

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de wetenschappelijke verantwoording en de Handleiding veelal letterlijk vermeld. De wetenschappelijke verantwoording heeft betrekking op de uitgangspunten van de toetsconstructie, de normen, de betrouwbaarheid en meetnauwkeurigheid en de validiteit. De Handleiding heeft betrekking op het gebruik van de toets, communicatie over de toetsgegevens en de inhoudsverantwoording.

Algemeen

Het Cito Volgstelsel primair en speciaal onderwijs beoogt de vorderingen van individuele leerlingen, groepen leerlingen en het onderwijs op school van groep 1 tot en met groep 8 te volgen en te evalueren. De toets Begrijpend lezen 3.0 voor groep 3 is een onderdeel van het Cito Volgstelsel primair en speciaal onderwijs en is bedoeld voor leerlingen in groep 3 van het primair onderwijs. De toetsen voor groep 3 tot en met groep 8 worden vanaf het schooljaar 2014/2015 geleidelijk vervangen door de nieuwe toetsen Begrijpend lezen 3.0. De toets voor groep 3 vormt dan samen met de LVS toetsen Begrijpend lezen 3.0 groep 4 tot en met 8 een systeem om vast te stellen hoe goed leerlingen begrijpend kunnen lezen en hoe hun vaardigheid op dit gebied zich in de basisschoolperiode ontwikkelt.

Onderstaande beschrijving is gebaseerd op de Handleiding.

Meetpretentie

De toets in de toets pakketten Begrijpend lezen 3.0 voor groep 3 van het Cito Volgstelsel primair en speciaal onderwijs is bedoeld om vast te stellen hoe goed een leerling geschreven teksten, met name zakelijke teksten en fictieve, narratieve en literaire teksten, kan begrijpen en hoe de vaardigheid in begrijpend lezen van de leerling zich in de loop van de jaren ontwikkelt.

Een belangrijk verschil met de vorige toetsen (LVS 2.0) is dat de nieuwe toetsen 3.0 voorspelopgaven bevatten en geen husselopgaven. Ook zijn er geen verschillende vervolgmodes meer, maar maken alle leerlingen die (bijvoorbeeld) de E3-toets maken dezelfde toets.

Doelgroep

De toets Begrijpend lezen 3.0 groep 3 is bedoeld voor leerlingen in groep 3 van het primair en speciaal onderwijs, maar kan ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 3 en voor leerlingen met een ontwikkelingsachterstand en/of extra onderwijsbehoeften. Voor deze groepen speciale leerlingen zijn geen afzonderlijke normen vastgesteld en de toetsresultaten van deze leerlingen worden geïnterpreteerd met behulp van de gemiddelde vaardigheidsscores voor leerlingen uit het regulier onderwijs.

Voor leerlingen bij wie het lezen nog niet geautomatiseerd is en/of die nog maar pas in Nederland verblijven, zijn de toetsen ongeschikt. Leerlingen moeten het Nederlands voldoende beheersen om de opgaven te kunnen maken voordat de toetsen Begrijpend lezen bij hen kunnen worden afgenomen.

Gebruiksdoel en functie

Doel van de toets Begrijpend lezen 3.0 voor groep 3 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van begrijpend

lezen. Hiervoor wordt de behaalde vaardigheidsscore normgericht geïnterpreteerd op basis van de vaardigheidsverdeling in een adequate, landelijke, referentiegroep. De vaardigheidsscore wordt uitgedrukt in de symmetrische niveau indeling I t/m V en in de asymmetrische niveau indeling A t/m E. De toets maakt het mogelijk om:

- De vaardigheid begrijpend lezen van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) te beoordelen via een vergelijking van de behaalde scores met de scores van een landelijke referentiegroep oftewel niveaubepaling.
- De ontwikkeling van de vaardigheid begrijpend lezen van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) door de leerjaren heen te volgen oftewel progressiebepaling.

