen	90	.95	.94	3.41	±5.61	±6.68
Lezen	40	.86	.84	2.42	±3.98	±4.74
Taalverzorging	50	.87	.84	2.56	±4.21	±5.02
Basis en Verdieping						
Rekenen	140	.97		4.03	±6.63	±7.91
Verhoudingen	30	.89		2.07	±3.40	±4.06
Getallen	50	.91		2.57	±4.23	±5.04
Verbanden	25	.84		1.88	±3.09	±3.69
Meten	35	.86		2.28	±3.76	±4.47
Lezen	80	.93		3.29	±5.42	±6.46
Begrijpend lezen	40	.86		2.43	±3.99	±4.76
Samenvatten	20	.79		1.75	±2.88	±3.43
Opzoeken	20	.77		1.75	±2.87	±3.42
Taalverzorging	90	.93		3.22	±5.30	±6.32
Grammatica	20	.86		1.49	±2.45	±2.92
Interpunctie	20	.79		1.54	±2.54	±3.02
Spelling niet-werkwoorden	25	.76		1.80	±2.96	±3.53
Spelling werkwoorden	25	.77		1.95	±3.21	±3.83
Verbreiding						
Luisteren	30	.79		2.06	±3.39	±4.04
Schrijven	30	.83		2.12	±3.49	±4.16
Woordenschat	40	.81		2.42	±3.99	±4.75
Wereldoriëntatie	60	.90		2.53	±4.16	±4.96
Aardrijkskunde	20	.76		1.79	±2.95	±3.52
Geschiedenis	20	.70		1.86	±3.07	±3.66
Natuur en Techniek	20	.75		1.83	±3.01	±3.59

**) Op basis van een simulatiestudie met N=1.00.000

De betrouwbaarheid in termen van interne consistentie is met een GLB van .97 voor de totaalscore zeer hoog (aanmerkelijk hoger dan de .90 die we als voorwaarde stelden om van een goede betrouwbaarheid te kunnen spreken). De betrouwbaarheid is daarmee hoog genoeg om de voorspelling en het advies over het best passende brugklatype voldoende grond te geven. De betrouwbaarheden van de onderdelen binnen de

basistoets voldoen met waarden van .95 (Rekenen), .86 (Lezen) en .87 (Taalverzorging) ruimschoots aan de gestelde eisen. Deze conclusies gelden ook als we zouden baseren op de (gesimuleerde) test-hertestbetrouwbaarheden voor de genoemde scores. De waarden liggen enkele honderdsten lager dan de GLB-waarden.

Wanneer naast de basistoets ook de module Verdieping wordt afgenomen stijgen de betrouwbaarheden van de deelscores naar waarden tussen .93 (voor Lezen en Taalverzorging) en .97 (voor Rekenen). Dat is zeer hoog. Voor de subonderdelen variëren de betrouwbaarheidswaarden tussen .77 en .93. Ze zijn dus overeenkomstig de COTAN-criteria te kwalificeren als voldoende tot goed, waarbij moet worden aangetekend dat geen enkele waarde lager is dan .75.

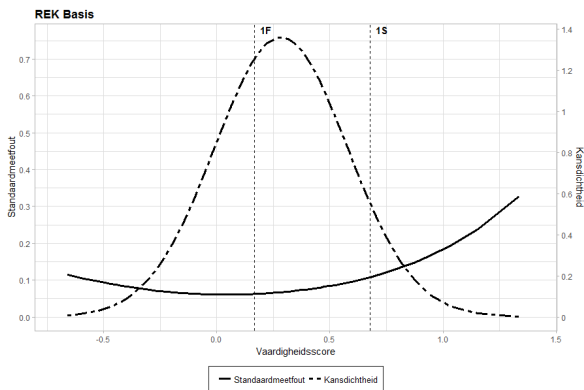
Men kan dezelfde conclusies trekken voor de taalonderdelen van de module Verbreding (waarden tussen .79 en .83). De GLB-waarden voor de domeinen van Wereldoriëntatie liggen tussen .70 en .76; het betrekkelijk geringe aantal opgaven (20 per onderdeel) zal hier aan hebben bijgedragen. De totaalscore Wereldoriëntatie over alle 60 opgaven kent een betrouwbaarheid van .90.

Van alle onderdelen/domeinen is de betrouwbaarheid groot genoeg om de leerling op een gedifferentieerde manier te kunnen beschrijven in een profiel van relatief sterke en zwakke punten.

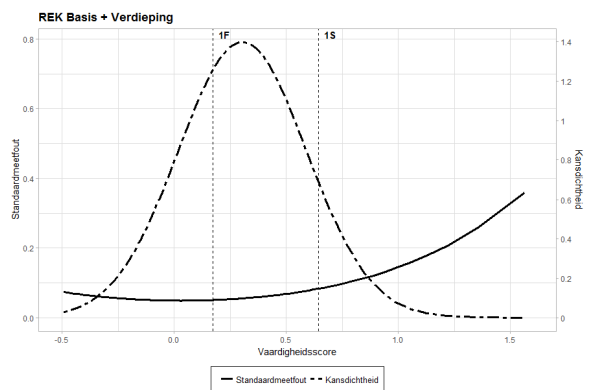
5.2 Nauwkeurigheid: lokale meetfouten

De in tabel 5.1 vermelde betrouwbaarheidscoëfficiënten en betrouwbaarheidsintervallen hebben alleen betrekking op de globale meetnauwkeurigheid van de toetsen. Omdat alle toetsonderdelen IRT-geschaald zijn, is het mogelijk ook iets te zeggen over de lokale meetnauwkeurigheid. We beperken ons hierbij tot de onderdelen Rekenen, Lezen en Taalverzorging die inzicht geven in al dan niet behaalde referentieniveaus. Dat doen we voor zowel de basistoets als voor de basistoets aangevuld met de module Verdieping. Figuur 5.1 laat zien hoe het gesteld is met de lokale meetnauwkeurigheid. In deze figuren waarin ook kansdichtheidfuncties zijn opgenomen, staat voor de genoemde toetsonderdelen de grootte van de meetfout afgebeeld. Hierbij past de kanttekening dat de standaardmeetfout onder meer afhankelijk is van het aantal opgaven in een onderdeel. Omdat de aantallen opgaven per onderdeel verschillen zijn de standaardmeetfouten in de figuren onderling niet direct vergelijkbaar. Daarnaast is de schaal waarop de kansdichtheid is aangegeven bij elke figuur steeds aangepast aan de gevonden waarden. Ook dit zorgt ervoor dat de figuren onderling niet overal rechtstreeks vergelijkbaar zijn.

Figuur 5.1a Grootte van de meetfout voor de score op het onderdeel Rekenen van de Entreetoets groep 7 (inclusief de kansdichtheidfunctie voor deze score)

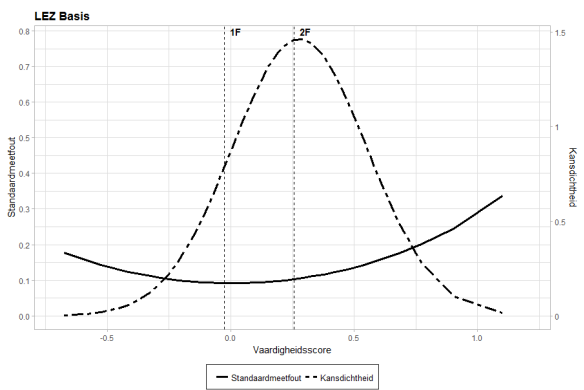


Module Basis

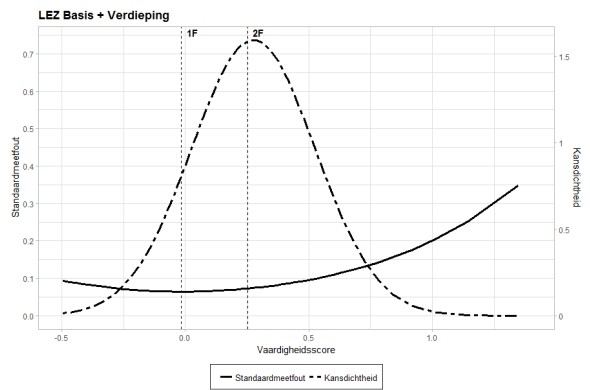


Module Basis en Verdieping

Figuur 5.1b Grootte van de meetfout voor de score op het onderdeel Lezen van de Entreetoets groep 7 (inclusief de kansdichtheidfunctie voor deze score)

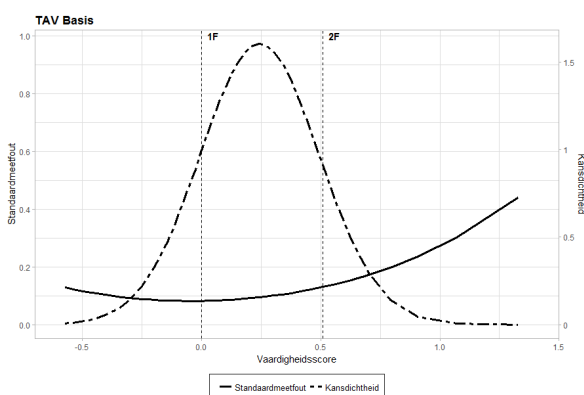


Module Basis

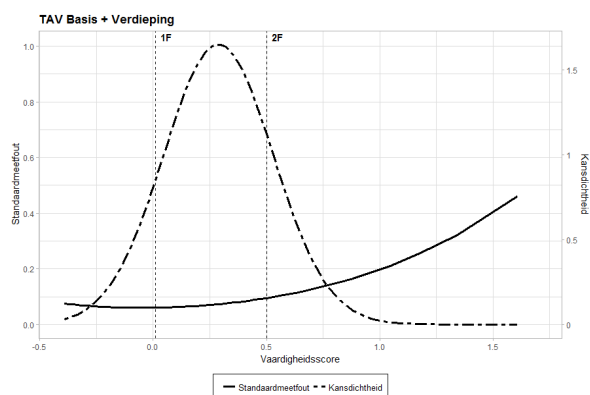


Module Basis en Verdieping

Figuur 5.1c Grootte van de meetfout voor de score op het onderdeel Taalverzorging van de Entreetoets groep 7 (inclusief de kansdichtheidfunctie voor deze score)



Module Basis



Module Basis en Verdieping

De figuren maken duidelijk dat de meetfout kleiner is in de lagere en gemiddelde vaardigheidsregionen dan in de hogere vaardigheidsregionen. Dit wordt vooral bepaald door de gemiddelde moeilijkheidsgraad van de toetsen en komt overeen met de bedoelingen van de toetsconstructeurs. Door de moeilijkheidsgraad voor de (onderdelen van) de basistoets en module Verdieping te kiezen in de buurt van .70, is het onderscheidend vermogen van de toets-(onderdelen) optimaal in het linkerdeel van de scoreverdeling, dus bij de zwakkere en gemiddelde scores. Door de gekozen moeilijkheidsgraad blijft de 'toetsbeleving' voor de meeste leerlingen prettig: zij worden niet geconfronteerd met een overmatig moeilijke toets.

Aangezien er bij deze onderdelen sprake is van rapportage op referentieniveau, heeft de lokale meetfout speciale betekenis. Het verdient namelijk de voorkeur als op de plaatsen waar sprake is van een cesuur (dat wil zeggen, referentieniveau wel gehaald versus referentieniveau niet gehaald) zo nauwkeurig mogelijk wordt gemeten. Daarmee wordt de kans op misclassificaties gereduceerd. Er is sprake van twee cesuren per onderdeel (1F en 1S voor Rekenen, 1F en 2F voor de onderdelen Lezen en Taalverzorging). Voor het onderdeel Rekenen in de basistoets komen de weergegeven scores op de vaardigheidsschaal (theta) (x-as) van .167 en .677 overeen met een ruwe score van respectievelijk 56 voor 1F en van 81 voor 1S (zie linkerdeel figuur 5.1a). Voor Rekenen in de basistoets en module Verdieping samen komen de weergegeven scores op de vaardigheidsschaal (x-as) van .174 en .643 overeen met een ruwe score van respectievelijk 88 voor 1F en van 124 voor 1S (zie rechterdeel figuur 5.1a). Soortgelijke figuren zijn gegeven voor Lezen (figuur 5.1b) en Taalverzorging (figuur 5.1c) met de referentiecesuren ingetekend in de figuren. De cesuren liggen voor Lezen in de basistoets op vaardigheidsniveau -.025 en .257 (overeenkomend met een ruwe score van 19 voor 1F en van 28 voor 2F) en voor Lezen in de basistoets en module Verdieping samen op vaardigheidsniveau -.017 en .252 (overeenkomend met een ruwe score van 39 voor 1F en 56 voor 2F). Voor Taalverzorging in de basistoets liggen de cesuren op vaardigheidsniveau .001 en .509 (overeenkomend met een ruwe score van 26 voor 1F en van 42 voor 2F) en voor Taalverzorging in de basistoets en module Verdieping samen op vaardigheidsniveau .009 en .500 (overeenkomend met een ruwe score van 47 voor 1F en 75 voor 2F).

Zoals te zien is in de figuren, liggen de referentiecesuren soms redelijk ver uit elkaar. Dat is met name bij Rekenen en Taalverzorging het geval. Hier gaat de cesuur voor 1S, respectievelijk 2F samen met een redelijk hoge ruwe toetsscore. Dit impliceert dat het lastig is om beide plaatsen even nauwkeurig te meten. Vanwege de gekozen moeilijkheidsgraad en het feit dat in samenhang daarmee de meetfouten voor de lagere en gemiddelde vaardigheidsscores kleiner zijn dan voor de hogere scores, wordt op het 1F-niveau nauwkeuriger gemeten dan op niveau 2F en 1S. In de buurt van 1F is dus de meetfout van *beide* toetsversies klein. We kunnen dus concluderen dat met de Entreetoets groep 7 optimaal kan worden vastgesteld of referentieniveau 1F al dan niet is behaald. Echter ook op niveau 1S, respectievelijk 2F is de nauwkeurigheid zeer behoorlijk, zeker als men in acht neemt dat de toetsen niet op twee heel verschillende niveaus even nauwkeurig hun werk kunnen doen. Voor Taalverzorging is de meetfout rond 2F relatief laag, maar hier is de afstand tussen 1F en 2F, zoals gezegd, klein.

Hoe geringer de meetfout, hoe nauwkeuriger de classificatie. De nauwkeurigheid van de classificaties is het onderwerp van de volgende paragraaf.

5.3 Classificaties

Leerlingen worden op basis van (onderdelen van) de Entreetoets groep 7 ingedeeld in een groep of categorie. De belangrijkste classificaties betreffen de vraag of de leerling met zijn score een bepaald referentieniveau heeft behaald of niet en de vraag welk brugklatype voor hem of haar het meest geschikt is. Geen enkele toets is onfeilbaar, ook al is de betrouwbaarheid psychometrisch gesproken optimaal. Er is altijd een zekere kans op fouten: misclassificaties. In het navolgende bespreken we deze kwestie eerst voor de rapportage op referentieniveaus en vervolgens voor de bepaling van het meest geschikte brugklatype.

5.3.1 Rapportage op referentieniveau: nauwkeurigheid van de classificatie

Zoals eerder besproken wordt er ten aanzien van drie vaardigheden een referentieniveau bepaald:

- Rekenen: referentieniveau 1F en 1S;
- Lezen: referentieniveau 1F en 2F;
- Taalverzorging: referentieniveau 1F en 2F.

Bij elk referentieniveau hoort een referentiecesuur (een score die aangeeft of een leerling het betreffende referentieniveau al dan niet heeft behaald). Omdat de vaardigheidsscore niet 100% nauwkeurig is, kan men bij een dergelijke beslissing fouten maken: de leerling krijgt onterecht het referentieniveau toegekend terwijl hij dit niveau in feite niet beheerst. Of omgekeerd, hij krijgt het referentieniveau niet toegekend, terwijl hij in feite wel over voldoende vaardigheid beschikt. In deze gevallen is er sprake van misclassificaties.

Omdat het steeds gaat om twee referentiecesuren per vaardigheid is er in feite sprake van een classificatie op drie niveaus:

- niveau 1F niet behaald;
- niveau 1F behaald, maar niveau 2F (dan wel 1S) niet;
- niveau 2F (dan wel 1S) behaald.

Tabel 5.2 Verwarringsmatrix voor Rekenen, conditioneel voor het werkelijke vaardigheidsniveau (in procenten)

a. Rekenen basistoets

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	30%	3%	0%	90%	10%	0%
1F wel, 1S niet	3%	52%	2%	6%	91%	3%
1S behaald	0%	3%	7%	0%	32%	68%

b. Rekenen basistoets en module Verdieping

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	32%	3%	0%	92%	8%	0%
1F wel, 1S niet	3%	51%	1%	5%	93%	3%
1S behaald	0%	3%	8%	0%	25%	75%

Om de nauwkeurigheid van de classificaties in beeld te brengen zijn simulaties uitgevoerd. Daarbij zijn per verwarringstabel 10.000 subjecten getrokken uit de (normaal verdeeld geachte) vaardigheidsverdeling ('ware vaardigheid') in de populatie, op basis van de itemparameters. Voor elk van de 10.000 subjecten is de ruwe score bepaald, waarna subjecten zijn ingedeeld in een 3 * 3 tabel op basis van zowel de 'ware vaardigheid' als de ruwe score. De aantallen in de cellen zijn vervolgens op twee manieren gepercenteerd, namelijk als percentage van het totaal (linkerdeel van tabel 5.2a) en rijgewijs (i.e. conditioneel op de ruwe score; rechterdeel van de tabel). De percentages op de diagonaal in het rechterdeel van de tabel geven

een indicatie van de accuratesse: de kans dat je gegeven de indeling op grond van de ruwe score met de 'ware vaardigheidsscore' in dezelfde scorecategorie terecht komt.

Het is goed te zien dat bij het onderdeel Rekenen in de basistoets verreweg de meeste leerlingen met hun werkelijke vaardigheidsscore in dezelfde referentieniveaugroep vallen als met hun geschatte vaardigheidsscore. Er zijn echter ook leerlingen die een niveaugroep te hoog of te laag geplaatst worden. Desalniettemin zijn de uitkomsten in lijn met het ambitieniveau zoals dat geformuleerd is door Pilliner (1969). Dat wil zeggen: praktisch gesproken scoren alle leerlingen in een scoregroep op basis van hun ware score, ook in werkelijkheid in die scoregroep, **of één scoregroep daarboven of één scoregroep daaronder**. De *Marginal Classification Accuracy* (d.w.z. de som van de diagonaal in het linkerdeel van de verwarringsmatrix als proportie van het totale aantal leerlingen) bedraagt 89%. Daarnaast is nog Cohens Kappa berekend als maat voor overeenstemming die gecorrigeerd is voor kans (dat wil zeggen, voor de frequentie van voorkomen van de onderscheiden categorieën). In het geval van het onderdeel Rekenen in de basistoets is de (ongewogen) Kappa gelijk aan .79.

