

## **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 onderwijs 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 voor groep 7 zijn een onderdeel van het Cito Volgsysteem primair onderwijs en zijn bedoeld voor leerlingen in groep 7 van het primair onderwijs. Onderstaande beschrijving is gebaseerd op de Handleiding.

### Meetpretentie

De toetsen in de toetspakketten Begrijpend lezen 3.0 voor groep 7 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 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.

### Doelgroep

De toetsen Begrijpend lezen 3.0 groep 7 zijn bedoeld voor leerlingen in groep 7 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 7 en voor leerlingen met een ontwikkelingsachterstand en/of extra onderwijsbehoeften (dit kan via toetsen op maat). 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 die nog maar pas in Nederland verblijven, zijn de toetsen ongeschikt. Leerlingen moeten het Nederlands voldoende beheersen en moeten over voldoende technische leesvaardigheden beschikken om de opgaven te kunnen maken voordat de toetsen Begrijpend lezen bij hen kunnen worden afgenomen.

Voor het toetsen van de begripsmatigheid kan bij dyslectische kinderen tevens de toets Begrijpend Luisteren afgenomen worden.

### Gebruiksdoel en functie

Doel van de toetsen Begrijpend lezen 3.0 voor groep 7 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van begrijpend lezen. De toetsen maken 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.

Naast de hierboven genoemde mogelijkheden heeft deze toets ook de mogelijkheid om de vaardigheidsscores af te zetten tegen het referentieniveau Lezen.

#### 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 omdat er 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 7 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
- Twee toetsen:
  - Toets M7 (Medio groep 7)
  - Toets E7 (Eind groep 7)
- Afnamekaarten met aanwijzingen voor de afname van de toetsen
- Nakijkkaarten
- Tabellen voor de twee 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 steekproef**

### S1.1. Is de steekproef representatief?

#### *Bevindingen:*

Sinds 2013/2014 wordt door Cito een nieuwe werkwijze voor het normeren van leerlingvolgsystemen toegepast. Een belangrijk aspect van die werkwijze is om de normeringssteekproef (ongeveer 50 procent) te baseren op gegevens uit het zogenaamde embedded field normeringsonderzoek en (ongeveer 50 procent) op gegevens uit Cito dataretour die afkomstig zijn van de tweede generatie toetsen. Bij de normering van de derde generatie LVS-toetsen is rekening gehouden met de variabelen regio, urbanisatiegraad, schooltype en sekse, waarbij de eerste drie variabelen op schoolniveau zijn gedefinieerd en de laatste variabele op leerlingniveau.

Op pagina 51-53 van de wetenschappelijke verantwoording wordt een algoritme beschreven dat een representatieve steekproef moet garanderen. Niettemin werd er naast die garantie een controle op de representativiteit uitgevoerd door de populatieverdelingen van gegevens uit DUO te vergelijken met de steekproefverdelingen van afnamemomenten/toetsen M7 en E7. Tabel 4.5 en 4.6 presenteren de resultaten van de representativiteitsanalyses van respectievelijk M7 en E7. De tabellen laten zien dat voor M7 de steekproefverdeling weinig afwijkt van de populatieverdeling. Voor E7 zijn de afwijkingen groter, vooral voor wat betreft urbanisatiegraad en schooltype. Uit de effectgroottes blijkt dat die voor afnamemoment M7 zeer klein (M7: ver onder .10) zijn en de conclusie is dan ook dat de normeringssteekproef voor afnamemoment M7 een representatieve afspiegeling vormt van de populatie. Dit laatste geldt ook voor E7, met uitzondering van de variabele schooltype (over- en ondervertegenwoordiging) en urbanisatiegraad. Bij de bepaling van de normering is deze variabele gewogen. De wegingsfactoren voldoen aan de criteria uit het COTAN-beoordelingssysteem (oververtegenwoordiging is geen probleem en bij ondervertegenwoordiging is een factor 2 acceptabel).

In tabel 4.4 van de wetenschappelijke verantwoording worden de aantallen leerlingen gegeven die meegenomen zijn bij de normering. Voor M7 betreft dat 2.451 leerlingen (1.073 leerlingen uit de steekproef en 1.378 leerlingen via dataretour) afkomstig van 108 scholen en voor E7 1.191 leerlingen uit de steekproef afkomstig van 56 scholen. Voor E7 was aanvulling met gegevens uit dataretour niet mogelijk omdat de tweede generatie toetsen geen toets voor afnamemoment E7 kent.