Inhoudelijke theoretische inkadering:

De inhoud van de toets Begrijpend lezen 3.0 is gebaseerd op het domein Lezen, waarbij onderscheid is gemaakt tussen het lezen van zakelijke teksten en het lezen van fictionele, narratieve en literaire teksten, beschreven in het Referentiekader Taal en Rekenen. De toets sluit aan bij de indeling die is gehanteerd in het Referentiekader Taal. In de publicatie 'Leerstoflijnen lezen beschreven' van de SLO, is aangegeven hoe de opbouw van de leerstoflijnen eruit kan zien voor de verschillende groepen. Voor de inhoud van de toetsen zijn deze uitwerkingen bepalend geweest voor zowel de theoretische basis als voor de indeling van de vaardigheden. De indeling van de vaardigheden is mede gebaseerd op basis van analyse van methoden (~~taal en lezen~~) voor begrijpend lezen die veel gebruikt worden in het basisonderwijs.

De koppeling die gemaakt wordt met het Referentiekader levert voor scholen een meerwaarde. Zeker wanneer de toetsen voor de bovenbouw gereed zijn en een directe koppeling met de referentieniveaus gemaakt kan worden, zodat inzicht wordt verkregen over wat de niveaus 1F en 1S voor Lezen inhouden en hoe leerlingen zich met hun toetsresultaten verhouden tot de referentieniveaus.

Inhoud van het toetspakket

Het toetspakket Begrijpend lezen 3.0 groep 3 bestaat uit de volgende documenten:

- Handleiding, deze bevat informatie over:
 - de afname van de toets (hfdst. 2),
 - nakijken en verwerken van toetsgegevens (hfdst. 3),
 - interpretatie van de toetsresultaten op leerling- en groepsniveaus (hfdst 4),
 - interpretatie van toetsresultaten op schoolniveau (hfdst 5),
 - theoretisch kader en achtergronden van de toets (hfdst 6),
 - communiceren over toetsresultaten met leerling en ouders (hfdst 7),
 - achtergrondinformatie en veelgestelde vragen (hfdst 8) en
 - enkele bijlagen
- Één toets:
 - Toets E3 (Eind groep 3)
- Afnamekaarten met aanwijzingen voor de papieren ~~of de digitale~~ afname van de toets
- Nakijkaarten
- Antwoordbladen
- Tabellen voor de ~~vier toetsen~~ voor het bepalen van de vaardigheidsscore en -niveau.

2. Beoordeling van de kwaliteitsaspecten

De beoordeling vindt plaats volgens het 'Beoordelingskader voor de psychometrische aspecten van (reeksen van) toetsen uit leerlingvolgsystemen (LOVS)', zoals opgesteld door de Expertgroep Toetsen PO. De Expertgroep Toetsen PO wordt gevormd door Prof. Dr. Cees Van der Vleuten (voorzitter), Prof. dr. Cees Glas (psychometrisch expert), Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en mevrouw Paulyn K. Berding-Oldersma MSc (secretaris).

De kwaliteit van de dataverzameling

S1.1. Is de steekproef representatief?

Bevindingen:

In juni 2012 zijn in een kalibratieonderzoek (proefonderzoek) 268 items voorgelegd aan 2.106 leerlingen van groep 3. De 268 items waren verdeeld over 15 boekjes (booklets). Elke boekje bestond uit 48 tot 51 opgaven verdeeld over 2 taken. Een van de twee taken betrof de starttaak uit LVS-II. Deze taak werd door alle 2.106 leerlingen gemaakt. De overige opgaven kwamen in een of twee boekjes voor en werden gemiddeld door 213 leerlingen gemaakt wat boven het minimum vereiste van 150 ligt.

Op grond van het kalibratieonderzoek (juni 2012) is voor het normeringsonderzoek een selectie gemaakt van 172 items, verdeeld over 6 boekjes, met elk 71 tot 76 opgaven verdeeld over 3 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. De steekproef is tegelijkertijd voor zowel normeringsonderzoek Begrijpend lezen E3 als voor normeringsonderzoek Spelling E3 benaderd. Het embedded field normeringsonderzoek is toegepast op de resultaten van 1.860 leerlingen uit groep 3 van 83 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.927 leerlingen uit groep 3 van 78 scholen uit Cito dataretour.