Voor Rekenen in basistoets, aangevuld met de module Verdieping; zie tabel 5.2b zijn de overall nauwkeurighedsindicatoren nog wat positiever. De *Marginal Classification Accuracy* komt hier op 91%, Cohens Kappa op .83. Gezien het feit dat er vijftig opgaven méér nodig waren (140 in plaats van 90) is de winst in termen van nauwkeurigheid vrij marginaal te noemen. Nauwkeurighedsverbetering met betrekking tot de totaalscore was ook niet het motief om de verdiepende module aan te bieden. Het gaat hier met name om de extra mogelijkheden om gedifferentieerd te rapporteren die ontstaan als de subonderdelen qua aantal items worden uitgebreid.

Tabel 5.3 Verwarringsmatrix voor Lezen, conditioneel voor het werkelijke vaardighedsniveau (in procenten)

a. Lezen basistoets

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	8%	3%	0%	74%	26%	0%
1F wel, 1S niet	2%	26%	6%	7%	77%	16%
1S behaald	0%	7%	48%	0%	12%	88%

b. Lezen basistoets en module Verdieping

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	10%	2%	0%	83%	17%	0%
1F wel, 1S niet	2%	28%	4%	6%	82%	13%
1S behaald	0%	5%	49%	0%	9%	91%

In tabel 5.3a is te zien dat voor het onderdeel Lezen in de basistoets de meeste leerlingen op basis van hun 'ware score' in dezelfde referentieniveaugroep vallen als waarin zij op basis van hun ruwe score terecht komen. Er zijn echter ook leerlingen die een niveaugroep te hoog of te laag geplaatst worden. Desalniettemin zijn de uitkomsten in lijn met het ambitieniveau zoals dat geformuleerd is door Pilliner (1969). De *Marginal Classification Accuracy* (d.w.z. de som van de diagonaal in de verwarringsmatrix als

proportie van het totale aantal leerlingen) bedraagt 82%. Voor Lezen in de basistoets is Cohens Kappa gelijk aan .69.

Voor het onderdeel Lezen in basistoets aangevuld met de module Verdieping zijn de indicatoren zoals te verwachten positiever. De *Marginal Classification Accuracy* bedraagt nu 87%, Cohens Kappa is gelijk aan .77.

In tabel 5.4 (a en b), ten slotte, zijn de cijfers opgenomen voor Taalverzorging. Ook voor het onderdeel Taalverzorging in de basistoets vallen verreweg de meeste leerlingen op grond van hun 'ware score' in dezelfde referentieniveaugroep als op grond van hun ruwe score. De *Marginal Classification Accuracy* is 85%, Cohens ongewogen Kappa .65.

Voor de basistoets aangevuld met de module Verdieping laten de cijfers, in analogie met Rekenen en Lezen, een vergelijkbaar beeld zien: de *Marginal Classification Accuracy* is 88%, Cohens Kappa .72.

Tabel 5.4 Verwarringsmatrix voor Taalverzorging, conditioneel voor het werkelijke vaardigheidsniveau (in procenten)

a. Taalverzorging basistoets

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	8%	3%	0%	74%	26%	0%
1F wel, 1S niet	2%	26%	6%	7%	77%	16%
1S behaald	0%	7%	48%	0%	12%	88%

b. Taalverzorging basistoets en module Verdieping

Indeling naar referentieniveau						
Ruwe score	Ware vaardigheid			Ware vaardigheid		
	1F niet	1F wel, 1S niet	1S behaald	1F niet	1F wel, 1S niet	1S behaald
1F niet	10%	2%	0%	83%	17%	0%
1F wel, 1S niet	2%	28%	4%	6%	82%	13%
1S behaald	0%	5%	49%	0%	9%	91%

5.3.2 Rapportage van het meest geschikte brugklatype: nauwkeurigheid van de classificatie

De leerling krijgt op basis van zijn of haar gerealiseerde totaalscore voor de basistoets een brugklatype toegewezen. Daarbij wordt het indelingsschema in acht categorieën gehanteerd dat eerder is beschreven in paragraaf 4.4.1. In deze paragraaf gaan we – opnieuw in de vorm van een verwarringsmatrix (zie tabel 5.5) – na, hoe nauwkeurig deze classificatie is.

Tabel 5.5 is gebaseerd op een simulatie met 10.000 fictieve leerlingen, waarbij we uitgaan van de vaardigheden en de grenswaarden tussen de adviescategorieën (beide in termen van theta), zoals die eerder in hoofdstuk 4 aan de orde zijn geweest. Er was een overeenstemming van 67%. Dat wil zeggen dat in 67% van de gevallen het in de simulatie geobserveerde en werkelijke advies exact met elkaar overeenkwamen. Wanneer we (in lijn met het Pilliner ambitieniveau) ook plaatsing in de aanpalende adviescategorie als correct interpreteerden, steeg het percentage overeenstemming tot 97%. In de tabel zijn percentages conditioneel op de werkelijke vaardigheid en advies weergegeven (de percentages tellen

dus van links naar rechts op tot 100%). In de tabel is te zien dat de meeste leerlingen met hun score op de basistoets totaal hetzelfde brugklatype krijgen toegewezen als met hun geschatte score op de basistoets totaal. Er zijn echter ook leerlingen die op basis van hun geschatte score een hoger of lager brugklatype krijgen toegewezen. Desalniettemin zijn de uitkomsten zo goed als in lijn met het ambitieniveau zoals dat geformuleerd is door Pilliner (1969). Steeds wordt 95% (bij havo/gt) tot 100% (bij bb en vwo) van de leerlingen in een brugklatype op basis van hun ware score, ook in werkelijkheid in dat brugklatype, **of** één brugklatype daarboven **of** één brugklatype daaronder geclassificeerd.

De *Marginal Classification Accuracy* bedraagt .671. Dit betekent dat de classificatie op basis van de geschatte score op de basistoets gemiddeld gezien in ruim 65 procent van de gevallen overeenstemt met de classificatie op basis van de (gesimuleerde) werkelijke score op de basistoets. De gewogen kappa bedraagt .856.

Tabel 5.5 Classificatie in acht brugklatypes: verwarringsmatrix

Indeling naar best passende brugklatype									Accuracy +/- 1 niveau
geschat type	Type gebaseerd op gerealiseerde standaardscore								
	bb	bb/kb	kb	gt	gt/h	havo	h/v	vwo	
bb	80%	20%	0%	0%	0%	0%	0%	0%	100%
bb/kb	8%	71%	16%	4%	0%	0%	0%	0%	96%
kb	0%	26%	36%	34%	4%	0%	0%	0%	96%
gt	0%	4%	17%	50%	29%	0%	0%	0%	96%
gt/havo	0%	0%	1%	13%	67%	14%	4%	0%	95%
havo	0%	0%	0%	1%	31%	38%	31%	0%	99%
havo/vwo	0%	0%	0%	0%	3%	13%	74%	11%	97%
vwo	0%	0%	0%	0%	0%	0%	14%	86%	100%

Deze resultaten zijn als zeer bevredigend te duiden.

6 Validiteit

Het hoofddoel van de Entreetoets groep 7 is het *beschrijven* van het niveau van leerlingen op het gebied van Rekenen, Lezen en Taalverzorging zoals omschreven in het referentiekader Nederlandse Taal en Rekenen. Met de basistoets kan de leerkracht in kaart brengen in welke mate de leerlingen de referentieniveaus Rekenen, Lezen en Taalverzorging beheersen.

Een nadere analyse op de domeinscores die deel uitmaken van Rekenen, Lezen en Taalverzorging is mogelijk wanneer ook de module Verdieping is afgenomen. Het gedifferentieerde overzicht van de leervorderingen dat op deze manier ontstaat, kan door het afnemen van de module Verbreding verder worden uitgebreid. Men kan zo de talige onderdelen Luisteren, Schrijven en Woordenschat aan het overzicht toevoegen, evenals de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek die deel uitmaken van het onderdeel Wereldoriëntatie.

Door middel van rapportages is in één oogopslag te zien waar de leerling goed in is en waar ondersteuning nodig is. Daarmee kan de leerkracht hiaten op het spoor komen op een moment dat deze nog kunnen worden 'bijgespijkerd' zo lang het kind nog op de basisschool zit. In de communicatie met de ouders is het van belang een objectief beeld van de leerling te kunnen schetsen.

Daarnaast is het mogelijk om de resultaten op de Entreetoets te gebruiken als informatie die – via het rapport Vooruitblik – een bijdrage kan leveren aan het formuleren van een advies voor het brugklastype dat het beste past bij het niveau van leervorderingen van elke leerling. Men kent hiertoe een zekere predictieve waarde toe aan de toets.

Ten aanzien van beide functies, de beschrijvende zowel als de voorspellende, is het noodzakelijk dat het instrument aantoonbaar inhouds- en begripsvalide is. Voor de tweede functie is empirisch onderzoek naar de criteriumvaliditeit noodzakelijk.

In dit hoofdstuk besteden we aandacht aan zowel de begripsvaliditeit (paragraaf 6.2) als de criteriumvaliditeit (paragraaf 6.3) van de Entreetoets Groep 7. Een leervorderingentoets ontleent zijn validiteit echter vooral aan de vraag hoe de toets is samengesteld, de vraag naar inhoudsvaliditeit dus. In paragraaf 6.1 vatten we daarom nog eens samen wat er in hoofdstuk 3 is geschreven over de wijze waarop de communale onderwijsdoelen in de toets zijn geoperationaliseerd.

6.1 Inhoudsvaliditeit

De inhoudsvaliditeit van een toets heeft betrekking op de vraag in hoeverre de opgaven in een toets een welomschreven en afgebakend universum representeren van mogelijk in de toets op te nemen opgaven. De opgaven in de Entreetoets groep 7 sluiten nauw aan bij het doel en de inhoud van het onderwijs op de basisschool. Dat moet ook, want zij dienen een goede afspiegeling te vormen van wat leerlingen in het basisonderwijs tot aan die afnamedatum geleerd hebben. Daarbij moeten het opgaven zijn waarin leerstof behandeld wordt die elke basisschoolleerling zou kunnen beheersen, uitgaande van de communale onderwijsdoelen en -inhouden. Ook voor een optimaal voorspellend vermogen is dit bovendien noodzakelijk.

Een uitgebreide inhoudelijke analyse van de inhoud van de Entreetoets groep 7 staat in hoofdstuk 3 van deze verantwoording. Hier vatten we nog eens samen hoe ervoor gezorgd is dat de toetsinhoud representatief is voor de communale onderwijsdoelen en -inhouden.

Aansluiten bij (theorieën over) vaardigheden die in het basisonderwijs tot ontwikkeling komen

Er is gekozen voor een verantwoording op basis van te realiseren onderwijsdoelen. Daarbij moeten we niet uit het oog verliezen dat de toetsonderdelen betrekking hebben op vaardigheden en deelvaardigheden. Deze – veronderstelde – deelvaardigheden ('latente trekken'), zoals rekenvaardigheid, (begrijpend) lezen, vormen de basis voor de toetsonderdelen en zijn te beschrijven in termen van de ontwikkeling die het kind gedurende de basisschool doormaakt. We hebben er in deze verantwoording niet voor gekozen om

uitgebreid in te gaan op de theorie rondom deze vaardigheden en vaardigheidsontwikkeling. Dat zou deze wetenschappelijke verantwoording onleesbaar gemaakt hebben. De geïnteresseerde lezer wordt op dit punt daarom voor meer informatie verwezen naar de wetenschappelijke verantwoordingen van de toetsen die deel uitmaken van het LVS, met name de verantwoordingen die betrekking hebben op de leerjaren 6 tot en met 8. Uiteraard blijft het wel zaak om te laten zien dat alle opgaven van een deoltoets, ook al representeren zij verschillende categorieën van doelstellingen, desondanks één en dezelfde vaardigheid representeren. We komen hierop in paragraaf 6.2.1 nog terug.

Aansluiten bij kerndoelen voor het basisonderwijs

De Entreetoets groep 7 sluit dus enerzijds aan bij de vaardigheidsgroei die zich in basisschoolperiode voltrekt en zoals die te beschrijven, te evalueren en te volgen is in de diverse LVS-toetsen, anderzijds bij de kerndoelen voor het basisonderwijs: de kerndoelen van het Ministerie van Onderwijs, Cultuur en Wetenschap (Ministerie van Onderwijs, Cultuur en Wetenschap, 2004, 2006)), de Toetswijzer Eindtoets Basisonderwijs (College voor Examens, 2012) en het Referentiekader Nederlandse Taal en Rekenen voor het basisonderwijs (Expertgroep doorlopende leerlijnen taal en rekenen, 2008b). Met name over de communale kerndoelen die betrekking hebben op de module Basis (dat wil zeggen de leerdoelen voor Rekenen, Lezen en Taalverzorging) en de module Verdieping bestaat een zekere consensus. Echter ook voor de onderdelen en domeinen in de module Verbreding konden de constructeurs bij het maken van de opgaven terugvallen op soortgelijke publicaties.

Inhoudsanalyse gericht op vaststelling van de optimale toetsinhoud

In de inhoudsverantwoording zoals deze is beschreven in hoofdstuk 3 is gedetailleerd te volgen hoe we bij de constructie de genoemde basis- en andere documenten hebben geanalyseerd in hun consequenties voor de toetsinhoud. Daarbij is steeds (vooraf) geëxpliciteerd hoe de optimale toetsinhoud eruit zou moeten zien in de verdeling van opgaven over diverse categorieën, bijvoorbeeld in de vorm van toetsmatrijzen. Cito analyseert bij het maken van zijn toetsen voortdurend doelstellingen en methoden en gaat steeds na of de inhoud ervan nog voldoende aansluit bij de inhoud van het onderwijs.

Afstemming van de gerealiseerde op de gewenste (optimale) toetsinhoud.

De toetsmatrijzen vormen steeds het kader van waaruit (specifieke typen en aantallen) opgaven worden geconstrueerd. Dit vormt echter geen garantie voor inhoudsvaliditeit. Ook al wordt (op grond van de verwachte mortaliteit) een veelvoud van het benodigde aantal opgaven geconstrueerd om uit te testen in de proeftoetsen, in de praktijk blijken soms na de proeftoetsing onvoldoende (op psychometrische gronden adequate) opgaven van een bepaald type beschikbaar te zijn. Ook het omgekeerde komt voor, namelijk dat er meer uitstekende opgaven van een bepaald type beschikbaar zijn dan op basis van de inhoudsanalyse noodzakelijk werd geacht. Kortom, bij de uiteindelijke selectie van de opgaven is er altijd sprake van afweging van inhoudelijke en psychometrische aspecten. Dit impliceert dat niet altijd de exacte gewenste inhoudelijke verdeling van opgaven wordt gerealiseerd. Over het algemeen kunnen we echter concluderen dat de gerealiseerde toetsinhoud dicht in de buurt van de gewenste inhoud, zoals af te leiden valt uit de inhoudsverantwoording in hoofdstuk 3. Daarbij moet men zich ook realiseren dat in de gevallen waarin de afwijkingen van de gewenste toetsinhoud wat groter zijn, er altijd sprake is van opgaven die uitstekend passen bij de onderliggende latente trek (zoals blijkt uit de kalibratieresultaten).

Afgezien van bovenstaande overwegingen kunnen we constateren dat de gewenste aantallen opgaven voor de onderdelen van een domein (bijvoorbeeld de aantallen opgaven voor de domeinen Begrijpend lezen, Samenvatten en Opzoeken bij het onderdeel Lezen) overal precies zo zijn gerealiseerd. Waar het gaat om meer ondergeschikte indelingen waren de wensen soms minder stringent geformuleerd, en zijn afwijkingen van de gewenste situatie ook minder van belang.

Evaluatie van de moeilijkheidsgraad

Opgaven moeten zijn afgestemd op het niveau van de leerlingen. Ze mogen dus niet te moeilijk en niet te gemakkelijk zijn. Het RCEC Beoordelingssysteem (Sanders et al., 2016) noemt bijvoorbeeld grenzen van .90 (niet te gemakkelijk, dat wil zeggen met een P-waarde $>.90$) en .20 (niet te moeilijk, dat wil zeggen met een P-waarde $<.20$), waarbij zich 90% van de opgaven binnen deze grenzen zou moeten bewegen.

Anderzijds zijn (binnen de genoemde grenzen) gemakkelijke en moeilijke opgaven nodig om op elk vaardigheidsniveau onderscheid te kunnen maken tussen leerlingen.

Afstemming van de moeilijkheidsgraad op de verwachte vaardigheid van *alle* leerlingen is een belangrijk uitgangspunt voor de Entreetoets Groep 7. De moeilijkheidsgraad van de opgaven wordt verderop als aspect van de itemkwaliteit (die door de COTAN gezien wordt als criterium voor de begripsvaliditeit van een toets) aan de orde gesteld. We lopen daarop vooruit en vatten hier de belangrijkste resultaten samen. De gemiddelde P-waarden voor de toetsonderdelen liggen tussen .57 en .76. Er zijn geen te moeilijke opgaven met P-waarden lager dan .20; iets te gemakkelijke opgaven, met P-waarden van .91 of meer, komen bij enkele onderdelen voor. Voor alle toetsonderdelen en de toets als geheel wordt aan de eis voldaan dat 90% van de opgaven of meer voldoet aan de gestelde criteria voor de moeilijkheidsgraad. De Entreetoets Groep 7 is dus goed afgestemd op het niveau van de doelgroep.

Deskundige opgavenconstructie

De benodigde aantallen opgaven werden geconstrueerd in zogeheten constructiegroepen bestaande uit drie à vier mensen 'uit het veld': leerkrachten van groep 7 of 8, PABO-docenten en/of IB'ers, onder leiding van een toetsdeskundige van Cito. Hun belangrijkste taak is echter om vanuit hun expertise op vakinhoudelijk en toetstechnisch gebied commentaar te geven op de geconstrueerde opgaven. In totaal werden, rekening houdend met de te verwachten mortaliteit voor alle onderdelen van de toets, 1174 nieuwe opgaven geconstrueerd.

Proeftoetsing van alle opgaven

Alle 1174 opgaven werden getest in empirisch proefonderzoek. Opgaven die niet bleken te voldoen, zoals bleek uit de empirische gegevens, werden niet opgenomen in de toetsen. Hierbij werd de inhoud bewaakt vanuit de omschreven doelstellingencategorieën: de toetsen dienen representatief te zijn ten aanzien van de geformuleerde specificatietabel (toetsmatrijs).