#### *Conclusie:*

De representativiteit van de steekproef is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. Op aspect S1.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 het oordeel '**voldoende**' toegekend.

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

#### *Bevindingen:*

Bij de ontwikkeling van de toetsen Begrijpend lezen 3.0 is geprobeerd om door middel van embedded field designs bias in de normen te vermijden door de afnamesituatie waarin de toets wordt afgenomen tijdens het normeringsonderzoek zoveel mogelijk te laten lijken op de afnamesituatie na uitgave van de toets.

Het normeringsonderzoek gaat uit van een onvolledig dataverzamelingsdesign. In totaal lieten scholen hun leerlingen op afnamemoment M7 vier taken maken; een taak uit LVS 2.0 en drie taken met nieuw materiaal voor de derde generatie. Omdat er voor afnamemoment E7 geen toets Begrijpend lezen LVS 2.0 bestaat, maakten leerlingen op dit afnamemoment uitsluitend drie taken met nieuw materiaal. In figuur 4.3 en 4.4 worden de embedded field designs weergegeven voor respectievelijk toetsversie M7 en toetsversie E7. Op afnamemoment M7 maakten de leerlingen volgens het design een taak uit de M7-toets LVS 2.0. Daarnaast maakte elke leerling drie taken met nieuwe opgaven voor LVS 3.0. Op afnamemoment E7 maakten leerlingen drie taken met nieuwe opgaven voor LVS 3.0. Sommige leerlingen maakten twee taken met nieuwe opgaven voor LVS 3.0 en daarnaast een ankertaak 'referentieset' om de prestatiestandaard van de referentieniveaus over te kunnen brengen. De designs laten zien dat de (toets)boekjes stevig verankerd zijn.

In het normeringsonderzoek M7 zijn 228 items voorgelegd aan 1.221 leerlingen medio groep 7, verdeeld over 6 boekjes volgens het design in figuur 4.3. Elk boekje bestond uit 99 tot 103 opgaven verdeeld over 4 taken. De 25 opgaven in de LVS 2.0 taak werden door alle 1.221 leerlingen gemaakt. De overige opgaven kwamen in twee of drie boekjes voor en werden gemiddeld door 456 leerlingen gemaakt. Bij alle opgaven werd voldaan aan de eis dat deze minimaal bij 150 leerlingen moesten zijn afgenomen.

In het normeringsonderzoek E7 zijn 205 items voorgelegd aan 1.534 leerlingen eind groep 7, verdeeld over 6 boekjes volgens het design in figuur 4.4. Elk boekje bestond uit 75 tot 78 opgaven verdeeld over 3 taken. De opgaven kwamen in twee of drie boekjes voor en werden gemiddeld door 437 leerlingen gemaakt. Bij alle opgaven werd voldaan aan de eis dat deze minimaal bij 150 leerlingen moesten zijn afgenomen.

Uit het kalibratieonderzoek blijkt dat voor de twee toetsen zowel de items, gegeven de M- en S-toetsen, als de gehele toetsen, gegeven de R1c-toetsen, passen bij het itemresponsstheoriemodel OPLM. Dit betekent dat er sprake is van een eendimensionale vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden. Op grond van de gegevens uit de kalibratie zijn de definitieve selecties van items gemaakt voor de uitgaven van de toetsen Begrijpend lezen 3.0 M7 en E7.

#### *Conclusie:*

Het onvolledige dataverzamelingsdesign is adequaat. Op aspect S1.2 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 het oordeel '**voldoende**' toegekend.

## **Normering**

### N1.2.1. Zijn de normgroepen groot genoeg?

#### *Bevindingen:*

In tabel 4.9 van de wetenschappelijke verantwoording wordt voor M7 en E7 de normtabel met relatieve normen op leerling-niveau gepresenteerd. Voor aantallen leerlingen en scholen zie S1.1. Voor beide afnamemomenten worden vaardigheidsverdelingen gepresenteerd, d.w.z. gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75 en P80. Met behulp van deze percentielen kunnen de twee niveau-indelingen (asymmetrisch: A t/m E en symmetrisch: I t/m V) opgesteld worden.