De representativiteit van de steekproeven voor de normeringsonderzoeken is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. Bij regio is uitgegaan van de vier landsdelen / regio's van de CBS-indeling. Bij urbanisatiegraad is uitgegaan van de CBS-indeling naar vijf niveaus van verstedelijking. Bij schooltype is uitgegaan van de formatiegewichten volgens OCW. Hierin worden drie niveaus onderscheiden die gebaseerd zijn op het opleidingsniveau van de ouders. Bij sekse is een tweedeling gemaakt naar jongens en meisjes. De steekproefverdeling wijkt weinig af van de populatieverdeling. De effectgrootte phi is het grootst voor de variabele schooltype. Daar onderzoek heeft laten zien dat de verdeling naar etnische herkomst sterk samenhangt met de verdeling naar urbanisatiegraad en schooltype, wordt aangenomen dat de steekproef tevens voldoende representatief is voor de variabele etniciteit. Schoolgrootte, ingedeeld in acht strata, is gebruikt bij de werving van de scholen voor het onderzoek. Schoolgrootte wordt niet beschouwd als een – voor de representativiteit – relevante achtergrondvariabele. Het wordt gebruikt om een onevenwichtige verdeling op andere achtergrondkenmerken zoals regio en urbanisatiegraad te voorkomen.

Uit de ruwe scores van de individuele leerlingen uit het embedded field normeringsonderzoek en Cito dataretour werden plausible values gegenereerd op de nieuw ontwikkelde vaardigheidsschaal. De normering werd vervolgens gebaseerd op de plausible

values van de leerlingen in de normeringssteekproef. De plausible values voor het afnamemoment E3 bleken een normale verdeling te vormen. De schoolverdeling werd bepaald met het intercept-only multilevel model. Dit model werd geschat via een bootstrap procedure. Ondanks dat de percentielen van de normgegevens op schoolniveau dichter bij elkaar kwamen te liggen dan in de leerlingverdeling, waren de afstanden groot genoeg om scholen zinvol te classificeren in de verschillende niveaus.

Conclusie:

De steekproef is representatief, is adequaat gestratificeerd naar sekse, regio en urbanisatiegraad en geeft informatie over hoe de steekproef zich verhoudt tot de populatiewaarden. De procedure voor het samenstellen van de steekproef is onderbouwd en de omstandigheden waaronder data is verzameld, is redelijk vergelijkbaar met de omstandigheden waaronder de toets wordt afgenomen. Daarmee wordt aan aspect S1.1. het oordeel '**voldoende**' toegekend.

S1.2. In geval van een onvolledig dataverzamelingsdesign: is het design adequaat?

Bevindingen:

Om te komen tot een set van psychometrisch en inhoudelijk geschikte items zijn de opgaven uit het proefonderzoek van juni 2012 en de opgaven uit het daaropvolgende normeringsonderzoek gekalibreerd. Hiervoor is gebruik gemaakt van het IRT model OPLM. Met dit statistische model zijn de psychometrische kenmerken (moeilijkheidsparameters en discriminatie indices) van de items geschat.

In het kalibratieproces is uitgegaan van een onvolledig maar 'verbonden' design. In het kalibratieproces van juni 2012 zijn 268 items voorgelegd aan 2.106 leerlingen van groep 3. De 268 items waren verdeeld over 15 boekjes (booklets). Elke boekje bestond uit 48 tot 51 opgaven verdeeld over 2 taken. Een van de twee taken betrof de starttaak uit LVS 2.0. Deze taak werd door alle 2.106 leerlingen gemaakt. De overige opgaven kwamen in een of twee boekjes voor en werden gemiddeld door 213 leerlingen gemaakt wat boven het minimum vereiste van 150 ligt. Door de koppeling met LVS 2.0 kunnen de scores van LVS 2.0 en LVS 3.0 worden vergeleken.

Op basis van inhoudelijke en psychometrische criteria werden 40 items voor de toets Begrijpend lezen 3.0 groep 3 geselecteerd. De 40 items zijn verdeeld naar vaardigheid en inhoudsaspecten van de elementen begrijpen en interpreteren van zowel formele als informele tekstgenres als de componenten van de latente vaardigheid Begrijpend lezen. Van alle opgaven die zijn meegegaan in het normeringsonderzoek zijn de gekalibreerde p-waarde, de τ_{it} waarde bepaald. Voor het normeringsonderzoek E3 werden na het trekken van een representatieve steekproef, waarbij rekening werd gehouden met verdeling naar regio, urbanisatiegraad, schooltype en sekse, zowel scholen geworven als data gehaald uit Cito dataretour. Voor het normeringsonderzoek werd gebruik gemaakt van resultaten van 1.860 leerlingen uit groep 3 van 83 scholen en van dataretour van 1.927 leerlingen uit groep 3 van 78 scholen.