Ook kregen leerkrachten bij de proeftoetsen het verzoek om de gehanteerde onderwijsdoelen en de van daaruit geconstrueerde opgaven van commentaar te voorzien.

Op deze manier werd de inhoudsvaliditeit inhoudelijk als veilig gesteld.

6.2 Begripsvaliditeit

De begripsvaliditeit van een toets heeft betrekking op de vraag in hoeverre de toetsscores toe te schrijven zijn aan de concepten en constructen die deel uitmaken van het theoretische kader dat aan de ontwikkeling van de toets ten grondslag ligt. Het gebruik van de termen concepten en constructen in het kader van leervorderingentoetsen ligt misschien minder voor de hand. Toch zijn er ook bij instrumenten als Entreetoets 7 verschillende manieren om empirische evidentie te verzamelen die kan worden opgevat als onderbouwend voor de begripsvaliditeit ervan. We zullen deze hieronder achtereenvolgens bespreken.

6.2.1 Passing van het meetmodel

De opgaven die worden opgenomen in de Entreetoets, vormen na proeftoetsing en kalibratie een gekalibreerde opgavenbank. Bij de analyse van de antwoorden van de leerlingen op de opgaven wordt nagegaan of de verschillende opgaven en opgaventypen een beroep doen op hetzelfde complex aan vaardigheden. Items die niet voldoen aan de beschreven passingscriteria worden uit de verzameling verwijderd. Dit impliceert dat items die op basis van inhoudsanalyse zijn geconstrueerd vanuit de veronderstelling dat de leerling over een bepaalde specifieke vaardigheid dient te beschikken om de opgave te kunnen oplossen worden onderworpen aan een check of deze vaardigheid samenvalt met de unidimensionele latente trek waarvan men aanneemt dat deze de basis vormt van de (deel)toets. Op deze manier grijpen inhoudsvaliditeit en begripsvaliditeit in elkaar.

In paragraaf 4.1.2 is beschreven hoe het gehanteerde meetmodel is getoetst en zijn gegevens gepresenteerd waaruit af te leiden valt of en in welke mate de kalibraties geslaagd genoemd kunnen worden. Wij vatten hier de conclusies samen.

In de Entreetoets gaat het om de volgende onderdelen en bijbehorende vaardigheidsschalen: Rekenen, Lezen en Taalverzorging (zowel in de basistoets als in de module Verdieping), de aanvullende tailge onderdelen in de module Verbreding (Luisteren, Schrijven en Woordenschat) en de Werledoriëntatie-domeinen in de module Verbreding (Aardrijkskunde, Geschiedenis en Natuur en techniek). Bij alle deze onderdelen en domeinen wordt verondersteld dat er sprake is van een onderliggende latente trek (vaardigheid). Dat deze veronderstellingen correct waren, blijkt uit de volgende analyses en resultaten: Uit de S-toetsen en grafische weergaven daarvan is bij alle genoemde onderdelen en domeinen gebleken dat de afzonderlijke opgaven steeds één onderliggende vaardigheidsdimensie representeren. Tevens geldt dat de nauwkeurigheid van de itemparameterschattingen bij de betreffende opgaven in alle gevallen uitstekend is: 'constante c' (zie COTAN-beoordelingssysteem; Evers et al., 2010) is bij geen enkele opgave hoger dan .20; de hoogst waarde die werd aangetroffen bedraagt .161. De gemiddelde waarden per toetsonderdeel zijn altijd lager dan .10 (ze variëren tussen .016 en .045) en daarmee uitstekend te noemen. Alle waarden voldoen ruimschoots aan de door de COTAN geformuleerde minimumeisen.

Voor de onderzochte onderdelen hebben we kunnen constateren dat de rechteroverschrijdingskansen van de S-toetsen gelijkmatig verdeeld zijn over de hele range van mogelijke waarden, met een klein aantal significante waarden (in overeenstemming met wat men op basis van toeval zou mogen verwachten). Dit duidt op een goede modelfit.

Voor de onderzochte onderdelen als geheel geldt dat de modelfit in termen van R1c voldoet aan de verwachtingen.

Voor elk van de onderdelen waarvoor de assumptie van unidimensionaliteit geldt, kan dus worden geconcludeerd dat er sprake is van een unidimensionaal concept (respectievelijk latente trek of vaardigheidsscore) onder de opgaven in het desbetreffende toetsonderdeel. In termen van klassieke testtheorie sluiten de uitkomsten van de itemanalyses (de R_{it} -waarden) en hoge interne consistentie (GLB, respectievelijk Cronbachs alfa) hierbij aan.

De gunstige uitkomsten zijn op zichzelf nog geen bewijs voor begripsvaliditeit (meet de toets en elke deeltoets wat hij geacht wordt te meten?), maar vormen daarvoor wel een noodzakelijke voorwaarde als eerste stap in de bewijsvoering.

6.2.2 Structuur

Leervorderingen (i.e. scores op leervorderingentoetsen) zijn op te vatten als een complex van vaardigheden die onderling relatief sterk samenhangen. Elk van de veronderstelde vaardigheden is daarbij op te vatten als een onderliggende latente trek (zie de vorige paragraaf 6.2.1). De op basis van IRT per vaardigheid geconstrueerde toetsonderdelen hebben een sterk unidimensioneel karakter. Bij aanvang van de toetsconstructie waren er naast de genoemde assumpties geen duidelijk uitgesproken verwachtingen over de structuur van de Entreetoets Groep 7 in termen van de intercorrelaties tussen de onderdelen. De intercorrelaties vertonen een vrij stabiel patroon, zoals we bijvoorbeeld weten van de deelscores op de Eindtoets Basisonderwijs.

Tabel 6.1 Latente correlaties tussen de hoofdonderdelen van de module Basis (boven de diagonaal) en de module (Basis en) Verdieping (onder de diagonaal)

	Rekenen	Lezen	Taalverzorging
Rekenen	--	.75	.66
Lezen	.75	--	.72
Taalverzorging	.67	.73	--

N = 56217 voor de module Basis

N = 35606 voor de module (Basis en) Verdieping

Een andere goede bron is de wetenschappelijke verantwoording van de Centrale Eindtoets (zie CvTE, 2016). De Centrale Eindtoets bevat onderdelen die goed vergelijkbaar zijn met die van de Entreetoets Groep 7. Bovendien zijn in de betreffende verantwoordingen ook ranges van te verwachten correlaties opgenomen. Ten slotte zijn ook wetenschappelijke verantwoordingen van Cito LVS-toetsen als bron bruikbaar. We hebben in dit opzicht gebruikgemaakt van de verantwoording van de toetsen Taalverzorging groep 6 – 8 (Roumans & Engelen, 2016) en Begrijpend Lezen 3.0 voor groep 6 (Tomesen & Weekers, in voorbereiding).

Om inzicht te krijgen in de patronen presenteren we in tabel 6.1 eerst de intercorrelaties tussen de drie onderdelen Lezen, Rekenen en Taalverzorging (de module Verdieping wordt nog buiten beschouwing gelaten). We geven de correlaties zowel voor de basistoets als voor de basistoets en Verdieping. De betrouwbaarheden van de betreffende deeltoetsen verschillen door hun toetslengte, maar dat is ondervangen door te corrigeren voor attenuatie. Ten aanzien van deze samenhangen hadden we nog geen duidelijke hypothesen, omdat voorheen uitsluitend onderscheid werd gemaakt tussen Rekenen en Taal (en binnen Taal niet tussen Lezen en Taalverzorging). De correlatie tussen Rekenen en Taal in de Centrale Eindtoets van 2015 bedraagt .75. De in de Entreetoets aangetroffen correlaties tussen Rekenen enerzijds en de beide Taalonderdelen Lezen (.75) en Taalverzorging (.66 / .67) anderzijds liggen in dezelfde orde van grootte en zijn voor de basistoets en de basistoets en module Verdieping vrijwel hetzelfde. Hetzelfde geldt voor de correlatie tussen Lezen en Taalverzorging; deze is voor de basistoets .72, voor de basistoets en de module Verdieping .73.

Ook interessant zijn de samenhangen tussen domeinen die zijn ondergebracht in hetzelfde onderdeel. Alle intercorrelaties zijn weergegeven in tabel 6.2.

Op de eerste plaats kunnen we constateren dat het niet veel uitmaakt op welke verzameling van data we de intercorrelaties baseren wanneer voor onbetrouwbaarheid wordt gecorrigeerd. De correlaties voor de basistoets wijken bij Rekenen en Taalverzorging nergens meer dan .02 van de overeenkomende correlaties voor de basistoets en de module Verdieping. Voor Lezen zijn de verschillen iets groter (tussen .04 en .06). Dat betekent dat beide sets van intercorrelaties leiden tot dezelfde conclusies. Welke conclusies zijn dat? Voor *Rekenen* gingen we ervan uit dat we te maken hebben met een zeer homogene vaardigheid. Doorgaans kunnen de opgaven voor de vier onderscheiden onderdelen op één schaal worden gekalibreerd. We verwachtten dan ook zeer hoge intercorrelaties (tussen .85 en .99; zie de verwachte range in de tabel, hier gebaseerd op de gerapporteerde schattingen en vastgestelde waarden bij de Centrale Eindtoets). Alle veronderstellingen werden empirisch bevestigd met waarden tussen .89 en .95.

Ook voor de domeinen van *Lezen* verwachtten we hoge intercorrelaties, zij het minder hoog dan bij Rekenen. De verwachtingen (.85 - .90) zijn gebaseerd op de aangetroffen intercorrelaties voor de Toets E6

Begrijpend lezen 3.0 die bedoeld is voor leerlingen van groep 6 aan het einde van het leerjaar.

De correlaties die we vinden voor de Entreetoets Groep 7 zijn in dezelfde orde van grootte, maar iets lager voor de basistoets. Voor de basistoets en de module Verdieping vallen ze precies binnen de verwachte range van waarden.

Met *Taalverzorging* als verzamelbegrip voor de domeinen Spelling (werkwoorden en niet-werkwoorden), Grammatica en Interpunctie is nog niet veel ervaring opgedaan. Voor de CET werden intercorrelaties in leerjaar 8 aangetroffen die liggen tussen .70 en .72, maar het is de vraag of deze zonder meer toepasbaar zijn op de situatie in leerjaar 7. De mogelijkheid bestaat namelijk dat taalverzorging een vaardigheid is die zich in de bovenbouw van het basisonderwijs ontwikkelt tot grotere eendimensionaliteit. Al veel eerder is vastgesteld dat Spelling niet-werkwoorden en Spelling-werkwoorden moeilijk op één vaardigheidsschaal zijn onder te brengen. De reden is, dat het spellen van werkwoorden pas in groep 6 wordt geïntroduceerd. Bij dit onderdeel van spellen moeten leerlingen een beroep doen op grammatica- en interpunctieregels om tot een juiste spelling te kunnen komen. Met name kennis van grammatica lijkt dus een voorwaardelijke factor van belang (zie Roumans & Engelen, 2016). De veronderstelling is dat er in de leerjaren 6 tot en met 8 sprake is van toenemende intergratie van de deelvaardigheden. Om die reden hebben we de intercorrelaties tussen de domeinen zoals die werden gevonden voor groep 7 als referentie in de tabel opgenomen; deze zijn over het algemeen wat lager dan voor vergelijkbare intercorrelaties in leerjaar 8. We kunnen vaststellen dat de correlatie tussen Spelling-werkwoorden en Spelling-werkwoorden met .76 het hoogste is (maar iets hoger dan de verwachte .70). Voor de overige correlaties geldt dat deze vrijwel overeenkomen met de verwachte waarden (alleen de waarden van de correlatie tussen Spelling werkwoorden en Grammatica is met .59 iets hoger dan de verwachte .50). We veronderstellen dat de hoger dan verwachte correlaties van .76 en .59 voor Spelling-werkwoorden te danken zijn aan een verdere voortschrijding van het integratieproces in de loop van leerjaar 7.

Onze slotconclusie is, dat – de hele tabel 6.2 overziende – intercorrelaties tussen domeinen binnen dezelfde onderdelen overeenkomen met de verwachtingen. Dat geldt zeker voor de zeer hoge waarden binnen de homogene vaardigheid rekenen en ook voor de hoge samenhangen die werden gevonden bij de leesonderdelen. Bij Taalverzorging zijn minder uitgesproken verwachtingen. De uitkomsten zijn overeenkomstig de ervaringen bij LVS.

Tabel 6.2 Latente correlaties tussen de deoltoetsen van de onderdelen Rekenen, Lezen en Taalverzorging: module Basis (boven de diagonaal) en de module (Basis en) Verdieping (onder de diagonaal); tussen haken de verwachte (range van) correlaties

Onderdeel Rekenen

	Getallen	Met en meetkunde	Verhoudingen	Verbanden
Getallen	--	.93 (.85-.99)	.92 (.85-.99)	.89 (.85-.99)
Met en meetkunde	.92	--	.94 (.85-.99)	.91 (.85-.99)
Verhoudingen	.90	.93	--	.94 (.85-.99)
Verbanden	.90	.92	.95	--

N = 56224 voor de basistoets

N = 35687 voor de basistoets en module Verdieping

Onderdeel Lezen

	Begrijpend lezen	Opzoeken	Samenvatten
Begrijpend lezen	--	.84 (.85-.90)	.81 (.85-.90)
Opzoeken	.88	--	.84 (.85-.90)
Samenvatten	.87	.88	--

N = 56344 voor de basistoets

N = 35637 voor de basistoets en module Verdieping

Onderdeel Taalverzorging

	Spelling werkwoorden	Spelling niet-werkwoorden	Grammatica	Interpunctie
Spelling werkwoorden	--	.76 (.70)	.59 (.50)	.61 (.60)
Spelling niet-werkwoorden	.77	--	.65 (.63)	.67 (.68)
Grammatica	.58	.63	--	.66 (.67)
Interpunctie	.60	.65	.66	--

N = 56326 voor de basistoets

N = 35635 voor de basistoets en module Verdieping

Tot nu toe hebben we nog niet gesproken over de onderdelen van de module Verbreding. We gaan hier eerst in op de domeinen van Wereldoriëntatie (zie tabel 6.3). Daarover hebben we geen andere verwachtingen dan die welke we kunnen baseren op de analyses met de Centrale Eindtoets (CvTE, in voorbereiding). Deze zijn tussen haakjes in de tabel opgenomen.

Tabel 6.3 Latente correlaties tussen de domeinen van Wereldoriëntatie

	Aardrijkskunde	Geschiedenis	Natuur en techniek
Aardrijkskunde	--	.86 (.80-.85)	.87 (.80-.85)
Geschiedenis		--	.87 (.90-.95)
Natuur en techniek			--

N = 10364

De gevonden waarden liggen in de buurt van de hoge samenhangen die we op basis van CET-resultaten in groep 8 mochten verwachten. Soms zijn ze iets hoger, soms iets lager dan verwacht.

Tabel 6.4 Latente correlaties tussen de taalonderdelen van de module Verbreding en de domeinen van de onderdelen Lezen en Taalverzorging deelvaardigheden Taal

	Domeinen Lezen			Domeinen Taalverzorging			
	Begrijpend lezen	Opzoeken	Samenvatten	Spelling werkwoorden	Spelling niet-werkwoorden	Grammatica	Interpunctie
	Basistoets						
Luisteren	.76	.76	.73	.36	.45	.52	.55
Schrijven	.84	.78	.79	.55	.64	.64	.77
Woordenschat	.81	.82	.76	.47	.57	.55	.62
	Basistoets en module Verdieping						
Luisteren	.77	.78	.77	.36	.41	.51	.57
Schrijven	.87	.81	.83	.57	.61	.66	.77
Woordenschat	.82	.83	.79	.47	.51	.54	.63

N = 8398 voor de combinatie met de basistoets

N = 6742 voor de combinatie met de basistoets en module Verdieping

De overige onderdelen binnen de module Verbreding (Luisteren, Schrijven en Woordenschat) zijn alle toetsonderdelen die betrekking hebben op semantische aspecten van de taalvaardigheid. Uit validiteits-overwegingen zijn we met name geïnteresseerd in de vraag of semantische deelvaardigheden die worden gemeten in de module Verbreding hoger samenhangen met andere semantische deelvaardigheden zoals gemeten in de basistoets (dan wel in combinatie met de module Verdieping) dan met niet-semantische deelvaardigheden in de basistoets (en in combinatie met de module Verdieping). De betreffende correlaties tussen latente vaardigheden staan in tabel 6.4.

Opnieuw is vast te stellen dat de patronen van intercorrelaties voor de basistoets en de basistoets aangevuld met de module Verdieping hetzelfde zijn. De waarden zijn onderling vergelijkbaar. Het is duidelijk dat onze verwachtingen worden bevestigd. Alle correlaties tussen semantische deelvaardigheden van de taalvaardigheid onderling (linkerdeel van de tabel) zijn hoog en liggen tussen .73 en .87. Per vaardigheid (rijgewijs) zijn de correlaties met semantische onderdelen (links) systematisch hoger dan de correlaties met de niet-semantische onderdelen (van Taalverzorging; rechts). De verschillen zijn soms aanmerkelijk. Opvallend, maar te verwachten, is het feit dat de correlaties tussen semantische en niet-semantische onderdelen het hoogst zijn bij het onderdeel Schrijven. In dit onderdeel komen semantische en niet-semantische deelvaardigheden als het ware bij elkaar. Dat geldt vooral voor Interpunctie (.77), maar ook voor Grammatica en in iets mindere mate de spellingonderdelen. Aan de andere kant is ook heel duidelijk te zien dat de niet-semantische taalvaardigheden het zwakst samenhangen met Luisteren. Vooral de spellingvaardigheid doet er kennelijk weinig toe.

De slotsom is dat met deze resultaten onze hypothesen worden bevestigd en dat daar waar we het formuleren van hypothesen achterwege lieten, de resultaten goed te interpreteren zijn.

6.2.3 Soortgenootvaliditeit

Wanneer het er op aankomt bij leerlingen gegevens te verzamelen in het kader van de vaststelling van de convergente en discriminante validiteit, is dit in het geval van de Entreetoets 7 een lastige onderneming. Afname van de toets vergt immers behoorlijk wat tijd van leerkracht en leerlingen. Het ligt niet voor de hand om school en leerlingen dan nog verder te belasten met de afname van aanvullende toetsen. Het is in principe mogelijk om via de Cito-dataretourfunctie samenhangen te onderzoeken tussen LVS-toetsen en scores op onderdelen van de Entreetoets. Dit is ook gebeurd voor de Entreetoets groep 6 (van Boxtel, Engelen, Frissen en van der Palm, 2015). De resultaten vormden een ondersteuning voor de convergente en divergente validiteit van toetsonderdelen. Er is geen reden om aan te nemen dat dit niet het geval zal zijn voor de Entreetoets groep 7. Scores op onderdelen van de Entreetoets groep 7 kunnen echter ook worden gebruikt ter vervanging van afnames van LVS-toetsen. Om de toetslast van hun leerlingen beperkt te houden, nemen scholen daarom soms geen LVS-toetsen af als er al een Entreetoets is afgenomen.

