

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 Volgsysteem 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 toetsen Begrijpend Lezen 3.0 groep 5 zijn een onderdeel van het Cito Volgsysteem primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 5 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 toetsen voor groep 5 vormen dan met het de LVS toetsen Begrijpend lezen 3.0 groepen 3, 4 en de groepen 6 tot en met 8 een systeem om vast te stellen hoe goed leerlingen begrijpend kunnen lezen en hoe hun vaardigheid zich in de basisschoolperiode ontwikkelt. Onderstaande beschrijving is gebaseerd op de Handleiding.

Meetpretentie

De toetsen in de toets pakketten Begrijpend lezen 3.0 voor groep 5 van het Cito Volgsysteem primair en speciaal onderwijs zijn bedoeld om vast te stellen hoe goed een leerling geschreven teksten, met name zakelijke teksten en literaire en fictionele 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 M5-toets maken dezelfde toets.

Doelgroep

De toetsen Begrijpend lezen 3.0 groep 5 zijn bedoeld voor leerlingen in groep 5 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 5 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.

De toetsen zijn niet geschikt voor leerlingen bij wie het lezen nog niet geautomatiseerd is. Ook voor leerlingen die nog maar kort in Nederland verblijven, zijn de toetsen ongeschikt: leerlingen moeten het Nederlands voldoende beheersen en over voldoende technische leesvaardigheid beschikken om de opgaven te kunnen maken.

Gebruiksdoel en functie

Doel van de toetsen Begrijpend lezen 3.0 voor groep 5 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 toetsen 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 toetsen sluiten 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 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 5 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 groepsniveau (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
- Drie toetsen:
 - Toets E4M5 (makkelijke variant van de toets M5)
 - Toets M5 (Medio groep 5)
 - Toets E5 (Eind groep 5)
- Afnamekaarten met aanwijzingen voor de papieren afname van de toetsen
- Nakijkaarten
- Opgaveboekjes
- Tabellen voor de drie toetsen voor het bepalen van de vaardigheidsscore en – niveau
- Wetenschappelijke verantwoording

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 steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

In januari 2013 zijn in een kalibratieonderzoek M5 (proefonderzoek) 265 items voorgelegd aan 1.910 leerlingen van groep 5. De 265 items waren verdeeld over 14 boekjes (booklets). Elke deelnemende school maakte op het medio-moment één taak met 24 – 26 nieuwe opgaven, naast de LVS 2.0 toets. De nieuwe opgaven kwamen in een of twee boekjes voor en werden gemiddeld door 182 leerlingen gemaakt wat boven het minimum vereiste van 150 ligt.

In juni 2013 zijn in een kalibratieonderzoek E5 (proefonderzoek) 171 items voorgelegd aan 2.220 leerlingen van groep 5. De 171 items waren verdeeld over 6 boekjes (booklets). Elk boekje bestond uit 50 opgaven verdeeld over 2 taken. Beide taken betroffen nieuw geconstrueerde opgaven. De opgaven kwamen in één of twee boekjes voor en werden gemiddeld door 649 leerlingen gemaakt. Op het eind-moment kon de proeftoets niet gekoppeld worden aan de afname van de tweede generatie, omdat de toetsen Begrijpend lezen 2.0 geen einde-moment kennen voor groep 5.

Op grond van het kalibratieonderzoek M5 is voor het normeringsonderzoek M5 een selectie gemaakt van 202 items, verdeeld over 7 boekjes met elk 75 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. Het embedded field normeringsonderzoek M5 is toegepast op de resultaten van 1.607 leerlingen uit groep 5 van 71 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.940 leerlingen uit groep 5 van 86 scholen uit Cito dataretour.

Op grond van het kalibratieonderzoek E5 is voor het normeringsonderzoek E5 een selectie gemaakt van 127 items, verdeeld over 5 boekjes, met elk 50 tot 52 opgaven verdeeld over 2 taken. Voor E5 bestond het normeringsbestand uitsluitend uit data die verzameld werden in het embedded field normeringsonderzoek wat is toegepast op de resultaten van 1.724 leerlingen uit groep 5 van 71 scholen.

De representativiteit van de steekproeven voor de normeringsonderzoeken M5 en E5 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 voor M5 vormt een zeer goede afspiegeling van de populatieverdeling. Voor afnamemoment E5 geldt dit eveneens, uitgezonderd voor de variabele regio. De hiervoor gebruikte wegingsfactoren voldoen aan de COTAN criteria.