Uit het kalibratieonderzoek blijkt dat de items passen bij voornoemd IRT model en dat het model ook past voor de toets al geheel. Dit betekent dat er sprake is van één

unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Conclusie:

Het onvolledige maar 'verbonden' design van het proefonderzoek is adequaat. Het volledige design van toets E3 is eveneens adequaat. Aan aspect S1.2 wordt het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

De toets is genormeerd voor het afnamemoment E3.

Op grond van het kalibratieonderzoek (juni 2012) is een selectie gemaakt van 172 items, verdeeld over 6 boekjes, met elk 71 tot 76 opgaven verdeeld over 3 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. De aselecte steekproef is tegelijkertijd voor zowel normeringsonderzoek Begrijpend lezen E3 als voor normeringsonderzoek Spelling E3 benaderd. Het embedded field normeringsonderzoek is toegepast op de resultaten van 1.860 leerlingen uit groep 3 van 83 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.927 leerlingen uit groep 3 van 78 scholen uit Cito dataretour.

Voor afnamemoment E3 werden vaardigheidsverdelingen gepresenteerd op leerlingniveau en op schoolniveau. Dit betreft de gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75 en P80. Van hieruit kunnen de beide niveau indelingen (de symmetrische niveau indeling I t/m V en de asymmetrische niveau indeling A t/m E) worden bepaald. Een punt van kritiek is wel dat niet wordt vermeld tot wanneer de normen geldig zijn.

De normen voor de toets Begrijpend lezen groep 3 zijn geldig tot en met 2023.

Conclusie:

Er is sprake van relatieve normen, de steekproef is representatief en groot genoeg. Wat we nog missen in de wetenschappelijke verantwoording is tot wanneer deze normen geldig zijn. Daarmee wordt aan aspect N1.2.1. het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit van de steekproef is besproken bij punt S1.1. Hier wordt reeds geconstateerd dat deze representatief is. Een aandachtspunt is wel dat minder dan 3% van de scholen uit de aanvullende steekproef bereid bleken te zijn deel te nemen aan het normeringsonderzoek.

Conclusie:

Aan aspect N1.2.2. wordt het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1.1. Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen:

Om relevante gegevens bij de toets te genereren, is gebruik gemaakt van het programma OPLAT. Binnen dit programma wordt de coëfficiënt MAcc ('Accuracy of Measurement') berekend. Deze coëfficiënt vertoont qua interpretatie grote overeenkomst met de betrouwbaarheidscoëfficiënt uit de KTT. Deze coëfficiënt wordt in de psychometrische literatuur beschreven en als correct aangemaakt.

Conclusie:

Aan aspect B1.1 wordt het oordeel '**voldoende**' toegekend.

B1.2. Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

Bevindingen:

Er wordt verwezen naar de COTAN criteria voor toetsen voor minder belangrijke beslissingen. De interne consistentie betrouwbaarheid is, volgens deze criteria, voldoende bij een betrouwbaarheidscoëfficiënt tussen 0,70 en 0,80. Voor de toets Begrijpend lezen 3.0 groep 3 wordt deze coëfficiënt berekend als MAcc (zie B1.1) voor afname-moment E3. Aanvullend hierop wordt de standaardmeetfout vermeld. De afnamecontext van de toets leent zich, dankzij een OPLM kalibratie, voor een gesimuleerd test-hertest onderzoek onder ideale condities. De test-hertest coëfficiënt is identiek aan de MAcc. De resultaten worden aangemerkt als 'goed' (standaardmeetfout 2,43; MAcc 0,91; test-hertest 0,91).