Voorts is er sprake van equivalentie met de voorganger van de hier verantwoorde Entreetoets Groep 7. Ondanks de inhoudelijke verschillen tussen beide versies is hun onderline samenhang hoog (zie verderop in paragraaf 6.3 onder criteriumvaliditeit). De hoge samenhang kan worden aangemerkt als indicatie voor de soortgenootvaliditeit, temeer daar de vorige versie van de Entreetoets ook een sterke samenhang liet zien met de Cito Eindtoets Basisonderwijs.

6.2.4 De Entreetoets groep 7: relevante verschillen tussen subgroepen

In deze paragraaf wordt een aantal uitkomsten van analyses gepresenteerd waarbij steeds onderscheid is gemaakt naar een aantal subgroepen. We hebben ons daarbij beperkt tot (achtergrond)variabelen waarvoor op grond van theoretische overwegingen verwachtingen zijn te formuleren over verschillen en/ of overeenkomsten in score en die daarmee interessant zijn voor de begripsvaliditeit van de Entreetoets 7. We besteden aandacht aan verschillen tussen groepen met een verschillend leerlinggewicht, aan sekse en aan leeftijd (binnen leerjaar 7). Alle achtergrondkenmerken zijn al eerder in deze wetenschappelijke verantwoording geïntroduceerd en besproken (zie paragraaf 2.5.2 en 4.2.1). Bij leerlinggewicht gaat het om

de categorieën 0.0, 0.30 en 1.20 (op individueel niveau), bij sekse om jongens versus meisjes en bij leeftijd om de categorieën voorlijk, regulier en vertraagd.

Leerlinggewicht

In tabel 6.5 staan de gemiddelden en standaarddeviaties voor de categorieën 0.0, 0.30 en 1.20. De categorie 0.0 staat voor reguliere leerlingen met een standaard bekostiging. Voor de categorieën 0.30 en 1.20 ontvangen de scholen extra bekostiging om de betreffende leerlingen zonedig extra aandacht te kunnen geven. In de regel scoren leerlingen uit deze twee categorieën gemiddeld lager dan de reguliere leerlingen. Over het verschil tussen 0.30- en 1.20-leerlingen hebben we geen hypothese.

Tabel 6.5 *Verschillen tussen leerlingen naar leerlinggewicht voor de onderdelen Rekenen, Lezen en Taalverzorging, toetsing op significantie en effectgrootte*

Rekenen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
0.0	61.8	15.6	44217	938.5	0
0.30	52.1	16.5	2495	P<.001	-.68
1.20	51.2	15.9	2461		-.62
Lezen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
0.0	28.0	6.4	44301	1297.5	0
0.30	23.8	6.6	2510	P<.001	-.66
1.20	22.6	6.6	2474		-.85
Taalverzorging					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
0.0	34.2	7.0	44289	439.5	0
0.30	30.8	7.0	2506	P<.001	-.48
1.20	31.3	7.2	2473		-.41

Voor alle onderdelen geldt dat de groepen met een afwijkend leerlinggewicht (0.30 en 1.20) gemiddeld aanzienlijk lager scoren dan de groep reguliere leerlingen. Alle verschillen zijn significant ($p < .001$). De groepen met een afwijkend leerlinggewicht ontlopen elkaar niet veel. In termen van Cohens vuistregels (Cohen, 1992) zijn de effecten te beoordelen als middelgroot.

Sekse

Over het algemeen doen meisjes het wat beter dan jongens op talige onderdelen; daarbij zijn de effectgroottes in de regel bescheiden. Jongens doen het daarentegen beter dan meisjes bij rekenen, waarbij de effectgroottes in de regel iets robuuster zijn. De betreffende informatie is te vinden in tabel 6.6.

Tabel 6.6 Verschillen tussen leerlingen naar sekse voor de onderdelen Rekenen, Lezen en Taalverzorging, toetsing op significantie en effectgrootte

Rekenen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
Jongens	62.9	15.8	28015	925.5	0
Meisjes	58.8	15.9	28021	P<.001	-.26
Lezen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
Jongens	27.0	6.7	28052	456.6	0
Meisjes	28.2	6.4	28111	P<.001	.18
Taalverzorging					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
Jongens	32.8	7.3	28041	1359.6	0
Meisjes	34.9	6.7	28105	P<.001	.31

De resultaten zijn overeenkomstig de verwachtingen. Jongens doen het beter op rekenen (klein effect), meisjes beter op taal (lezen en taalverzorging, met verwaarloosbare tot kleine effecten). Alle verschillen zijn statistisch significant ($p < .001$).

Verschillen naar leeftijd

Niet alle leerlingen in groep 7 zijn even oud. Zijn er effecten van de verschillen in leeftijd op de hoogte van de scores op de Entreetoets? In tabel 6.7 staat de verdeling naar leeftijd van de leerlingen die in 2016 hebben deelgenomen aan de toets waarbij de scheiding in leeftijdsgroepen is gelegd op de peildatum van 1 oktober. Leerlingen die (tot op dat moment) zonder vertraging hun schoolloopbaan volbrachten en in juni 2016 in groep 7 zaten, zijn geboren tussen 1 oktober 2004 en 30 september 2005. Ze vormen de grote groep reguliere leerlingen. Oudere leerlingen vormen de groep leerlingen die om een of andere reden vertraging heeft opgelopen, deze leerlingen zijn bijvoorbeeld later begonnen in groep 3 of zijn gedoubleerd. Jongere leerlingen zijn 'voorlijke' leerlingen die waarschijnlijk eerder aan groep 3 begonnen zijn of die een klas hebben overgeslagen. De verwachting is dat de vertraagde leerlingen gemiddeld lager scoren dan de reguliere leerlingen. De 'voorlijke' leerlingen zullen naar verwachting gemiddeld wat hoger scoren, al zal het verschil niet zo hoog zijn als dat tussen vertraagde en reguliere leerlingen.

Tabel 6.7 Verschillen tussen leerlingen naar leeftijd voor de hoofdonderdelen Rekenen, Lezen en Taalverzorging, toetsing op significantie en effectgrootte

Rekenen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
voorlijk	65.9	14.3	5956	2054.9	.25
regulier	62.0	15.5	42310	P<.001	0
vertraagd	50.8	15.6	7642		-.73
Lezen					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
voorlijk	29.4	5.9	5962	1696.1	.22
regulier	28.0	6.4	42391	P<.001	0
vertraagd	23.8	6.5	7679		-.66
Taalverzorging					
<i>Categorie</i>	<i>Gemiddelde</i>	<i>SD</i>	<i>N</i>	<i>F</i>	<i>Cohens d</i>
voorlijk	36.4	6.4	5957	2148.6	.31
regulier	34.3	6.9	42381	P<.001	0
vertraagd	29.4	6.8	7677		-.72

Alle resultaten zijn conform de verwachtingen en significant. Effectgroottes zijn klein voor het verschil tussen 'voorlijke' en reguliere leerlingen, en middelgroot voor het verschil tussen reguliere en vertraagde leerlingen.

Daarmee zijn alle verschillen tussen categorieën leerlingen naar leerlinggewicht, sekse en leeftijd in overeenstemming met de (vooraf) geformuleerde verwachtingen. De resultaten vormen daarmee een extra ondersteuning voor de begripsvaliditeit van de Entreetoets groep 7.

6.2.5 Verschillen tussen subgroepen: intercorrelatiematrix en DIF-analyses

Invariantie van de structuur

In paragraaf 6.2.2 is het nodige gezegd over de structuur van de Entreetoets. We deden dat door de intercorrelatiematrix te bespreken en door na te gaan of het patroon van de intercorrelaties overeenkwam met wat we mochten verwachten op basis van theorie en ervaringen in het verleden. In verband met de vergelijkbaarheid van toetsresultaten voor verschillende subgroepen van leerlingen, zou het de voorkeur verdienen wanneer de structuur van de toets voor deze subgroepen ongeveer gelijk zou zijn (invariantie van structuur). We zijn dit nagegaan voor de intercorrelaties tussen de hoofdonderdelen, en wel voor categorieën met een verschillend leerlinggewicht en voor jongens versus meisjes (zie tabel 6.8).

De correlaties zijn weergegeven voor zowel de basistoets als de basistoets met de module Verdieping.

Tabel 6.8 Intercorrelatiematrices voor subgroepen van leerlingen naar leerlinggewicht en sekse voor module basistoets en basistoets en module Verdieping

Basistoets									
Intercorrelatiepatronen voor groepen naar leerlinggewicht									
Leerlinggewicht 0.0									
	<i>R</i>	<i>L</i>	<i>T</i>	Leerlinggewicht 0.30			Leerlinggewicht 1.20		
	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>
<i>Rekenen (R)</i>	--	.67	.59	--	.62	.54	--	.64	.59
<i>Lezen (L)</i>		--	.61		--	.59		--	.60
<i>Taalverzorging (T)</i>			--			--			--
Intercorrelatiepatronen jongens versus meisjes									
Jongens									
	<i>R</i>	<i>L</i>	<i>T</i>	Meisjes					
	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>			
<i>Rekenen (R)</i>	--	.70	.63	--	.70	.63			
<i>Lezen (L)</i>		--	.62		--	.61			
<i>Taalverzorging (T)</i>			--			--			
Basistoets en Verdieping									
Intercorrelatiepatronen voor groepen naar leerlinggewicht									
Leerlinggewicht 0.0									
	<i>R</i>	<i>L</i>	<i>T</i>	Leerlinggewicht 0.30			Leerlinggewicht 1.20		
	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>
<i>Rekenen (R)</i>	--	.73	.67	--	.68	.60	--	.67	.62
<i>Lezen (L)</i>		--	.68		--	.65		--	.68
<i>Taalverzorging (T)</i>			--			--			--
Intercorrelatiepatronen jongens versus meisjes									
Jongens									
	<i>R</i>	<i>L</i>	<i>T</i>	Meisjes					
	<i>R</i>	<i>L</i>	<i>T</i>	<i>R</i>	<i>L</i>	<i>T</i>			
<i>Rekenen (R)</i>	--	.71	.63	--	.74	.68			
<i>Lezen (L)</i>		--	.68		--	.68			
<i>Taalverzorging (T)</i>			--			--			

Per achtergrondkenmerk is er niet veel verschil te zien voor de van links naar rechts onderscheiden subgroepen. De verschillen zijn niet groter dan .05; alleen voor de basistoets en de module Verdieping is het verschil in correlatie tussen Rekenen en Taalverzorging voor de leerlingen met een gewicht van 0.0 en de leerlingen met een gewicht van 0.30 iets groter (.67 versus .60). Ook is de correlatie tussen Rekenen en Lezen per matrix steeds het grootst, met één uitzondering voor leerlingen met een leerlinggewicht van 1.20- bij de basistoetsen en de module Verdieping. Kortom, de verschillen in structuur zijn verwaarloosbaar en voor zover er sprake is van lichte afwijkingen, zijn deze niet te interpreteren.

DIF-analyses

De bruikbaarheid van een leervorderingstoets zoals de Entreetoets groep 7 voor verschillende doelgroepen heeft grote implicaties voor de (onderwijs)toekomst van grote groepen leerlingen. Uiterwijk (1994) en Van Schilt-Mol (2007) hebben dit probleem geformuleerd in termen van 'Differential Item Functioning' (DIF). Daarmee doelen zij op het verschijnsel dat toetsen opgaven kunnen bevatten die verschillend 'functioneren' voor verschillende subgroepen leerlingen, zelfs wanneer deze leerlingen een vergelijkbaar prestatieniveau laten zien. In de IRT houdt dit concreet in dat leerlingen die behoren tot verschillende categorieën (bijvoorbeeld jongens versus meisjes, leerlingen behorend tot groepen met een verschillend leerlinggewicht) dezelfde kans moeten hebben om op een item met een bepaalde respons te reageren. Er is sprake van DIF (ook wel aangeduid als 'onbedoelde moeilijkheden') als leerlingen uit verschillende subgroepen met een vergelijkbaar prestatieniveau een ongelijke kans hebben om een

toetsopgave juist te beantwoorden. Wanneer de oorzaak daarvan niet behoort tot het construct dat de opgave beoogt te meten, is er sprake van itembias.

Uit eerder onderzoek naar de Eindtoets Basisonderwijs (Uiterwijk, 1994; Uiterwijk & Vallen, 1997, 2005) is gebleken dat de Eindtoets (in jaren voorafgaand aan de genoemde publicatiejaren) opgaven bevatte waarbij sprake is van DIF en dus mogelijk sprake is van itembias. Van Schilt-Mol (2007) heeft nauwkeurig onderzoek verricht naar DIF en itembias voor Turkse en Marokkaanse leerlingen in de Eindtoets van 1997. De analyses leidden tot de conclusie dat de Eindtoets leerlingen van allochtone afkomst niet benadeelt en ook niet bevoordeelt, althans voor zover het de onderzochte versie betreft. Ondanks deze (voor de versie van 1997) gunstige uitkomsten heeft Cito inmiddels een aantal opgaventypen aangepast, ook in de Entreetoets. Het onderzoek van Van Schilt-Mol bevatte daartoe een aantal bruikbare aanwijzingen en aanbevelingen. Ook werd het onderzoek waarop de selectie van opgaven is gebaseerd zodanig uitgebreid dat het nog beter mogelijk is de selectie van opgaven met onbedoelde moeilijkheden te voorkomen. Dit laatste geldt ook voor de opgaven die voor de Entreetoets groep 7 werden geconstrueerd.

Ondanks de zorgvuldige itemconstructie werd besloten om DIF-analyses uit te voeren voor sekse en leerlinggewicht. Helaas was geen informatie beschikbaar over de etnische herkomst of thuistaal van de leerlingen. Maar omdat de groepen met afwijkende leerlinggewichten (0.30 en 1.20) relatief veel leerlingen van allochtone herkomst herbergen, kunnen we de analyse naar leerlinggewicht beschouwen als een bruikbaar alternatief.

Op alle 180 verplichte opgaven (basistoets) van de Entreetoets groep 7 zijn DIF-analyses uitgevoerd die vergelijkbaar zijn met de analyses van Van Schilt-Mol (2007) (zie hierboven). Dit gebeurde op 'klassieke' wijze door toepassing van de Mantel Haenszel procedure. Ook de uitgebreidere verdiepende module onderwierpen we aan dit type analyses. Hierbij wordt per item en per vergelijking tussen subgroepen de Mantel Haenszel (MH) DIF-statistiek berekend en op significantie getoetst. We voerden de analyses uit op de gehele ruwe datasets voor de 2016-afnames. Het probleem is dan dat toetsen op significantie niet erg zinnig is vanwege de zeer grote aantallen leerlingen. Een alternatief voor toetsen is het berekenen van effectgroottes die onafhankelijk zijn van de steekproefomvang. We hebben daartoe de MH DIF statistiek via het pakket R (procedure 'difR') omgezet in een zogeheten DeltaMH-schaal. Hierover valt in de handleiding van het pakket het volgende te lezen: "In addition, the Mantel-Haenszel estimates of the common odds ratios α_{MH} are used to measure the effect sizes of the items. These are obtained by $\Delta_{MH} = -2.35 \log \alpha_{MH}$ (Holland and Thayer, 1985). According to the ETS delta scale, the effect size of an item is classified as negligible if $|\Delta_{MH}| \leq 1$, moderate if $1 < |\Delta_{MH}| \leq 1.5$, and large if $|\Delta_{MH}| \geq 1.5$. The values of the effect sizes, together with the ETS classification, are printed with the output."

In overeenstemming met deze richtlijnen zijn in de tabellen 6.9 en 6.10 waarin de resultaten van de analyses zijn opgenomen, de schaalwaarden als volgt gedefinieerd en gecodeerd:

A = verwaarloosbaar ($|\Delta_{MH}| \leq 1$)

B = matig ($1 < |\Delta_{MH}| \leq 1.5$)

C = groot ($|\Delta_{MH}| > 1.5$)

Aantallen A-waarden zijn ongeacht de richting van het verschil aangeduid met de noemer "Geen DIF". Voor de B- en C-waarden is ook de richting van het verschil in de tabel opgenomen in de vorm van een plus- of minteken.

Tabel 6.9 DIF naar sekse

Onderdeel	Rekenen		Lezen		Taalverzorging	
	B	V	B	V	B	V
Geen DIF	81	129	37	79	50	89
-B	3	7	2	0	0	0
+B	4	3	0	0	0	1
-C	-	-	1	1	0	0
+C	2	1	0	0	0	0
Totaal	90	140	40	80	50	90
% items zonder DIF	90.0%	92.1%	92.5%	98,8%	100%	98.9%

Resultaten van het onderzoek van DIF naar sekse staan in tabel 6.9. Daarbij vormen de meisjes steeds de focale groep. Dit impliceert dat een negatieve waarden (-B of -C) te interpreteren is als matige, respectievelijk grote DIF die in het nadeel van de meisjes uitvalt. Positieve waarden impliceren op analoge wijze DIF in het nadeel van jongens. De aantallen items zijn afzonderlijk vermeld voor de basistoets en de basistoets en de module Verdieping onder 'respectievelijk de noemers 'B' en 'V. Ze worden niet opgeteld: bij de leerling wordt immers de basistoets afgenomen óf de basistoets aangevuld met de module Verdieping, niet beide. De items met forse DIF (+C en -C) zijn extra gemarkeerd door middel van een vetgedrukte omkadering. Cellen met positieve waarden (in het nadeel van de jongens) zijn gearceerd, cellen met negatieve waarden (in het nadeel van de meisjes) niet.

De meeste DIF-items zijn te vinden bij *Rekenen*: 10% van de rekenopgaven in de basistoets bevat matige of grote DIF; in de basistoets aangevuld met de module Verdieping is dat 7.9%. Dit is niet onverwacht; in de vorige versie van de Entreetoets betrof dit 10% van de verplichte opgaven. Bij meeste DIF-items is sprake van matige DIF. Bovendien is de DIF soms in het nadeel van de meisjes, soms in het nadeel van de jongens; wel zijn jongens relatief wat vaker in het nadeel. Het aantal opgaven met grote DIF is beperkt: slechts twee in de basistoets (2.2%) en maar één in de basistoets en module Verdieping (.7%).