De effectgrootte phi is het grootst voor de variabele Regio voor het afnamemoment E5 (.174). 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. Voor afnamemoment E5 vindt hierbij een weging plaats. De plausible values voor de afnamemomenten M5 en E5 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 steekproeven zijn representatief, zijn adequaat gestratificeerd naar sekse, regio en urbanisatiegraad en geven informatie over hoe de steekproeven zich verhouden tot de populatiewaarden. De procedure voor het samenstellen van de steekproeven 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 de proefonderzoeken van januari en juni 2013 (M5 en E5) en de opgaven uit de daaropvolgende normeringsonderzoeken 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 M5 van januari 2013 zijn 265 items voorgelegd aan 1.910 leerlingen van groep 5. De 265 items waren verdeeld over 14 boekjes (booklets). Elke deelnemende school maakte op het medio-moment één taak met 24 – 26 nieuwe opgaven, naast de LVS 2.0 toets. De nieuwe opgaven kwamen in een of twee boekjes voor en werden gemiddeld door 182 leerlingen gemaakt wat boven het minimum vereiste van 150 ligt.

In het kalibratieproces E5 van juni 2013 zijn 171 items voorgelegd aan 2.220 leerlingen van groep 5. De 171 items waren verdeeld over 6 boekjes (booklets). Elk boekje bestond uit 50 opgaven verdeeld over 2 taken. Beide taken betroffen nieuw geconstrueerde opgaven. De opgaven kwamen in één of twee boekjes voor en werden gemiddeld door 649 leerlingen gemaakt, wat boven het minimum vereiste van 150 ligt. Op het eind-moment kon de proeftoets niet gekoppeld worden aan de afname van de tweede generatie, omdat de toetsen Begrijpend lezen 2.0 geen einde-moment kennen voor groep 5.

Op basis van inhoudelijke en psychometrische criteria werden 50 items voor elk van de toetsen Begrijpend lezen 3.0 groep 5 geselecteerd (M5, E5 en E4M5). De 50 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 en de rit-waarde bepaald. Voor de normeringsonderzoeken M5 en E5 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 M5 werd gebruik gemaakt van resultaten van 1.607 leerlingen uit groep 5 van 71 scholen en van dataretour van 1.940 leerlingen uit groep 5 van 86 scholen. Voor het normeringsonderzoek E5 werd uitsluitend gebruik gemaakt van data uit een embedded field normeringsonderzoek, wat is toegepast op de resultaten van 1.724 leerlingen uit groep 5 van 71 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 de proefonderzoeken is adequaat. Het volledige design van de toets M5 en E5 zijn 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 de afnamemomenten M5 en E5.

Op grond van het kalibratieonderzoek M5 is een selectie gemaakt van 202 items, verdeeld over 7 boekjes met elk 75 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. Het embedded field normeringsonderzoek M5 is toegepast op de resultaten van 1.607 leerlingen uit groep

5 van 71 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.940 leerlingen uit groep 5 van 86 scholen uit Cito dataretour.

Op grond van het kalibratieonderzoek E5 is een selectie gemaakt van 127 items, verdeeld over 5 boekjes, met elk 50 tot 52 opgaven verdeeld over 2 taken. Voor E5 bestond het normeringsbestand uitsluitend uit data die verzameld werden in het embedded field normeringsonderzoek wat is toegepast op de resultaten van 1.724 leerlingen uit groep 5 van 71 scholen.

Voor de afnamemomenten M5 en E5 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.

De normen voor de toetsen Begrijpend lezen groep 5 zijn geldig tot en met 2024.

Conclusie:

Er is sprake van relatieve normen, de steekproeven zijn representatief en groot genoeg. Daarmee wordt aan aspect N1.2.1. het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit van de steekproeven is besproken bij punt S1.1. Hier wordt reeds geconstateerd dat deze representatief zijn. Een aandachtspunt is wel dat 4% van de scholen uit de aanvullende steekproef M5 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 5 wordt deze coëfficiënt berekend als MAcc (zie B1.1) voor de afname momenten M5 en E5. 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 laten zich lezen als standaardmeetfout (E4M5 2,715; M5 2,885; E5 2,978), MAcc (E4M5 0,90; M5 0,89; E5 0,88) en test-hertest simulatie (E4M5 0,90; M5 0,89; E5 0,88) en worden aangemerkt als 'goed'.

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 toetsen Begrijpend lezen 5 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft.