Aanvullend hierop is de lokale meetnauwkeurigheid (misclassificaties) weergegeven in betrouwbaarheidstabellen. Uitgaande van de betrouwbaarheidstabellen worden twee indices voor de nauwkeurigheid van de classificaties gerapporteerd: de plus/minus 1 niveau-index en de marginal classification index. Uit de hoogte van de indices blijkt dat de laagst en de hoogst scorende leerlingen accuraat te classificeren zijn, maar dat tussen leerlingen in de niveaugroepen B, C en D, respectievelijk II, III en IV, minder duidelijk onderscheid te maken is.

Conclusie:

De betrouwbaarheid van de toets Begrijpend lezen 3 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft en ingestemd wordt met de beoordelingscriteria voor de betrouwbaarheid van de COTAN.

Op aspect B.1.2. wordt aan de toets Begrijpend lezen groep 3 het oordeel '**voldoende**' toegekend.

Validiteit

V1. Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items)

Bevindingen:

In groep 3 is 1 toets voor begrijpend lezen die aan het eind van groep 3 wordt afgenomen. Dit is een logische keuze omdat dan de meeste kinderen de elementaire leeshandeling hebben geleerd en geautomatiseerd. De toets voor groep 3 bestaat uit twee taken die kunnen worden afgenomen op 2 verschillende dagdelen. Iedere taak bestaat uit een aantal korte teksten met 20 opgaven (meerkeuzeopgaven). De toets kan zowel handmatig als via de computer via het computerprogramma LOVS nagekeken worden.

In de handleiding wordt door toetsmatrijzen inzicht geboden hoe de toets is samengesteld. In de toets worden twee deelvaardigheden getoetst: begrijpen en interpreteren van geschreven teksten. Dit zijn ook twee van de drie kenmerken van de taakuitvoering die uitgewerkt zijn bij de referentieniveaus voor leesvaardigheid. Het kenmerk evalueren vindt niet plaats. Dit is ook een goede keuze. Het gefundeerd leren waarderen van teksten, leren kinderen vooral in de onderbouw door in gesprek met elkaar te zijn over teksten/boeken.

In de toetsmatrijzen wordt duidelijk dat meer dan 50% van de opgaven bij de teksten gaat over vragen die direct uit de tekst te halen zijn (begrijpen en interpreteren) en in mindere mate openplaatsopgave en voorspelopgave. Gezien het leesniveau dat de meeste kinderen op die leeftijd hebben is dit goed.

De teksten in de toetsen zijn voor de helft verhalend en voor de andere helft informatief en betogend. De informatieve teksten in de toetsen zijn herkenbaar qua onderwerpen voor kinderen van die leeftijd (bijvoorbeeld kuiken, lente, pretpark). De structuur van zowel de formele (informatief, betogend) als informele (verhalende) teksten is helder. Het verhaalschema in de teksten is eenvoudig. De zinnen zijn kort, de teksten niet te lang en het woordgebruik concreet. De opgaven en meerkeuzeantwoorden zijn kort en krachtig geformuleerd en zijn goed terug te leggen op de teksten daar waar het gaat over begrijpen/interpreteren.

Kinderen lezen voor iedere taak 7 of 8 verschillende teksten achter elkaar. Kinderen in groep 3 zullen daar niet aan gewend zijn. Dit vraagt wel een voorbereiding op de toetshouding van leerlingen door leraren. Anders zou het van invloed kunnen zijn op de te behalen toetsresultaten.

Merk op dat de beoordeling van dit aspect zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat is verricht:

De toets Begrijpend lezen 3.0 groep 3 is niet bedoeld voor voorspellend gebruik. Daarmee is de criteriumvaliditeit niet van toepassing. De (psychometrische) begripsvaliditeit wordt uitgewerkt in unidimensionaliteit, itemkwaliteit, itembias, convergente en divergente validiteit en in verschillen tussen relevante subgroepen.

De resultaten van de uitgevoerde kalibratie maken het aannemelijk dat er sprake is van unidimensionaliteit. Dit betekent dat met elke willekeurige subset van items uit de gekalibreerde itembank dezelfde onderliggende vaardigheid begrijpend lezen kan worden

vastgesteld. Dit wordt tevens bevestigd door de nauwkeurigheid van de itemparameterschattingen (c is voor alle items ≤ 0.20).