In de taalonderdelen komt weinig DIF voor. In de basistoets troffen we voor Lezen twee opgaven aan met matige DIF en één met grote DIF, steeds in het nadeel van meisjes (7.5% van de items). In de basistoets en module Verdieping is er slechts één opgave met grote DIF, eveneens in het nadeel van meisjes (1.3%). In tegenstelling tot de situatie bij Rekenen komen meisjes dus bij het onderdeel Lezen iets vaker DIF tegen in hun nadeel dan jongens. Het onderdeel Taalverzorging kent nauwelijks DIF; bij slechts één opgave in de basistoets en module Verdieping is sprake van matige DIF (1.1%). Ter vergelijking: in de vorige versie van de Entreetoets was er bij 7.7% van de talige opgaven sprake van significante DIF.

Tabel 6.10 DIF naar leerlinggewicht

	Rekenen		Lezen		Taalverzorging	
	B	V	B	V	B	V
Geen DIF	89	137	40	78	50	90
-B	1	3	0	2	0	0
+B	0	0	0	0	0	0
-C	0	0	0	0	0	0
+C	0	0	0	0	0	0
Totaal	90	140	40	80	50	90
% items zonder DIF	98.8%	97.9%	100%	97.5%	100%	100%

Resultaten van het onderzoek van DIF naar leerlinggewicht staan in tabel 6.10. In dit onderzoek zijn de beide categorieën met een afwijkend leerlinggewicht (0.30 en 1.20) samengenomen. De resulterende categorie (met afwijkend leerlinggewicht) is gedefinieerd als focale groep. Dat houdt in dat negatieve waarden moeten worden geïnterpreteerd als in het nadeel van deze groep.

Ten aanzien van leerlinggewicht is er weinig DIF vast te stellen. Slechts één van de 180 opgaven in de basistoets (bij Rekenen) vertoont matige DIF (dat is .6% van de opgaven). In de basistoets en module Verdieping ligt dit percentage met 1.9% wat hoger: drie opgaven van Rekenen en twee van Lezen in de module Verdieping laten matige DIF zien. Alle DIF-opgaven laten een effect zien in het nadeel van leerlingen met een afwijkend leerlinggewicht. Doordat een leerling maar weinig DIF-items tegenkomt (één, respectievelijk vijf opgaven in de basis- en de uitgebreide versie), waarbij het bovendien gaat om matige vormen van DIF, blijft het effect op de scores zeer beperkt.

De DIF-analyses gaven geen aanleiding om items buiten beschouwing te laten in de scoring. Leerlingen die van elkaar verschillen in leerlinggewicht of geslacht worden immers in de Entreetoets groep 7 niet op oneigenlijke wijze benadeeld door de opgaven, dat wil zeggen door itemkenmerken die met de te meten vaardigheid zelf niets van doen hebben.

6.3 Criteriumvaliditeit

De score op de basistoets van de Entreetoets groep 7 wordt gebruikt om een advies te onderbouwen over het best passende brugklatype voor de leerling (of meerdere best passende brugklatypes indien daarvan sprake is); zie verder paragraaf 4.4.1 voor een beschrijving van de werkwijze en voor een beschrijving van het Rapport Vooruitblik dat is ontworpen om te communiceren over het geadviseerde of de geadviseerde brugklatype(n). Voor een afbeelding van het Rapport Vooruitblik verwijzen we naar figuur 4.5. Om de Entreetoets groep 7 op deze manier in predictieve zin te kunnen gebruiken, is het nodig te verantwoorden dat dit gebruik van een solide onderbouwing is voorzien. In paragraaf 6.3.1 plaatsen we het predictieve gebruik van de Entreetoets groep 7 in historisch perspectief en laten we zien wat over de kwaliteit hiervan eerder is gerapporteerd. Ook eerder al was deze brede toepassing gebaseerd op de assumptie van equivalentie tussen verschillende toetsversies. In paragraaf 6.3.2 laten we zien wat dit betekent en zullen we ook evidentie presenteren met betrekking tot de hier verantwoorde verise van de Entreetoets groep 7 en diens voorloper. In paragraaf 6.3.3, ten slotte, bespreken we de resultaten van toelatings- en doorstroomonderzoek die dienen ter onderbouwing van de criteriumvaliditeit.

6.3.1 Het predictieve gebruik van de Entreetoets groep 7 in historisch perspectief

In de wetenschappelijke verantwoording van de vorige versie van de Entreetoets groep 7 uit 2012; (van Boxtel et al., 2013) is besproken hoe de adviezen die gebaseerd waren op de Entreetoetsscore zich verhielden tot de adviezen op basis van de Cito Eindtoets Basisonderwijs van een jaar later. De overlap in adviezen was aanzienlijk. In bijna 60% (59.5) van de gevallen kwamen het geschatte en het uiteindelijk afgegeven advies exact met elkaar overeen. In nog eens 36.5% van de gevallen wijken de adviezen niet meer dan één adviescategorie van elkaar af. Daarmee kan de voorspellende waarde van deze versie van de vorige versie van de Entreetoets groep 7 uit 2012 als zeer goed worden aangemerkt. De genoemde samenhang tussen in wezen twee inhoudelijk geheel verschillende instrumenten (geen enkele opgave van beide toetsen was hetzelfde) die met een interval van bijna een jaar werden afgenomen, had als indicator van de test-hertestbetrouwbaarheid (voor één en hetzelfde instrument met een interval van – bijvoorbeeld – enkele weken), absoluut niet misstaan.

Laatstgenoemde correlatie van .88 is overigens niet irrelevant, omdat deze is berekend met de vorige Entreetoets groep 7 uit 2012 versie die we vóór de huidige Entreetoets groep 7 hebben gebruikt om toelatings- en doorstroomonderzoek (TDO) uit te voeren in 2013. De resultaten van dit TDO kunnen gelden ter onderbouwing van de criteriumvaliditeit van de hier verantwoorde versie. Dit hebben we mogelijk gemaakt door equivalentie tot stand te brengen tussen de huidige versie en de vorige versie uit 2012. Over deze equivalering in de volgende paragraaf meer. We kunnen in ieder geval alvast constateren dat de voorspellende waarde van de Entreetoets groep 7 uit 2012 ten aanzien van latere toetsresultaten bijzonder goed was. Wat we nog nooit eerder hebben kunnen verwezenlijken is het rechtstreeks koppelen van Entreetoetsscores aan de feitelijke instroom in het voortgezet onderwijs. In paragraaf 6.3.3 zullen we de resultaten bespreken van dit onderzoek.

6.3.2 Equipercntiel-equivalering

De Entreetoets groep 7 uit 2012 is, in 2013, gebruikt om een relatie te leggen tussen de score op de Entreetoets groep 7 en de feitelijke instroom in het voortgezet onderwijs. Om deze resultaten te kunnen gebruiken voor het maken van het Rapport Vooruitblik voor de huidige Entreetoets is gebruikgemaakt van een equipercntiel-equivalering. Met behulp van deze equipercntiel-equivalering kunnen de grenzen van de adviescategorieën van de voorganger van de Entreetoets groep 7, omgezet worden op de scoreschaal van de basistoets van de hier verwantwoorde Entreetoets groep 7.

Deze equipercntiel-equivalering is toe te passen wanneer beide Entreetoetsversies ontwikkeld zijn volgens dezelfde toetsspecificaties. De beide versies van de Entreetoetsen verschillen echter op een belangrijk punt van elkaar. Het Rapport Vooruitblik bij de huidige Entreetoets is gebaseerd op de onderdelen die aan bod komen in de basistoets, terwijl het bij zijn voorganger gebaseerd werd op alle verplichte onderdelen. Dit wil zeggen dat de onderdelen Woordenschat en Schrijven bij de huidige Entreetoets groep 7 niet worden meegenomen in het bepalen van de score voor het Rapport Vooruitblik terwijl dit bij zijn voorganger wel het geval was. Deze verschillen maken het noodzakelijk om aan te tonen dat er een hoge samenhang is tussen de score op Entreetoets groep 7 uit 2012 en de score op de Entreetoets groep 7 2012 waarbij de onderdelen Woordenschat en Schrijven niet meegenomen zijn. De correlatie tussen beide scores is .98, waaruit blijkt dat de samenhang tussen de scores erg hoog is. Het gebruik van de equipercntiel-equivalering is daarmee gerechtvaardigd.

6.3.3 Resultaten van het toelatings- en doorstroomonderzoek

In het geval van eindtoetsen die in het basisonderwijs worden gebruikt om het advies met betrekking tot het best passende brugklatype te onderbouwen, is het gebruikelijk dat informatie over de voorspellende waarde van een eerdere versie van de bewuste toets wordt verzameld om deze te kunnen toepassen op de meest recente versie. Want een nieuwe versie van zo'n eindtoets is zelf nog nooit eerder afgenomen (afgezien van de elementen ervan waarover in proeftoetsen informatie is verzameld), laat staan dat er follow-up onderzoek is uitgevoerd waarin is vastgesteld naar welk brugklatype de leerling is uitgestroomd. Kortom, het is noodzakelijk om onderzoek uit te voeren met een voorganger om informatie te verzamelen,

deze van toepassing te maken op de huidige afnameversie en de verzamelde informatie te benutten om adviescategorieën op te stellen die gelden bij de rapportage op het moment dat de huidige versie is afgenomen. Deze werkwijze is in het verleden toegepast op de Cito Eindtoetsen Basisonderwijs en wordt momenteel ook met succes toegepast toegepast op de Centrale Eindtoets (zie CvTE, in voorbereiding).

Een noodzakelijke voorwaarde om deze procedure te kunnen uitvoeren is, dat er sprake is van equivalentie tussen elke toetsversie en zijn voorganger. Pas als bewezen is dat de assumptie van equivalentie houdbaar is, is het mogelijk om de bedoelde informatie daadwerkelijk toe te passen. Het succes van de werkwijze is rechtstreeks afhankelijk van de kwaliteit van de equivalering.

Deze werkwijze is ook toegepast op de hier verantwoorde versie van de Entreetoets groep 7. Voor de onderbouwing van de equivalering verwijzen we naar de informatie die we in de vorige paragraaf hebben gepresenteerd.

Tabel 6.11 *Instroom in brugklas VO in relatie tot advies op basis van Entreetoets groep 7 in absolute aantallen (N = 86565)*

Advies ET7	Grenzen	Instroom in brugklatype voortgezet onderwijs								Totaal
		bb	bb/kb	kb	gt	gt/h	havo	h/v	vwo	
bb	0 – 79	2428	2062	783	307	96	12	20	9	5717
bb/kb	79 – 97	1588	2607	2447	1974	765	38	88	17	9524
kb	97 – 104	172	662	827	1275	595	53	77	10	3671
gt	104 – 113	180	907	1145	3699	2295	213	553	43	9035
gt/havo	113 – 127	39	445	569	4315	4899	1103	3008	217	14595
havo	127 – 133	10	70	65	964	1835	826	3037	351	7158
havo/vwo	133 – 148	5	42	42	598	1799	1652	9343	2825	16306
vwo	148 – 180	3	13	10	93	253	443	7164	12580	20559
Totaal		4425	6808	5888	13225	12537	4340	23290	16052	86565

Tabel 6.12 *Instroom in brugklas VO in relatie tot advies op basis van Entreetoets groep 7 in percentages (N = 86565)*

Advies ET7	Grenzen	Instroom in brugklatype voortgezet onderwijs								Totaal
		bb	bb/kb	kb	gt	gt/h	havo	h/v	vwo	
bb	0 – 79	3%	2%	1%	0%	0%	0%	0%	0%	5717
bb/kb	79 – 97	2%	3%	3%	2%	1%	0%	0%	0%	9524
kb	97 – 104	0%	1%	1%	1%	1%	0%	0%	0%	3671
gt	104 – 113	0%	1%	1%	4%	3%	0%	1%	0%	9035
gt/havo	113 – 127	0%	1%	1%	5%	6%	1%	3%	0%	14595
havo	127 – 133	0%	0%	0%	1%	2%	1%	4%	0%	7158
havo/vwo	133 – 148	0%	0%	0%	1%	2%	2%	11%	3%	16306
vwo	148 – 180	0%	0%	0%	0%	0%	1%	8%	15%	20559
Totaal		4425	6808	5888	13225	12537	4340	23290	16052	86565

De meest recente versie waarmee toelatings- en doorstroomonderzoek is uitgevoerd, is de versie die (onder meer) in 2012 is afgenomen (dus in het schooljaar 2011-2012). Het is de versie die we in 2013 hebben verantwoord (van Boxtel et al., 2013). De betreffende leerlingen zaten in het schooljaar 2012-2013 in groep 8 en zijn in augustus/september 2013 ingestroomd in het voortgezet onderwijs. TDO is het resultaat van samenwerking tussen Cito en CBS (Centraal Bureau voor de Statistiek). Er zijn intussen meer recente toelatings- en doorstroomonderzoeken uitgevoerd, maar daarvan zijn de resultaten nog niet beschikbaar. Ook met de versie van de Entreetoets groep 7 die we hier verantwoorden is inmiddels toelatingsonderzoek gaande. Hierover is in paragraaf 4 gerapporteerd. Eerder in dit hoofdstuk hebben we beschreven hoe de grenzen van adviescategorieën zijn overgebracht van de vorige versie van de Entreetoets groep 7 uit 2012 (met een scoreschaal die varieerde van 0-420) naar de basistoets van de Entreetoets groep 7 die een scoreschaal kent die loopt van 0-180. In tabel 6.11 worden dezelfde grensscores gebruikt om de adviescategorieën af te bakenen. Deze zijn rijgewijs weergegeven. Kolomgewijs staan het brugklatype beschreven waarin een leerling is ingestroomd in het voortgezet onderwijs. Gegevens zijn verzameld over 86565 leerlingen. Tabel 6.12 kent dezelfde opbouw, maar nu op basis van percentages. Uit de tabel blijkt dat er in 37% van de gevallen sprake is van exacte overeenstemming (het totaal van de donker gearceerde cellen op de diagonaal). Rekenen we ook aanpalende brugklatype als correct, dan stijgt het percentage naar 79% (het totaal van alle gearceerde cellen). We kunnen de percentages ook conditioneel weergeven. De percentages tellen dan rijgewijs op tot 100 (zie tabel 6.13).

Tabel 6.13 *Instroom in brugklas VO in relatie tot advies op basis van Entreetoets groep 7 in percentages (N = 86565) conditioneel op het gegeven advies*

Advies ET7	grenzen	Instroom in brugklatype voortgezet onderwijs								Totaal
		bb	bb/kb	kb	gt	gt/h	havo	h/v	vwo	
bb	0 – 79	42%	36%	14%	5%	2%	0%	0%	0%	5717
bb/kb	79 – 97	17%	27%	26%	21%	8%	0%	1%	0%	9524
kb	97 – 104	5%	18%	23%	35%	16%	1%	2%	0%	3671
gt	104 – 113	2%	10%	13%	41%	25%	2%	6%	0%	9035
gt/havo	113 – 127	0%	3%	4%	30%	34%	8%	21%	1%	14595
havo	127 – 133	0%	1%	1%	13%	26%	12%	42%	5%	7158
havo/vwo	133 – 148	0%	0%	0%	4%	11%	10%	57%	17%	16306
vwo	148 – 180	0%	0%	0%	0%	1%	2%	35%	61%	20559
Totaal		4425	6808	5888	13225	12537	4340	23290	16052	86565

Uit de tabel blijkt (bijvoorbeeld), dat van de 20559 leerlingen die een brugklasadvies 'vwo' kregen, 61% daadwerkelijk is ingestroomd in het vwo. Daarnaast kwam 35% terecht in een brugklatype havo/vwo, een plaatsing die als conform het advies kan worden betiteld. Daarnaast stroomde 2% door naar de havo en 1% naar brugklas gt/havo. De rest van de tabel is op soortgelijke wijze te interpreteren. De 'average conditional accuracy' is 37%. Rekenen we de aanpalende cellen op de diagonaal als correct, dan bedraagt de 'average conditional accuracy + 1' 79%. De gewogen kappa bedraagt .67.

De tabellen laten zien dat er voldoende evidentie is voor de criteriumvaliditeit van de Entreetoets groep 7, zeker wanneer men zich realiseert dat er een volledig schooljaar zit tussen de bepaling van het best passende brugklatype en de feitelijke instroom in het voortgezet onderwijs. Daarnaast houden de tabellen geen rekening met het feit dat er in veel gevallen meerdere brugklatypes tegelijk als best passend worden gerapporteerd.

7 Samenvatting

Deze wetenschappelijke verantwoording betreft de Entreetoets groep 7; de analyses zijn uitgevoerd op de gegevens van de afname van deze toets in 2016.

De Entreetoets groep 7 is een brede school- of leervorderingentoets waarvan de inhoud nauw aansluit bij het onderwijsprogramma voor de basisvaardigheden op het gebied van taal, rekenen en wereldoriëntatie. De opgaven zijn een operationalisering van doelstellingen die door het gehele basisonderwijs worden nagestreefd, zogenoemde communale doelstellingen. Voor zover het om taal en rekenen gaat sluit de inhoud van de toets aan op het Referentiekader Taal en Rekenen.

De opzet van de hier verantwoorde Entreetoets wijkt nadrukkelijk af van die van zijn voorganger.

De gebruiker kan kiezen tussen drie verschillende modules.

De module *Basis* (basistoets) vormt het hart van de toets. De inhoud van deze module heeft betrekking op drie belangrijke elementen in het Referentiekader Nederlandse Taal en Rekenen, te weten Rekenen en de talige onderdelen Lezen en Taalverzorging. Voor Rekenen zijn vier domeinen relevant: Getallen, Meten en meetkunde, Verbanden en Verhoudingen. Bij *Lezen* staan de domeinen Begrijpend lezen, Opzoeken en Samenvatten centraal. Bij alle drie deze talige domeinen ligt de focus op tekstbegrip en semantiek.

Bij *Taalverzorging* is dit niet of veel minder het geval. Het gaat bij Taalverzorging vooral om het kunnen produceren van een verzorgde tekst, door het correct kunnen toepassen van regels en strategieën. Bij alle vier de domeinen die we hierbij onderscheiden (Grammatica, Interpunctie, Spelling niet-werkwoorden en Spelling werkwoorden) spelen aangeleerde automatisen een grote rol.

De gebruiker die een uitvoerigere en gedetailleerdere rapportage wenst over de genoemde domeinen, kan de *basistoets en Verdieping* afnemen. Door in de deeltaetsen meer opgaven onder te brengen dan in de basistoets, is het mogelijk gedifferentieerder en met adequate betrouwbaarheid te rapporteren en op die manier een gedetailleerd beeld te schetsen van de huidige leervorderingen.