Op aspect B.1.2. wordt aan de toets Begrijpend lezen 3.0 groep 5 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:

Zoals uit de algemene beschrijving hierboven blijkt, is de Handleiding voor de leerkracht zeer leesbaar en volledig, en doet hij recht aan vragen bij de vorige editie vanuit het veld. De Handleiding gaat uitgebreid in op het voorbereiden, afnemen, nakijken en verwerken van de toetsen en daarmee op hun plek in het onderwijsleerproces. De keuze voor en gebruiksmogelijkheden van de toetsversies voor diverse deeldoelgroepen en meetmomenten in groep 5 wordt grondig toegelicht. Daarnaast is aandacht voor de mogelijkheid tot foutenanalyse, interpreteren en analyseren van resultaten op leerling- en groepsniveau, en voor communicatie met leerling en ouders over de resultaten.

Elk van de drie toetsen bestaat uit twee taken met teksten; elke taak bevat in totaal 25 meerkeuzeopgaven. Afname is in principe klassikaal. De wijze waarop een leerkracht de toetsen 'op maat' in kan zetten en evalueren wordt uitgebreid toegelicht en tevens verbonden met de rapportage van de reguliere afname en eventueel naar een handelingsplan en schoolplan.

Bij de toetsconstructie is aantoonbaar aangesloten bij diverse relevante bronnen, waaronder de tussendoelen en leerlijnen en de Referentieniveaus Taal. De keuze van teksten in de toetsboekjes is een goede afspiegeling van het aanbod aan teksten waar leerlingen in het onderwijs mee in aanraking komen, en van de tekstsoorten zoals die in het Referentiekader genoemd worden. In de wetenschappelijke verantwoording is de keuze en verdeling van tekstsoorten en soorten opgaven (openplaats-opgaven (cloze); tekstopgaven en voorspelopgaven) onderbouwd. Niveau van de teksten en vraagstellingen in het toetsmateriaal geven geen aanleiding tot fundamentele vragen over of twijfel aan de validiteit.

Hoewel de opmerking geplaatst kan worden dat de bevraging in de openplaatsopgaven naar 'wat past het best...' in een groot deel van de gevallen naast het correcte antwoord ook een alternatief geeft dat zeer wel verdedigbaar is, laat de psychometrische onderbouwing zien dat de itemkwaliteit van de opgenomen openplaatsopgaven goed is. Verder discrimineren deze opgaven goed tussen vaardige en minder vaardige lezers. Bij die items waarbij de meer vaardige lezers structureel het foute antwoord kozen, zijn de alternatieven aangepast. Deze aangepaste items zijn nogmaals onderzocht en de itemkwaliteit is door deze aanpassing verbeterd.

Merk op dat de beoordeling van dit aspect zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat verricht is inzake validiteit en dan specifiek begripsvaliditeit.

De toets Begrijpend lezen 3.0 groep 5 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 moeilijkheidsgraden van de toetsen voldoen (p-waarden E4M5 met range van .55 - .92 en een gemiddelde van .76; M5 range .45 - .87 en gemiddelde .71; E5 range .45 - .85 en gemiddelde .69). De toets E4M5 is geschikt gemaakt voor leerlingen met een vertraagde ontwikkeling. Dit is terug te vinden in de verdeling linksscheef van de gemiddelde moeilijkheidsgraad.

De gemiddelde itemkwaliteit voldoet eveneens. Voor twee items uit de toets E4M5, zes items uit de toets M5 en twee items uit de toets E5 ligt de Rit-waarde onder de .30. De gemiddelden komen uit in de range van .38 tot .41. In het onderzoek naar itembias is er

bij één item van elk van de toetsen E4M5, M5 en E5 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 (M5: $r=0,89$; $N=1.767$) 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. Voor E5 was deze analyse niet mogelijk, omdat er in LVS 2.0 geen vergelijkbare toets bestaat.

De hoge correlaties tussen Begrijpend lezen 3.0 en de meeste Taaltoetsen (m.u.v. Technisch lezen – DMT M5, $r=0,40$) en de lagere 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:

Aan aspect V1. wordt 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 5 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 toetsen M5 en E5 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 vooral 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. Vooral 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.

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 5 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

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

4. Literatuurlijst

- Tomesen, M., Weekers, A., Hilde, M., Jolink, A. & Engelen R. (2016). *Wetenschappelijke verantwoording Begrijpend lezen 3.0 groep 5*. Arnhem: Cito B.V.
- Cito (2015). *Cito Volgstelsel primair en speciaal onderwijs. Begrijpend Lezen 3.0 groep 5*. Arnhem: Cito