De gemiddelde moeilijkheidsgraad voldoet met p -waarden die liggen tussen de .54 en .88 en een gemiddelde van .73. De toets is, bewust, niet te moeilijk gemaakt om frustratie bij de jonge leerling te voorkomen. Dit is terug te vinden in de verdeling linksscheef van de gemiddelde moeilijkheidsgraad.

De gemiddelde itemkwaliteit voldoet eveneens met R_{it} waarden die liggen tussen .34 en .62 en een gemiddelde van .47. In het onderzoek naar itembias is geen sprake van DIF (Differential Item Functioning) naar sekse.

De constructvaliditeit is uitgewerkt in convergente validiteit door de samenhang te onderzoeken met de voorgaande versie Begrijpend lezen 2.0. De divergente validiteit is onderzocht door de samenhangen met andere taalonderdelen en rekenen-wiskunde te onderzoeken. De hoge correlatie ($r=0,93$; $N=1.901$) tussen de gewogen scores gebaseerd op items uit LVS 2.0 en LVS 3.0 sterkt de eerdere COTAN waardering 'goed' of 'voldoende' voor de begripsvaliditeit. De hoge correlaties tussen Begrijpend lezen 3.0 en de meeste Taaltoetsen en de lage correlatie tussen Begrijpend lezen 3.0 en Rekenen-wiskunde vormen een ondersteuning voor de validiteit van de toets. Wat betreft de verschillen tussen relevante subgroepen scores jongere leerlingen naar verwachting iets beter dan oudere leerlingen en scores meisjes iets hoger dan jongens.

Conclusie:

Op aspect V1.1 wordt aan de toetsen begrijpend lezen 3.0 groep 3 op dit aspect het oordeel '**voldoende**' toegekend.

Het volg-aspect

VA1.1. Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Bevindingen:

Het algemene (inhoudelijke) uitgangspunt voor de toets Begrijpend lezen 3.0 groep 3 is dat de (latente) vaardigheid Begrijpend lezen kan worden opgevat als een unidimensioneel continuüm en dat elke leerling kan worden voorgesteld als een punt op dit continuüm. Hierbij wordt opgemerkt dat de elementen Begrijpen en Interpreteren van zowel formele als informele tekstgenres als de componenten van de latente vaardigheid Begrijpend lezen niet opgevat kunnen worden als te isoleren vaardigheden. Daarmee wordt de vaardigheid Begrijpend lezen, als interactie tussen deze componenten, beschouwd als één unidimensionele vaardigheid.

Uit het kalibratieonderzoek blijkt dat de items passen bij het gehanteerde IRT model en dat het model ook past voor de toets E3 als geheel. Dit betekent dat er sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Afhankelijk van het aantal items dat een leerling goed maakt, wordt er een vaardigheidsscore toegekend. Jongere leerlingen scoren iets beter dan oudere leerlingen. Tevens scoren meisjes iets hoger dan jongens.

Voor Begrijpend lezen 3.0 is een nieuwe vaardigheidsschaal ontwikkeld, waarop alle uitgebrachte en nog uit te brengen toetsen uit het Cito Volgsysteem primair en speciaal basisonderwijs Begrijpend lezen 3.0 worden gekalibreerd. Vanwege het volgmodel en de gevolgde dataverzamelingsstrategie voor de normering, worden de nieuw ontwikkelde toetsen gefaseerd uitgebracht, d.w.z. in elk schooljaar een toetspakket voor een hogere groep.

Conclusie:

Aan aspect VA1.1. wordt het oordeel **'voldoende'** toegekend.

VA1.2. Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden? Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Bevindingen:

In hoofdstuk 7 van de handleiding ('Communiceren over toetsresultaten') wordt beschreven hoe er met de verschillende gebruikers over de toetsresultaten kan worden gecommuniceerd. Hierin wordt onderscheid gemaakt tussen 'niveau' en 'groei', wat wordt onderbouwd met diverse rapportage mogelijkheden.