Door afname van ook de module Verbreding kan dit toch al gedetailleerde beeld nog verder worden uitgebreid. Uitbreiding is mogelijk naar de talige domeinen Luisteren, Schrijven en Woordenschat. Deze domeinen kennen alle een semantisch karakter; betekenisverlening is bij alle drie van belang. Daarnaast is sprake van verbreding in de richting van Wereldoriëntatie, kort te omschrijven als 'kennis van de wereld', met de domeinen Aardrijkskunde, Geschiedenis en Natuur en techniek. Bij het verwerven van deze kennis doet de leerling een beroep op bepaalde talige en grafische studievaardigheden en bouwt hij deze vaardigheden verder uit.

De Entreetoets groep 7 is bedoeld voor leerlingen in groep 7 van het primair en speciaal onderwijs. Voor leerlingen die nog maar pas in Nederland verblijven en de Nederlandse taal onvoldoende beheersen, zijn de toetsen ongeschikt. Leerlingen met een vertraagde ontwikkeling (achterstand van een jaar of meer) kunnen beter de Entreetoets groep 6 maken; deze toets past beter bij hun vaardigheidsniveau. Voor leerlingen met een visuele beperking of dyslexie zijn diverse aangepaste versies beschikbaar. Afname van de toets vindt hoofdzakelijk plaats in de periode april – juni, maar is ook mogelijk begin groep 8 in de maanden augustus en september.

Het hoofddoel van de Entreetoets groep 7 is het *beschrijven* van het niveau van leerlingen op het gebied van Rekenen, Lezen en Taalverzorging zoals omschreven in het referentiekader Nederlandse Taal en Rekenen. Dit kan in meer of minder gedetailleerde, en al dan niet uitgebreide vorm. Afname van de Entreetoets groep 7 past ook in de functie van het Cito Volgstelsel primair en speciaal onderwijs (het *volgen* van leerlingen in hun groei, ook wel aangeduid als *progressiebepaling*). Scores op bepaalde deeltaetsen kunnen als vervanger gelden voor de desbetreffende LVS-toets of –toetsen. Bij deze beschrijvende en volgende functies passen twee typen normering, namelijk een relatieve normering in termen van percentielscores en niveau-scores (I tot en met V) en een domeingeoriënteerde normering op de referentieniveaus 1F en 2F (dan wel 1S).

Naast deze functies kent de Entreetoets ook een voorspellend gebruik. De totaalscore op de basistoets vormt de grondslag voor een advies over het best passend brugklatype (of de best passende

brugklastype) na afsluiting van het basisonderwijs. In dit kader is de criteriumgerichte normering van belang die is onderbouwd door toelatings- en doorstroomonderzoek in het voortgezet onderwijs.

In hoofdstuk 2 en 3 is de toets uitgebreid beschreven en inhoudelijk verantwoord. Daarbij is veel aandacht besteed aan inhoudelijke samenstelling van de onderdelen van de toets. *Het uitgangspunt voor de constructie van de Entreetoets 7* was, dat deze de communale onderwijsdoelstellingen voor het einde van leerjaar 7 zou moeten representeren. Op grond van een gedetailleerde analyse hiervan is de toetsinhoud minutieus voorgeschreven aan de opgavenconstructeurs. In de inhoudsverantwoording konden we in grote lijnen terugrijpen op de tekst die ook al in de leerkrachthandleiding is gepubliceerd. We hebben deze tekst aangevuld met matrixen waarin steeds werd aangegeven wat de meest wenselijke spreiding van opgaven over opgaventypen was, met daarnaast de daadwerkelijk gerealiseerde verdeling. Uit de vergelijking van gewenste en gerealiseerde inhoud kan worden afgeleid dat de Entreetoets groep 7 een inhoudsvalide toets is. Daarnaast is weergegeven hoeveel opgaven van elk type in de basistoets en de basistoets aangevuld met de module Verdieping zijn opgenomen ten behoeve van de (domeingerichte) normering op referentieniveau. De betreffende overzichten laten zien dat ook in dit opzicht de inhoudsvaliditeit op orde is. Voor de Entreetoets groep 7 is niet alleen de inhoudsvaliditeit belangrijk, ook constructvaliditeit speelt een essentiële rol. Tot de uitgangspunten voor de constructie behoren ook meet-technische principes. Aan de onderscheiden (deel-)toetsen ligt een itemresponsmodel ten grondslag dat veronderstelt dat elke (deel-)toets een uni-dimensionale latente trek of vaardigheid representeert. We hebben ruim aandacht geschonken aan de beschrijving van dit meetmodel (OPLM) en aan de stappen die in het kader van geslaagde kalibraties moesten worden genomen om adequate opgavenbanken te creëren.

In het verlengde van de inhoudsverantwoording werd – rekeninghoudend met de verwachte mortaliteit – een veelvoud van het aantal benodigde opgaven geconstrueerd. Dit gebeurde in constructiegroepen van ervaren leerkrachten van groep 7 of 8, PABO-docenten en/of IB'ers, onder leiding van een toetsdeskundige van Cito. In totaal werden 1174 nieuwe opgaven geconstrueerd, die allemaal werden uitgetest in proefonderzoek. Het doel van de proeftoetsing was de selectie van de meest geschikte items, ankering naar de Entreetoets groep 6 en naar de vorige versie van de Entreetoets groep 7 (dit laatste met het oog op equivalering in verband met het vaststellen van de criteriumvaliditeit), voor sommige onderdelen ook een ankering naar de betreffende LVS-toetsen en ten slotte ankering naar referentiesets met het oog op het vaststellen van referentieniveaus voor Rekenen, Lezen en Taalverzorging (overbrenging van de referentiecesuur). Om dit alles mogelijk te maken werd een groot aantal bestaande opgaven in het proefonderzoek opgenomen en werd een complex afnamedesign ontwikkeld. Om zoveel opgaven uit te testen en met elkaar te kunnen verbinden waren zeer veel scholen en leerlingen nodig. In totaal toonden in 2015 206 scholen met 5035 leerlingen zich bereid om deel te nemen; 189 scholen deden daadwerkelijk mee. Representativiteit is voor zo'n proefonderzoek van ondergeschikt belang: een zekere spreiding over achtergrondkenmerken volstaat. Uit controleanalyses blijkt dat de onderzoeksgroep aan die voorwaarde voldoet. Ook blijkt dat elk van de in totaal 2070 opgaven in het proefonderzoek door voldoende leerlingen is gemaakt om nauwkeurig te kunnen kalibreren; de minimumaantallen leerlingen variëren van 364 tot 685.

In hoofdstuk 4 deden we verslag van de kalibraties en de normering.

De Entreetoets groep 7 kent drie typen normering, (1) een relatieve normering in de vorm van percentielscores (en een op deze percentielscores gebaseerde indeling in niveaugroepen I – V), (2) een domeingeoriënteerde normering naar referentieniveaus en (3) een criteriumgeoriënteerde normering ten behoeve van de advisering met betrekking tot het best passende brugklatype in het voortgezet onderwijs. Alle normen werden bepaald op basis van de afnamedata van 2016. In totaal werd de module basistoets afgenomen bij 56392 leerlingen van 2085 scholen. Voor de basistoets aangevuld met de module Verdieping waren de afnamebestanden kleiner; daar moesten we volstaan met 35689 leerlingen. Voor de module Verbreding lag de omvang van de bestanden, afhankelijk van de deelttoets, tussen 8771 en 13323. Kalibraties werden uitgevoerd voor Rekenen, Lezen en Taalverzorging (zowel voor de basistoets als voor de basistoets aangevuld met Verdieping), voor Luisteren, Schrijven en Woordenschat, en voor de domeinen van Wereldoriëntatie: Aardrijkskunde, Geschiedenis en Natuur en techniek. Analyses op een random deel van de data (met N omstreeks 5400) volstonden daartoe. Bij toetsing op significantie werden

de bestanden op grond van toeval opgesplitst in tien deelbestanden van telkens omstreeks $N = 540$. Via (voorbeelden van) grafische weergaven van de S-toetsen per opgave, de verdeling van de rechteroverschrijdingskansen van S-toetsen, de model-fit van de deoltoetsen als geheel op basis van $R1c$ en de nauwkeurigheid van de itemparameterschatting op basis van de zogeheten constante 'c', konden we aantonen dat alle kalibraties geslaagd genoemd kunnen worden. Geslaagde kalibraties zijn hier op te vatten als noodzakelijke voorwaarde voor de constructvaliditeit van de deoltoetsen.

Voor een adequate relatieve normering (1) is het nodig om aan te tonen dat de data representatief zijn voor de populatie. We deden dit voor vier achtergrondkenmerken op schoolniveau (regio, mate van verstedelijking, schoolgrootte en percentage achterstandsl leerlingen) en drie op leerlingniveau (seks, leeftijd en formatiegewicht). Afwijkingen van de populatieverdeling bleken dusdanig klein, dan wel verwaarloosbaar, dat statistische weging niet noodzakelijk werd geacht. Dit maakte het mogelijk om percentielnormen af te leiden uit de data voor de basistoets (voor de totaalscore en voor de scores op Rekenen, Lezen en Taalverzorging). Voor de basistoets aangevuld met de module Verdieping was dit – door de lengte van de deoltoetsen – ook mogelijk voor de domeinscores. Voor de module Verbreding konden percentielscores worden berekend voor Luisteren, Schrijven en Woordenschat, en voor de drie domeinen van Wereldoriëntatie. De grenzen van de niveaugroepen I tot en met V zijn rechtstreeks af te leiden uit de percentielscores. Volledige normtabellen zijn opgenomen in de bijlage.

De domeingerichte normering (2) betreft de referentieniveaus 1F en 2F voor Lezen en Taalverzorging, respectievelijk 1F en 1S voor Rekenen. Ten behoeve van deze normering werd gebruikgemaakt van de openbare referentiesets die door het CvTE zijn gepubliceerd. Voldoende representatieve ankeropgaven uit deze referentiesets werden opgenomen in de proeftoetsing om de referentiecesuren te kunnen overbrengen voor de genoemde onderdelen. Dit gebeurde zowel voor de basistoets als basistoets aangevuld met de module Verdieping; cesuren in termen van de vaardigheidsscore theta waren voor beide modulen vergelijkbaar. Wel worden de onderscheiden niveaus voor Rekenen, Lezen en Taalverzorging door zeer verschillende percentages leerlingen behaald.

De criteriumgerichte normering (3) is gebaseerd op toelatings- en doorstroomonderzoek (TDO) met de vorige versie van de Entreetoets groep 7 uit 2012. Dit is legitiem voor zover de toetsversies equivalent genoemd kunnen worden. Dat beide versie equivalent zijn, wisten we aannemelijk te maken, zodat voor de acht onderscheiden brugklustypen grenswaarden konden worden bepaald aan de hand van de TDO-data.

In hoofdstuk 5 evalueerden we de betrouwbaarheid en nauwkeurigheid van de Entreetoets groep 7. Voor de scores op de basistoets werd de GLB bepaald. Voor de totaalscore is deze zeer hoog (.97), voor de score op de hoofonderdelen hoog (tussen .86 en .95). Gesimuleerde test-hertest-betrouwbaarheden liggen slechts enkele honderdsten lager. In de basistoets aangevuld met de module Verdieping zijn de GLB-waarden voor de donderdelen – vanwege de langere toetsen – hoger dan in de basistoets: tussen .93 en .97. Hier werd de GLB ook bepaald voor de domeinen van Rekenen (tussen .84 en .91), Lezen (tussen .77 en .86) en Taalverzorging (tussen .76 en .86). Voor de module Verbreding ligt GLB tussen .70 en .83 (voor de gehele toets Wereldoriëntatie is GLB .90). De laagste waarden vinden we bij relatief korte deoltoetsen met aantallen opgaven tussen 20 en 30. Daarnaast werden de overall standaardmeetfout en 90%- en 95%-betrouwbaarheidsintervallen bepaald.

Uit afbeeldingen die de lokale meetfouten grafisch weergeven blijkt, dat de meetfout kleiner is in de lagere en gemiddelde vaardigheidsregionen dan in de hogere vaardigheidsregionen. Dit wordt vooral bepaald door de gemiddelde moeilijkheidsgraad van de toetsen (in de buurt van .70) en komt overeen met de bedoelingen van de toetsconstructeurs. Het onderscheidend vermogen van de toetsen is dus het hoogst in het linkerdeel van de scoreverdelingen, terwijl tegelijkertijd de 'toetsbeleving' voor de meeste leerlingen over het algemeen prettig blijft. Uit de figuren blijkt tevens dat de lokale meetfout relatief klein is op de plaatsen waar de referentiecesuren liggen. Hierdoor wordt de kans op misclassificaties beperkt. Door de keuze van de moeilijkheidsgraad kan referentieniveau 1F het meest nauwkeurig worden bepaald, maar ook voor referentieniveau 2F (dan wel 1S) is de nauwkeurigheid nog zeer aanvaardbaar te noemen.

Om de kans op misclassificaties in beeld te brengen zijn verwarringsmatrices opgesteld. Daarnaast zijn de *Marginal Classification Accuracy* en Cohens (ongewogen) Kappa berekend als indicatoren voor de nauwkeurigheid van de classificaties. De nauwkeurigheid (bij afname van de basistoets) is het hoogst voor Rekenen (Kappa is .79), wat lager voor Lezen (.69) en nog iets lager voor Taalverzorging (.65). Een hogere

classificatie-nauwkeurigheid is te bereiken door afname van de aanzienlijk langere verdiepende module. De waarden van Kappa nemen echter slechts marginaal toe (naar achtereenvolgens .83, .77 en .72). Ten slotte is ook een verwarringsmatrix opgesteld met betrekking tot de classificatie in brugklastype op basis van de totaalscore op de basistoets. Hierbij werden simulaties ingezet. In 67% van de gevallen kwam het geobserveerde brugklastype exact overeen met het geschatte brugklastype. Wanneer het ambitie-niveau van Pillner wordt aangehouden neemt het percentage overeenstemming toe tot 97% met een (gewogen) Kappa van .86.

In hoofdstuk 6 hebben we veel aandacht geschonken aan de validiteit.

Inhoudsvaliditeit

Voor een leervorderingstoets als de Entreetoets groep 7 is de inhoudsvaliditeit buitengewoon belangrijk. Een goede inhoudsvaliditeit werd bereikt door een voortdurende studie van relevante literatuur en lesmethoden, door een uitgebreide inhoudsverantwoording te hanteren, door aan te sluiten bij het geldende Nederlandse referentiekader voor Taal en Rekenen, door de inhoud van de toetsen minutieus voor te schrijven aan de toetsconstructeurs en de uiteindelijke opgavenselectie daar zo dicht mogelijk bij te laten aansluiten, door de opgaven te laten construeren in constructiegroepen van praktijkdeskundige leerkrachten en toetsdeskundigen, door over alle opgaven empirische gegevens te verzamelen in proeftoetsen en door leerkrachten tijdens de proeftoetsing de opgaven te laten becommentariëren.

Voorwaarden voor constructvaliditeit

De toetsen worden niet alleen geconstrueerd aan de hand van inhoudelijke overwegingen, er gelden tegelijkertijd ook psychometrische overwegingen die voortvloeien uit het gehanteerd meetmodel (OPLM). Nauwkeurige kalibraties hebben gezorgd dat er bij alle deelttoetsen sprake is van uni-dimensionaliteit: alle opgaven representeren per toets dezelfde onderliggende latente trek (zie hierboven).

Kwaliteit van de toetsopgaven

De homogeniteit van de toetsopgaven in termen van gemiddelde r_{it} -waarden (ze variëren tussen .29 en .44 voor de verschillende deelttoetsen) sluit hierbij aan (evenals de eerder genoemde GLB-waarden die indicatief zijn voor de homogeniteit). De meeste r_{it} -waarden zijn als goed te kwalificeren ($> .30$), lagere waarden ($\leq .30$) komen weinig voor en onvoldoende waarden ($< .20$) zijn een uitzondering. Ook sluit de moeilijkheidsgraad van de opgaven goed aan bij het niveau van de doelgroep (gemiddelde P-waarden voor relevante deelttoetsen liggen tussen .59 en .76 en sluiten daarmee goed aan bij de gekozen uitgangspunten). P-waarden van individuele items liggen alleen in uitzonderingsgevallen niet in de nagestreefde range van .40 - .90.

De beschrijvende gegevens op het niveau van de toets en de onderdelen laten zien dat er in de regel sprake is van een licht links-scheve verdeling. Deze hangt uiteraard samen met de gekozen moeilijkheidsgraad.

Structuur: intercorrelaties

Latente intercorrelaties tussen de hoofonderdelen variëren tussen .66 en .75.

Correlaties tussen onderdelen binnen elke hoofdrubriek sluiten aan bij verwachtingen die daarover konden worden geformuleerd. Ze laten voor Rekenen zien dat er ondanks het onderscheid dat wordt gemaakt naar typen opgaven sprake is van een zeer homogene vaardigheid. De (latente) intercorrelaties zijn met waarden tussen .89 en .95 erg hoog; dit impliceert dat de domeinen op één en dezelfde vaardigheidsschaal liggen. Ook de intercorrelaties tussen de domeinen van Lezen (tussen .81 en .88) zijn in overeenstemming met de verwachtingen hoog (maar iets minder hoog dan bij Rekenen). De intercorrelaties tussen de domeinen van Taalverzorging zijn met waarden tussen .58 en .77 aanzienlijk lager. Zij weerspiegelen de opvatting dat er in groep 7 nog geen sprake is van een volledig uitgekristalliseerde en geïntegreerde vaardigheid.

De intercorrelaties tussen de scores voor Aardrijkskunde, Geschiedenis en Natuur en techniek (.86 à .87) laten zien dat het onderbrengen van de domeinen in één rubriek Wereldoriëntatie gerechtvaardigd is.

Samenhangen tussen alle lexicaal-semantiche deelvaardigheden van de taalvaardigheid onderling zijn hoog (correlaties tussen .73 en .87). Overeenkomstig de verwachtingen zijn de correlaties met semantische onderdelen systematisch hoger dan de correlaties met niet-semantische onderdelen (i.e. van Taalverzorging).

Soortgenootvaliditeit

Er is sprake van equivalentie met de vorige versie van de Entreetoets groep 7 uit 2012, dit ondanks het feit dat beide toetsen qua inhoud niet exact dezelfde onderdelen kennen. De onderlinge samenhang kunnen we op basis van de proeftoetsing inschatten als hoog en als indicatief voor de soortgenootvaliditeit, temeer daar de vorige versie van de Entreetoets groep 7 uit 2012 ook een sterke samenhang liet zien met de Cito Eindtoets Basisonderwijs.