In de wetenschappelijke verantwoording wordt toegelicht hoe de toetsen ingezet kunnen worden om de ontwikkeling van leerlingen te volgen in de tijd, namelijk door het toetsresultaat van een leerling te vergelijken met andere leerlingen en door het toetsresultaat van een leerling te vergelijken met diens andere toetsresultaten. Voor alle vergelijkingen geldt dat uitspraken over de voortgang van leerlingen gerelativeerd moeten worden vanwege de (on)betrouwbaarheid van de toetsen. Door betrokkenen bij de toetsen Begrijpend lezen moet beseft worden dat vaardigheidsgroei zich langzaam in de tijd voltrekt.

Conclusie:

Aan aspect VA1.2. wordt het oordeel **'voldoende'** toegekend.

Inzicht in leervorderingen

I1. Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/voogden/verzorgers begrijpelijk is?

Bevindingen:

Via de portal van Cito B.V. kan gebruik worden gemaakt van rapportage-/registratieformulieren voor een leerlingrapport, groepsrapport, groepsoverzicht (overzicht van één groep leerlingen tijdens hun schoolperiode) en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken). Voor ouders is met name

het leerlingrapport of alternatief leerlingrapport informatief omdat deze rapporten van hun kind individueel de vaardigheid en de groei weergeven.

In de Handleiding wordt in hoofdstuk 7 aandacht besteed aan de wijze waarop met ouders over de toetsresultaten gecommuniceerd kan/moet worden. Met name wordt daarbij gewezen op het leerlingrapport waarin zowel het niveau van de leerling als de progressie van de leerling numeriek en grafisch gepresenteerd worden.

Daarnaast wordt de leraar gewezen op misverstanden die zich bij de interpretatie van de niveau-indelingen bij de ouders kunnen voordoen. Ook moeten zij aan ouders het verschil tussen methode-onafhankelijke en methodegebonden toetsen duidelijk maken en erop wijzen dat deze toetsen leerlingen anders (kunnen) beoordelen. De informatie biedt goede handvatten voor de gesprekken met ouders. In hoofdstuk 8 worden veelgestelde vragen behandeld die weliswaar voor de leraren bestemd zijn maar waar de antwoorden voor een deel ook informatief zijn tijdens bijvoorbeeld de tienminutengesprekken.

In de communicatie naar ouders toe over de resultaten van begrijpend lezen, is het van belang om ook de koppeling te maken naar de resultaten op technisch lezen en woordenschat omdat die van invloed zijn op de scores van begrijpend lezen.

De taalprofielen die worden weergegeven in bijlage 4, bedoeld als een eerste hulpmiddel waarmee de toetsresultaten in een breder perspectief kunnen worden geplaatst, kunnen ook behulpzaam zijn in de communicatie naar ouders toe.

Voor het interpreteren, analyseren en benutten van de toetsresultaten zijn er een tweetal bijlagen ingevoegd. In bijlage 3 bevindt zich een stappenplan voor interpretatie en analyse van toetsresultaten op leerling- en groepsniveau. En in bijlage 4 bevindt zich een stroomdiagram met zeven taalprofielen t.b.v. het opstellen van een individueel handelingsplan bij zwakke resultaten.

Over de interpretatie van toetsresultaten is ook een folder ouderinformatie beschikbaar die men via de website van het Cito kan downloaden.

Conclusie:

Op aspect I1.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 3 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

Kwaliteitsaspect	Code	Oordeel
De kwaliteit van de steekproef	<i>S1.1</i>	Voldoende
	<i>S1.2</i>	Voldoende
Normering	<i>N1.1</i>	Voldoende
	<i>N1.2</i>	Voldoende
Betrouwbaarheid	<i>B1.1</i>	Voldoende
	<i>B1.2</i>	Voldoende
Validiteit	<i>V1.1</i>	Voldoende
Volg-aspect	<i>VA1.1</i>	Voldoende
	<i>VA1.2</i>	Voldoende
Inzicht in leervorderingen	<i>I1.1</i>	Voldoende

4. Literatuurlijst

De beoordeling is gebaseerd op het volgende, door Cito B.V. aangeleverde, materiaal:

- Jolink A., Tomesen M., Hilde M., Weekers A. & Engelen R. (2015). *Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 3*. Arnhem: Cito B.V.
- Cito B.V. (2014). *Leerkrachtmap Begrijpend lezen 3.0 voor groep 3*. Arnhem: Cito B.V.