Relevante verschillen tussen subgroepen

Analyses van verschillen tussen diverse groepen leidden tot een aantal bevindingen die aansluiten bij theoretische verwachtingen vooraf ten aanzien van deze verschillen:

Voor alle onderdelen geldt dat de groepen met een afwijkend leerlinggewicht (0.30 en 1.20) gemiddeld aanzienlijk lager scoren dan de groep reguliere leerlingen (effecten zijn middelgroot). Alle verschillen zijn significant ($p < .001$). De groepen met een afwijkend leerlinggewicht ontlopen elkaar onderling niet veel. In overeenstemming met de verwachtingen doen jongens het beter op rekenen (klein effect), meisjes beter op taal (lezen en taalverzorging, met verwaarloosbare tot kleine effecten). Alle verschillen zijn statistisch significant ($p < .001$).

Verwachtingen met betrekking tot subgroepen van leerlingen die qua leeftijd van elkaar verschillen, kwamen erop neer dat 'voorlijke' leerlingen het wat beter doen dan reguliere leerlingen en 'vertraagde' leerlingen juist slechter. Alle resultaten zijn conform de verwachtingen en significant. Effectgroottes zijn klein voor het verschil tussen 'voorlijke' en reguliere leerlingen, en middelgroot voor het verschil tussen reguliere en 'vertraagde' leerlingen.

Invariantie van structuur en DIF naar sekse en leerlinggewicht

Groepen die te onderscheiden zijn naar sekse (jongens versus meisjes) en naar leerlinggewicht (regulier; 0.0 versus afwijkend; 0.30 en 1.20) laten geen wezenlijk verschillende intercorrelaties zien tussen de scores op de hoofdrubrieken. Voor zover er verschillen zijn, zijn deze verwaarloosbaar en niet duidelijk te interpreteren.

Er zijn DIF-analyses uitgevoerd met betrekking tot sekse en groepen met een verschillend leerlinggewicht. Ten aanzien van sekse komt de meeste DIF voor bij Rekenen: bij acht à tien procent van de opgaven. Meestal is er sprake van matige DIF (bij 17 van de 20 DIF-items). De effecten voor jongens en meisjes heffen elkaar voor een deel op (10 items in het voordeel van meisjes, 10 in het voordeel van de jongens). Bij drie opgaven is sprake van forse DIF. In de taalonderdelen komt erg weinig DIF voor: bij vijf opgaven is er sprake van matige DIF, ernstige DIF kon in de taalopgaven niet worden vastgesteld. DIF naar leerlinggewicht is er nauwelijks (bij één opgave in de basistoets en bij vijf opgaven in basistoets en de module Verdieping is sprake van matige DIF (die wel steeds in het nadeel uitpakt van de leerlingen met een afwijkend leerlinggewicht. Ernstige DIF komt niet voor.

De conclusie is gerechtvaardigd dat leerlingen die van elkaar verschillen in leerlinggewicht of geslacht op de Entreetoets 7 niet of nauwelijks op oneigenlijke wijze worden benadeeld door de opgaven, dat wil zeggen door itemkenmerken die met de te meten vaardigheid zelf niets van doen hebben.

Criteriumvaliditeit

Omdat de Entreetoets groep 7 ook ondersteuning wil bieden ten aanzien van de keuze van de brugklas(sen) in het voortgezet onderwijs die het beste past of passen bij de leerlingen, zijn gegevens over de predictieve of criteriumvaliditeit van de toets relevant en belangrijk. Gegevens over de instroom van leerlingen in het voortgezet onderwijs zijn in 2012 verzameld met de vorige versie van de Entreetoets groep 7. Dat is een legitieme aanpak die ook gebruikelijk is bij eindtoetsen in het basisonderwijs, maar dan slechts voor zover er sprake is van equivalente toetsen. Die equivalentie kon worden aangetoond door gebruik te maken van equipercentiel-equivalering.

Uit het onderzoek naar de instroom blijkt dat het brugklasadvies (bepaald via de totaalscore op de basistoets) in 37% van de gevallen exact overeenkomt met het brugklatype waarin de leerling instroomt. De adviescategorieën kennen echter grenzen die dikwijls erg dicht bij elkaar liggen. In veel gevallen krijgt de leerling dan ook twee (soms zelfs drie) 'aanpalende' adviezen tegelijk. Als ook 'aanpalende' adviezen en instroom als overeenstemming worden geïnterpreteerd, stijgt de mate van overeenstemming tot 79%. Daarbij moeten we echter de ogen niet sluiten voor het feit dat er een schooljaar ligt tussen de afname van de Entreetoets groep 7 en de afname van de eindtoets in groep 8, respectievelijk de instroom in het voortgezet onderwijs. Aan de andere kant is het predictief vermogen van de Entreetoets groep 7 aanzienlijk te noemen als we een ander criterium dan feitelijke instroom hanteren. De correlatie tussen de score op de vorige versie van de Entreetoets groep 7 en de toenmalige Cito Eindtoets Basisonderwijs bedroeg in twee verschillende cohorten .87 en .88. Wanneer we er van uit mogen gaan dat er sprake is van equivalentie zeggen deze cijfers zeker ook iets over het voorspellend vermogen van de hier verantwoorde toetsversie met betrekking tot de gerealiseerde leervorderingen van een jaar later (in termen van toetsscores).

8 Literatuur

Aarnoutse, C., Verhoeven, L. (2003). *Tussendoelen gevorderde geletterdheid*. Leerlijnen voor groep 4 tot en met 8. Nijmegen: Expertisecentrum Nederlands.

Beek, A. van der, Hoogeveen, M. en Prenger, J. (2015). *Leerstoflijnen mondelinge taalvaardigheid beschreven: Uitwerking van het referentiekader Nederlandse taal voor het onderwijs in mondelinge taalvaardigheid op de basisschool*. Enschede, SLO.

Beek, A. van der & Paus, H. (2011). *Leerstoflijnen begrippenlijst en taalverzorging beschreven. Uitwerking van het referentiekader Nederlandse taal voor het domein begrippenlijst en taalverzorging op de basisschool*. Enschede: SLO.

Boxtel, H.W. van, Engelen, R., & Wijs, A. de (2011). *Wetenschappelijke verantwoording van de Eindtoets 2010*. Arnhem: Cito.

Boxtel, H.W. van, Engelen, R., Frissen, S. & Benthem, M. van (2013). *Wetenschappelijke verantwoording Entreetoets Groep 7*. Arnhem: Cito.

Boxtel, H.W. van, Engelen, R., Frissen, S. & Palm, D. van der (2015). *Wetenschappelijke verantwoording Entreetoets Groep 6*. Arnhem: Cito.

Cito (2010). *Inhoudsverantwoording Entreetoets groep 7. Basisvaardigheden Taal, Rekenen-Wiskunde en Studievaardigheden*. Arnhem: Cito.

Cito (2015). *Cito Volgstelsel primair en speciaal onderwijs. Toetsen Taalverzorging groep 6 t/m 8*. Arnhem: Cito.

Cito (2016). *Entreetoets Groep 7. Handleiding*. Arnhem: Cito.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112,155-159.

College voor Examens (2012). *Toetswijzer bij de centrale eindtoets PO taal en rekenen*. Utrecht: College voor Examens.

CvTE (2014). *Toetswijzer bij de centrale eindtoets PO wereldoriëntatie*. Inhoudsverantwoording van de centrale eindtoets voor de wereldoriënterende vakken aardrijkskunde, geschiedenis en natuur en techniek. Utrecht: CvTE.

CvTE (2014). *Openbare referentiesets taal en rekenen*. Utrecht: CvTE.

CvTE (2016). *Centrale Eindtoets 2015. Terugblik*. Utrecht: CvTE.

CvTE (2016). *Rapportage ankersets taalverzorging. Verantwoording project*. Utrecht: CvTE.

CvTE (in voorbereiding). *Centrale Eindtoets 2015. Wetenschappelijke verantwoording*. Utrecht: CvTE.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Geheel herziene versie met gewijzigde herdruk. Amsterdam: NIP/COTAN

- Engen, T.J.H.M. (1993). Itemresponstheorie en onvolledige gegevens. In: T.J.H.M. Engen & P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 239-284). Arnhem: Cito.
- Engelen, R., Roumans, P., Keizer, M. & Jongen, I. (2016). *Wetenschappelijke verantwoording Taalverzorging groep 6 tot en met 8*. Arnhem: Cito
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008a). *Over de drempels met taal. De niveaus voor de taalvaardigheid*. Enschede: SLO.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008b). *Over de drempels met taal en rekenen. Hoofdrapport van de Expertgroep Doorlopende Leerlijnen Taal en Rekenen*. Enschede: SLO.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2009a). *Referentiekader taal en rekenen. De referentieniveaus*. Enschede: SLO.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2009b). *Een nadere beschouwing. Over de drempels met taal en rekenen*. Enschede: SLO.
- Gelderens, A. van (2010). *Leerstoflijnen schrijven beschreven*. Enschede: SLO.
- Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika* 53. pp. 525–546
- Glas, C.A.W. & Verhelst, N.D., (1993). Een overzicht van itemresponsmodellen. In: T.J.H.M. Engen & P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 179-238). Arnhem: Cito.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item response Theory*. Newbury Park, CA: Sage.
- Hemker, B.T., Kordes, J. & Weerden, J.J. van (2011). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2010 – Jaarlijks Peilingsonderzoek van het Onderwijsniveau*. Arnhem: Cito.
- Holland, P. W. & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty. Research Report RR-85-43*. Princeton, NJ: Educational Testing Service.
- Huizenga, H. (2005). *Taal en didactiek. Woordenschat*. Groningen: Wolters-Noordhoff.
- Huizenga, H. (2010). *Taal en didactiek. Spelling*. Groningen: Noordhoff Uitgevers.
- Keuning, J., Boxtel, H.W. van, Lansink, L., Visser, J., Weekers, A., & Engelen, R. (2015). Actualiteit en kwaliteit van normen. Een werkwijze voor het normeren van een leerlingvolgsysteem. Arnhem: Cito.
- Krom, R., Berkel, S. van, Schoot, F. van der, Sijstra, J., Hemker, B., & Marsman, M. (2011). *Balans van het luisteronderwijs in het basis en speciaal basisonderwijs 4. Uitkomsten van de vierde taalpeiling in 2007*. PPON-reeks nummer 46. Arnhem, Cito.
- Kuhlemeier, H., Til, A. van, Hemker, B., Klijn, W. de & Feenstra, H. (2013). *Balans van de schrijfvaardigheid in het basis- en speciaal basisonderwijs 2. Uitkomsten van de peiling in 2009 in groep 5, groep 8 en de eindgroep van het SBO*. PPON-reeks nummer 53. Arnhem: Cito.
- Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Ministerie van Onderwijs, Cultuur en Wetenschappen (2004). *Voorstel herziene kerndoelen basisonderwijs*. Den Haag: Ministerie van OCW.

Ministerie van Onderwijs, Cultuur en Wetenschap (2006). *Kerndoelen primair onderwijs*. Den Haag, Ministerie van OCW.

Ministerie van Onderwijs, Cultuur en Wetenschap (2009). *Referentiekader taal en rekenen. De referentieniveaus*. Den Haag: Ministerie van OCW.

Nederlandse Taalunie (2009). *Technische handleiding. Regels voor de officiële spelling van het Nederlands*. Geraadpleegd op 19 juni 2014 via [www.taalunie.nl](#).

Pilliner, A. (1969). *Estimation of number of grades to be awarded in an examination by consideration of its reliability coefficient*. Edinburgh: The Godfrey Thomson Unit for Educational Research.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.

Sander, P., Dijk, P. van, Eggen, Th., Otter, D. den, Veldkamp, B. (2016). *RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens*. Enschede: RCEC.

Schilt-Mol, T.M.M.L. van (2007). *Differential Item Functioning en itembias in de Cito-Eindtoets Basisonderwijs*. Amsterdam: Aksant.

Schrooten, W. & Vermeer, A. (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen*. Tilburg: University Press.

Staphorsius, G. (1994). *Leesbaarheid en leesvaardigheid: de ontwikkeling van een domeingericht meetinstrument*. Enschede: Universiteit Twente.

Staphorsius, G. (red.) (2009). *Doelenboek*. Inhoudsverantwoording van de Eindtoets Basisonderwijs vanaf 2003. Arnhem: Cito.

Staphorsius, G., Krom, R.S.H., Kleintjes, F.G.M. & Verhelst, N. (2001). *Begrijpend Lezen. Verslag van het kalibratie-, validerings- en normeringsonderzoek (wetenschappelijke verantwoording)*. Arnhem: Cito.

Staphorsius, G., Verhelst, N. & Kleintjes, F.G.M. (2001). *Didactische Toetsen Schrijfvaardigheid. Verslag van het kalibratie-, validerings- en normeringsonderzoek (wetenschappelijke verantwoording)*. Arnhem: Cito.

Staphorsius, G., & Verhelst, N.D. (2001). *Achtergronden*. In: Verantwoording Toelatingstoets LWOO en PROO. Volg- en adviessysteem, Eerste Fase VO. Cito: Arnhem.

TAL-team (2007) *Metten en meetkunde in de bovenbouw*. Groningen: Wolters Noordhoff.

TAL-team (2005). *Breuken, procenten, kommagetallen en verhoudingen. Tussendoelen Annex Leerlijnen*. Groningen: Wolters Noordhoff.

TAL-team. (2004). *Jonge kinderen leren meten en meetkunde*. Groningen: Wolters Noordhoff.

TAL-team. (2001). *Kinderen leren rekenen*. Groningen: Wolters Noordhoff.

TAL-team. (1999). *Jonge kinderen leren rekenen: hele getallen*. Groningen: Wolters Noordhoff.

Tomesen, M., Weekers, A., Hiddink, L., & Jolink, A. (2017). *Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor Groep 6*. Arnhem: Cito.

Uiterwijk, H. (1994). *De bruikbaarheid van de Eindtoets Basisonderwijs voor allochtone leerlingen*. Arnhem: Cito [proefschrift KUB]

Uiterwijk, H., & Vallen, T. (1997). Onderzoek naar bias voor allochtone leerlingen in de Cito-Eindtoets Basisonderwijs. *Pedagogische Studiën*, 74, pp. 21-32.

Uiterwijk, H., & Vallen, T. (2005). Linguistic sources of item bias for second generation immigrants in Dutch tests. *Language Testing*, 22 (2), pp. 211-234.

Verhelst, N.D. (1992). *Het één parameter model (OPLM). Een theoretische inleiding en een handleiding bij het computerprogramma*. Arnhem: Cito.

Verhelst, N.D. (1993). Itemresponstheorie. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 83-178). Arnhem: Cito.

Verhelst, N.D. & Eggen T.J.M.M. (1989). *Psychometrische en statistische aspecten van Peilingsonderzoek (PPON-rapport, nr. 4)*. Arnhem: Cito.

Verhelst, N.D., & Glas, C.A.W. (1995) The one parameter logistic model. In: G.H. Fischer & I.W. Molenaar (Eds.). *Rasch models: Foundations, recent developments and applications* (pp. 215-239). New York: Springer.

Verhelst, N.D., Glas, C.A.W. & Verstralen, H.H.F.M. (1995). *OPLM: One Parameter Logistic Model. Computer program and manual*. Arnhem: Cito.

Verhelst, N.D. & Kleintjes, F.G.M. (1993). Toepassingen van itemresponstheorie. In: T.J.H.M. Eggen en P.F. Sanders (Red.). *Psychometrie in de praktijk*. Arnhem: Cito.

Verhelst, N.D., Verstralen, H.H.F.M., & Eggen, T.H.J.M. (1991). *Finding starting values for the item parameters and suitable discrimination indices in the one-parameter logistic model*. Measurement and Research Department Reports 91-10. Arnhem: Cito.

Verhoeven, L., Biemond, H., & Litjens, P. (2007). *Tussendoelen mondelinge communicatie. Leerlijnen voor groep 1 tot en met 8*. Nijmegen: Expertisecentrum Nederlands.

Bijlagen

Bijlage 1 Rechteroverschrijdingskansen van S-toetsen en R1c per toetsonderdeel

Basistoets

Basis – Rekenen (90 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	249.401	267	.7742	1	5	3	13	8	10	12	11	11	6	8	2
2	251.677	267	.7426	1	7	9	4	11	10	10	5	7	3	11	12
3	260.995	267	.5947	2	4	4	8	7	9	6	10	10	14	10	6
4	260.368	267	.6053	0	4	5	12	6	10	5	11	7	12	9	9
5	266.214	267	.5050	1	2	5	9	11	11	10	8	12	6	8	7
6	287.123	267	.1903	2	4	2	12	5	17	8	8	6	9	9	8
7	226.415	267	.9646	2	1	4	5	9	6	6	13	14	13	8	9
8	283.106	267	.2393	1	10	6	8	7	11	8	8	9	3	8	11
9	274.882	267	.3593	2	4	4	11	7	9	11	6	10	11	5	10
10	300.776	267	.0750	0	7	7	12	8	3	9	8	13	5	8	10

Basis – Lezen (40 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	95.608	117	.9245	0	0	2	1	3	5	9	3	2	7	3	5
2	114.134	117	.5619	1	2	3	5	0	6	4	5	4	5	3	2
3	103.932	117	.8015	1	3	0	4	3	6	3	4	3	4	3	6
4	110.048	117	.6659	0	2	3	3	5	3	7	2	4	5	4	2
5	109.698	117	.6745	0	3	2	4	2	5	4	3	6	4	4	3
6	98.068	117	.8961	0	1	1	6	3	6	4	2	5	4	4	4
7	126.390	117	.2628	1	0	3	2	5	7	1	3	2	4	7	5
8	96.827	117	.9112	0	0	2	6	2	3	4	6	4	2	7	4
9	75.516	117	.9985	0	0	3	0	3	6	5	6	11	2	1	3
10	97.417	117	.9042	0	2	4	2	4	1	6	1	7	4	3	6

Basis – Taalverzorging (50 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	144.375	147	.5496	1	0	1	2	8	4	3	4	6	9	4	8
2	152.733	147	.3593	0	2	3	4	6	3	9	2	4	6	2	9
3	148.617	147	.4510	0	4	2	8	4	4	5	6	3	4	7	3
4	114.179	147	.9776	1	0	1	7	5	2	7	5	7	5	3	7
5	145.053	147	.5338	1	2	4	3	6	2	6	7	3	8	5	3
6	155.885	147	.2947	0	1	2	6	7	5	7	3	8	3	5	3
7	120.005	147	.9479	1	2	4	3	6	7	2	3	5	6	8	3
8	142.190	147	.6001	0	3	5	7	3	8	0	4	4	5	4	7
9	152.086	147	.3732	1	3	1	4	4	6	8	7	4	3	4	5
10	141.983	147	.6049	0	5	0	5	5	4	3	7	1	9	4	7

Module Verdieping

Verdieping – Rekenen (140 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	770.595	773	.5193	3	7	3	14	11	17	11	14	20	13	12	15
2	742.391	773	.7804	1	3	7	10	13	15	13	16	18	13	12	19
3	775.020	773	.4745	2	8	4	18	18	14	15	13	14	15	10	9
4	781.651	773	.4082	0	7	6	11	15	6	17	14	21	15	16	12
5	782.946	773	.3956	0	5	7	13	21	9	17	15	17	11	14	11
6	835.977	773	.0567	2	7	7	17	11	9	18	18	6	12	20	13
7	801.652	773	.2312	1	8	5	15	9	5	13	18	12	15	19	20
8	783.845	773	.3868	0	9	7	10	16	10	11	11	17	20	12	17
9	815.716	773	.1391	1	7	6	12	14	8	21	16	14	11	15	15
10	772.899	773	.4960	2	1	6	15	13	8	14	15	22	21	9	14

Verdieping – Lezen (80 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	323.751	393	.9949	2	4	2	9	4	6	10	8	5	7	11	12
2	357.536	393	.8992	0	2	2	8	9	8	12	7	6	13	7	6
3	378.076	393	.6983	0	1	4	8	9	5	9	12	8	8	5	11
4	344.931	393	.9602	0	2	8	8	4	9	6	10	11	8	8	6
5	375.497	393	.7303	2	1	3	6	8	8	6	12	9	10	6	9
6	358.475	393	.8929	2	1	1	9	9	6	8	7	5	10	8	14
7	315.105	393	.9982	0	0	2	12	7	4	5	9	8	11	12	10
8	341.594	393	.9699	0	2	5	9	7	6	6	9	5	9	15	7
9	363.333	393	.8557	0	7	2	8	5	4	4	12	11	7	13	7
10	332.660	393	.9869	0	3	7	7	6	11	7	3	9	11	10	6

Verdieping – Taalverzorging (90 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	409.093	463	.9648	1	4	5	9	11	10	6	8	10	9	10	7
2	458.791	463	.5486	1	7	6	7	5	9	12	9	7	7	13	7
3	459.255	463	.5426	2	2	4	10	9	5	9	12	13	2	11	11
4	443.334	463	.7379	1	3	7	6	7	10	7	9	6	14	9	11
5	492.068	463	.1693	0	2	2	6	12	5	8	10	10	13	11	11
6	441.611	463	.7564	1	2	4	6	16	4	7	9	11	6	9	15
7	435.344	463	.8176	1	3	4	4	11	6	13	11	6	8	14	9
8	367.498	463	.9995	1	1	4	7	7	8	10	14	12	6	9	11
9	372.644	463	.9991	1	2	2	6	4	11	6	14	18	12	7	7
10	418.090	463	.9328	1	2	3	10	10	8	4	13	11	10	3	15

Module Verbreding

Taalonderdelen

Verbreding – Taal – Luisteren (30 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	91.804	87	.3416	0	1	2	6	2	2	5	2	2	5	0	3
2	58.074	87	.9928	1	0	1	4	3	0	3	3	2	5	6	2
3	73.484	87	.8489	0	2	0	5	1	4	5	4	0	3	3	3
4	77.907	87	.7466	1	0	1	2	4	1	3	0	4	4	5	5
5	87.028	87	.4790	1	1	0	3	6	3	2	1	2	2	3	6
6	68.952	87	.9229	0	0	0	6	4	4	2	2	4	3	3	2
7	77.339	87	.7613	0	3	0	1	2	2	2	5	4	1	4	6
8	91.457	87	.3510	1	1	1	5	2	4	0	3	2	1	6	4
9	90.966	87	.3644	0	3	4	2	2	1	3	5	2	3	4	1
10	92.467	87	.3241	0	2	0	2	2	7	2	2	2	4	5	2

Verbreding – Taal – Schrijven (30 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	79.502	87	.7035	0	2	1	2	2	5	2	5	1	3	5	2
2	89.831	87	.3963	0	0	2	6	3	5	3	2	4	2	1	2
3	88.809	87	.4259	1	0	2	3	2	3	7	1	4	4	1	2
4	92.648	87	.3193	1	1	3	1	2	2	3	5	2	6	2	2
5	86.314	87	.5006	0	1	1	4	2	2	1	4	4	6	3	2
6	81.271	87	.6530	1	2	4	3	2	3	0	1	4	1	5	4
7	84.175	87	.5658	1	1	2	5	2	1	1	2	3	5	2	5
8	88.920	87	.4227	0	3	1	2	2	4	2	3	1	6	3	3
9	77.820	87	.7489	0	0	2	1	5	0	4	2	3	6	3	4
10	97.642	87	.2044	0	4	2	5	2	0	2	4	6	3	2	0

Verbreding – Taal – Woordenschat (40 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	115.687	117	.5213	0	0	1	7	6	5	3	5	1	3	6	3
2	152.986	117	.0129	0	3	1	5	4	4	7	3	4	3	3	3
3	102.981	117	.8194	1	0	2	3	3	3	6	5	7	6	1	3
4	119.391	117	.4253	1	2	4	1	3	4	5	4	0	8	5	3
5	126.575	117	.2590	1	1	2	4	3	1	2	4	7	5	5	5
6	126.590	117	.2587	0	0	0	7	3	3	6	1	6	6	5	3
7	100.178	117	.8664	0	1	1	4	2	6	6	1	3	6	3	7
8	75.074	117	.9987	0	2	1	0	2	2	1	10	2	5	7	8
9	116.065	117	.5114	0	2	4	7	1	2	2	3	1	5	7	6
10	110.148	117	.6635	0	2	2	6	2	7	5	0	7	3	4	2

Module Verbreding

Onderdelen Wereldoriëntatie

Verbreding – Wereldoriëntatie – Aardrijkskunde (20 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	48.171	57	.7912	0	3	1	1	3	0	1	2	2	3	2	2
2	49.590	57	.7465	0	0	0	0	3	4	2	3	2	1	2	3
3	68.057	57	.1499	0	1	0	3	0	1	3	2	3	0	6	1
4	44.732	57	.8810	0	0	0	2	2	2	2	2	4	1	1	4
5	48.481	57	.7818	0	2	0	1	1	4	1	2	4	2	2	1
6	50.815	57	.7049	0	0	2	2	1	3	2	4	0	1	2	3
7	62.132	57	.2984	1	1	1	2	3	2	3	2	1	0	1	3
8	30.853	57	.9982	0	0	0	2	0	1	1	2	4	5	4	1
9	40.661	57	.9498	0	1	0	1	4	3	0	1	4	2	2	2
10	61.259	57	.3258	1	1	3	1	1	2	2	3	1	0	4	1

Verbreding – Wereldoriëntatie – Geschiedenis (20 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	64.356	57	.2348	1	3	1	4	3	3	1	0	2	1	1	0
2	41.603	57	.9374	0	1	1	0	3	1	3	3	0	3	3	2
3	70.833	57	.1030	0	1	1	2	2	3	1	4	2	1	2	1
4	54.438	57	.5718	0	2	0	2	2	4	0	2	0	0	7	1
5	39.670	57	.9609	0	3	0	1	1	2	0	2	2	3	3	3
6	60.813	57	.3403	0	0	0	4	3	2	0	0	3	4	2	2
7	100.629	57	.0003	1	3	2	1	2	4	1	1	1	0	2	2
8	50.350	57	.7209	0	0	3	1	1	3	1	3	0	1	3	4
9	42.562	57	.9227	0	0	0	0	4	2	3	2	2	3	1	3
10	66.586	57	.1805	0	3	3	1	2	1	2	2	1	2	3	0

Verbreding – Wereldoriëntatie – Natuur en techniek (20 opgaven)

#	R1c	df	p	0	1	2	3	4	5	6	7	8	9	10	
				[0,1] interval van p-waarden											
1	45.517	57	.8630	0	0	1	3	1	0	0	5	4	4	0	2
2	55.460	57	.5330	0	1	1	1	3	1	2	1	4	1	3	2
3	54.357	57	.5748	0	2	2	3	2	0	0	3	1	1	1	5
4	46.221	57	.8456	0	0	0	1	2	2	2	5	4	2	2	0
5	49.166	57	.7602	1	3	0	1	2	0	1	1	3	2	2	4
6	49.645	57	.7447	0	0	1	2	3	1	1	3	2	3	2	2
7	58.536	57	.4188	2	1	1	1	0	1	2	1	3	2	4	2
8	61.680	57	.3124	0	0	0	4	3	3	2	4	1	2	0	1
9	69.178	57	.1293	0	1	1	1	3	1	2	2	1	2	1	5
10	47.208	57	.8192	0	0	1	1	2	1	4	3	3	2	1	2

Bijlage 2 Normtabellen (omzetting van ruwe naar percentielscores)

Omzettingstabel van ruwe scores naar percentielscores voor de totaalscore (TOT), Rekenen (REK), Lezen (LEZ) en Taalverzorging (TAV) voor basistoets

Perc	Ruwe scores				Perc	Ruwe scores			
	TOT	REK	LEZ	TAV		TOT	REK	LEZ	TAV
01	0-57	0-22	0-11	0-17	51	126	64	29	-
02	58-63	23-25	12-13	18-19	52	127	-	-	36
03	64-67	26-27	14	20	53	-	65	-	-
04	68-71	28-29	15	21	54	128	-	-	-
05	72-74	30-31	16	22	55	129	-	-	-
06	75-77	32-33	-	23	56	-	66	-	-
07	78-79	34	17	24	57	130	-	30	-
08	80-81	35-36	18	-	58	131	67	-	37
09	82-83	37	-	25	59	-	-	-	-
10	84-85	38	19	-	60	132	68	-	-
11	86-87	39	-	26	61	133	-	-	-
12	88-89	40	-	-	62	-	-	-	-
13	90	41	20	27	63	134	69	-	-
14	91-92	42	-	-	64	135	-	31	-
15	93	43	21	-	65	-	70	-	38
16	94	44	-	28	66	136	-	-	-
17	95-96	45	-	-	67	137	-	-	-
18	97	-	22	-	68	-	71	-	-
19	98	46	-	29	69	138	-	-	-
20	99	47	-	-	70	139	-	-	-
21	100	48	-	-	71	-	72	32	39
22	101-102	-	23	-	72	140	-	-	-
23	103	49	-	30	73	141	73	-	-
24	104	50	-	-	74	-	-	-	-
25	105	-	24	-	75	142	-	-	-
26	106	51	-	-	76	143	74	-	-
27	107	52	-	31	77	144	-	33	-
28	108	-	-	-	78	-	75	-	40
29	109	53	-	-	79	145	-	-	-
30	-	-	25	-	80	146	-	-	-
31	110	54	-	32	81	-	76	-	-
32	111	55	-	-	82	147	-	-	-
33	112	-	-	-	83	148	77	-	41
34	113	56	26	-	84	149	-	34	-
35	114	-	-	-	85	-	-	-	-
36	115	57	-	33	86	150	78	-	-
37	-	-	-	-	87	151	-	-	-
38	116	58	-	-	88	152	79	-	-
39	117	-	27	-	89	153	-	35	42
40	118	59	-	-	90	154	80	-	-
41	119	-	-	34	91	155	-	-	-
42	120	60	-	-	92	156	81	-	-
43	-	-	-	-	93	157	-	-	43
44	121	61	-	-	94	158	82	36	-
45	122	-	28	-	95	159	-	-	-
46	123	62	-	35	96	160-161	83	-	44
47	-	-	-	-	97	162-163	84	-	-
48	124	-	-	-	98	164-165	85	37	45
49	125	63	-	-	99	166-179	86-89	38-39	46-49
50	-	-	-	-	100	180	90	40	50

Omzettingstabel van ruwe scores naar percentielscores voor Rekenen (REK), Lezen (LEZ) en Taalverzorging (TAV) voor module (Basis en) Verdieping

Perc	Ruwe scores				Perc	Ruwe scores			
	TOT ¹⁾	REK	LEZ	TAV		TOT ¹⁾	REK	LEZ	TAV
01		0-35	0-24	0-32	51		-	-	65
02		36-40	25-27	33-35	52		101	58	-
03		41-44	28-29	36-38	53		-	-	-
04		45-47	30-31	39-40	54		102	-	66
05		48-50	32	41	55		103	59	-
06		51-53	33	42-43	56		-	-	-
07		54-55	34-35	44	57		104	-	-
08		56-57	36	45	58		105	-	67
09		58-59	37	46	59		-	60	-
10		60-61	38	47	60		106	-	-
11		62	39	48	61		-	-	-
12		63-64	-	-	62		107	61	68
13		65	40	49	63		108	-	-
14		66-67	41	50	64		-	-	-
15		68	42	-	65		109	62	69
16		69-70	-	51	66		-	-	-
17		71	43	52	67		110	-	-
18		72	-	-	68		111	-	-
19		73	44	53	69		-	63	70
20		74	45	-	70		112	-	-
21		75-76	-	54	71		-	-	-
22		77	46	-	72		113	64	-
23		78	-	55	73		114	-	71
24		79	47	-	74		-	-	-
25		80	-	56	75		115	65	-
26		81	48	-	76		-	-	72
27		82	-	57	77		116	-	-
28		83	49	-	78		-	-	-
29		84	-	-	79		117	66	-
30		-	50	58	80		118	-	73
31		85	-	-	81		-	-	-
32		86	-	59	82		119	67	-
33		87	51	-	83		-	-	74
34		88	-	-	84		120	-	-
35		89	52	60	85		121	68	-
36		-	-	-	86		-	-	75
37		90	-	-	87		122	-	-
38		91	53	61	88		123	69	-
39		92	-	-	89		-	-	76
40		93	54	-	90		124	70	-
41		-	-	62	91		125	-	-
42		94	-	-	92		126	-	77
43		95	55	-	93		-	71	-
44		-	-	63	94		127	-	78
45		96	-	-	95		128	72	-
46		97	56	-	96		129	-	79
47		-	-	64	97		130	73	80
48		98	-	-	98		131-132	74	81
49		99	57	-	99		133-139	75-79	82-89
50		100	-	-	100		140	80	90

¹⁾ Niet van toepassing: Ook als de module Verdieping is afgenomen, wordt in de kopregel van het Leerlingprofiel altijd gerapporteerd over de totaalscore op grond van de module Basis.

Omzettingstabel van ruwe scores naar percentielscores voor de onderdelen van Rekenen, Lezen en Taalverzorging (module (Basis en) Verdieping

Ruwe score	Percentielscores										
	Rekenen				Lezen			Taalverzorging			
	Getallen	Meten en meetkunde	Verbanden	Verhoudingen	Begrijpend lezen	Opzoeken	Samenvatten	Grammatica	Interpunctie	Spelling niet-werkwoorden	Spelling werkwoorden
00	1	1	1	1	1	1	1	1	1	0	0
01	1	1	1	1	1	1	1	1	1	1	1
02	1	1	1	1	1	1	1	1	1	1	1
03	1	1	1	1	1	1	1	1	1	1	1
04	1	1	1	1	1	1	1	1	1	1	1
05	1	1	1	1	1	1	1	2	1	1	1
06	1	1	1	2	1	2	2	3	1	1	1
07	1	1	2	3	1	3	3	5	1	1	2
08	1	1	3	4	1	5	6	7	3	1	4
09	1	1	5	6	1	9	9	9	4	1	7
10	1	2	7	8	1	13	14	12	7	2	12
11	1	3	10	10	1	19	19	16	11	3	18
12	1	4	13	12	1	26	27	20	15	4	26
13	1	6	17	14	2	35	36	26	21	6	35
14	2	8	22	17	3	45	46	32	29	9	45
15	2	10	27	20	4	56	58	41	38	13	55
16	3	12	34	24	5	69	71	51	49	18	65
17	4	15	41	28	6	81	83	64	62	24	75
18	4	19	48	32	8	91	93	78	78	33	83
19	5	23	57	37	10	98	99	93	93	43	90
20	6	27	66	42	12	100	100	100	100	56	95
21	8	32	75	47	15					70	98
22	9	37	84	53	18					84	99
23	11	43	92	60	21					95	99
24	12	49	98	67	25					99	99
25	14	56	100	73	29					100	100
26	16	63		81	34						
27	18	70		87	39						
28	21	77		94	44						
29	24	83		98	50						
30	26	89		100	57						
31	29	94			64						
32	33	97			70						
33	36	99			77						
34	40	99			84						
35	44	100			90						
36	48				95						
37	52				98						
38	57				99						
39	61				99						
40	66				100						
41	71										
42	76										
43	81										
44	85										
45	90										
46	94										
47	97										
48	99										
49	99										
50	100										

Omzettingstabel van ruwe scores naar percentielscores voor de taalonderdelen en (onderdelen van) Wereldoriëntatie van de module Verbreding

Ruwe score	Percentielscores						
	Taalonderdelen			Wereldoriëntatie			
	Luisteren	Schrijven	Woordenschat	Wereldoriëntatie	Aardrijkskunde	Geschiedenis	werkwoorden
00	1	1	1	1	1	1	1
01	1	1	1	1	1	1	1
02	1	1	1	1	1	1	1
03	1	1	1	1	1	1	1
04	1	1	1	1	1	1	1
05	1	1	1	1	1	1	1
06	1	1	1	1	3	2	3
07	1	1	1	1	5	5	5
08	1	1	1	1	8	9	8
09	1	2	1	1	12	16	13
10	1	3	1	1	18	24	19
11	2	4	1	1	25	34	26
12	3	6	1	1	34	46	36
13	4	7	1	1	44	59	46
14	6	10	2	1	56	72	58
15	8	13	2	1	68	84	71
16	11	16	3	1	80	93	83
17	15	20	5	2	90	98	92
18	19	25	6	2	97	99	98
19	25	30	8	3	99	99	99
20	32	37	11	4	100	100	100
21	39	44	14	5			
22	48	52	18	6			
23	58	61	23	7			
24	68	70	28	8			
25	79	79	34	10			
26	88	88	41	11			
27	95	94	48	13			
28	99	99	56	15			
29	99	99	64	18			
30	100	100	72	20			
31			79	23			
32			86	26			
33			91	29			
34			95	32			
35			98	35			
36			99	39			
37			99	43			
38			99	47			
39			99	51			
40			100	55			

Voor Wereldoriëntatie : zie vervolgtabel

Omzettingstabel van ruwe scores naar percentielscores voor de taalonderdelen en (onderdelen van) Wereldoriëntatie van de module Verbreding [vervolg]

Ruwe score	Percentielscores						
	Taalonderdelen			Wereldoriëntatie			
	Luisteren	Schrijven	Woordenschat	Wereldoriëntatie	Aardrijkskunde	Geschiedenis	werkwoorden
41				59			
42				64			
43				68			
44				72			
45				77			
46				81			
47				84			
48				88			
49				91			
50				94			
51				96			
52				97			
53				99			
54				99			
55				99			
56				99			
57				99			
58				99			
59				99			
60				100			

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

Cito

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

Fotografie: Ron Steemers