In tabel 4.12 van de wetenschappelijke verantwoording wordt voor M7 en E7 de normtabel met relatieve normen op schoolniveau gepresenteerd. Voor beide toetsen en afnamemomenten worden vaardigheidsverdelingen gepresenteerd, d.w.z. gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75 en P80.

*Conclusie:*

De normgroepen zijn groot genoeg (zie S1.1). Op aspect N1.2.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

*Bevindingen:*

De representativiteit van de steekproeven werd in S1.1 en 2.1 besproken en daar werd geconstateerd dat de steekproeven representatief waren voor regio, urbanisatiegraad, schooltype en sekse. Voor Begrijpend lezen 3.0 groep 7 geldt dat de normen geldig zijn tot en met 2026. Daarnaast monitort Cito periodiek de normering. Jaarlijks wordt aan de hand van representatieve afnamedata nagegaan of er systematisch verschuivingen in het prestatieniveau plaatsvinden. Indien nodig wordt de normering aangepast.

*Conclusie:*

Op aspect N1.2.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 het oordeel '**voldoende**' toegekend.

**Betrouwbaarheid**

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

*Bevindingen:*

In hoofdstuk 5 van de wetenschappelijke verantwoording wordt beschreven hoe, gebruikmakend van het feit dat de items uit de toets geschaald zijn met behulp van de OPLM software, een betrouwbaarheidscoëfficiënt, de MAcc ('Accuracy of Measurement'), berekend kan worden die qua interpretatie grote overeenkomst vertoont met de betrouwbaarheidscoëfficiënt uit de klassieke testtheorie. Deze coëfficiënt wordt in de psychometrische literatuur beschreven en als correct aangemerkt.

*Conclusie:*

De betrouwbaarheidsgegevens worden correct berekend. Op aspect B1.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 het oordeel '**voldoende**' toegekend.

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

*Bevindingen:*

Voor M7 en E7 zijn drie betrouwbaarheidsgegevens berekend: standaardmeetfout, MAcc en een gesimuleerde test-hertest betrouwbaarheidscoëfficiënt. Voor M7 zijn de gegevens gelijk aan 11.01, .89 en .89, en voor E7 gelijk aan 10.25, .88 en .88. De auteurs van de wetenschappelijke verantwoording verwijzen naar het beoordelingssysteem van de COTAN waar voor tests die geen zware consequenties voor leerlingen hebben, zoals de toetsen Begrijpend lezen 3.0 groep 7, een betrouwbaarheidscoëfficiënt van meer dan .80 als 'goed' aangemerkt wordt.

Naast klassieke betrouwbaarheidscoëfficiënten is ook de lokale betrouwbaarheid en de meetnauwkeurigheid onderzocht. De betekenis van de meetnauwkeurigheid voor de beslissingen die met de toets genomen worden, is af te leiden uit de betrouwbaarheidstabellen van twee niveauverdelingen (I t/m V en A t/m E) voor M7 en E7. Tabel 5.2a en tabel 5.2b van de wetenschappelijke verantwoording laten zien dat de leerlingen met de hoogste en de laagste niveau-indicatie voldoende betrouwbaar geïnclassificeerd worden. Voor de andere niveaus geldt dat de leerlingen onvoldoende betrouwbaar geïnclassificeerd worden. Aan de betrouwbaarheidstabellen moet de eis gesteld worden dat alle waarden op de hoofddiagonaal minstens .70 zijn omdat de classificaties dan voldoende betrouwbaar zijn in het geval de uitslag op de toets geacht wordt geen belangrijke consequenties voor de leerling te hebben. Heeft de toets wel belangrijke consequenties voor de leerling dan moeten voornoemde waarden minstens .80 zijn.

De informatie uit de betrouwbaarheidstabellen kan worden samengevat met twee indices: de marginal classification accuracy en de plus/minus 1 niveau-index. Uit de indices blijkt dat gegeven het gebruiksdoel van de toetsen de classificaties voldoende betrouwbaar zijn, maar dat er wel rekening mee moet worden gehouden dat er altijd sprake zal zijn van misclassificaties ter grootte van veelal maximaal 1 niveau.

#### *Conclusie:*

De betrouwbaarheid van de toetsen Begrijpend lezen 3.0 groep 7 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft. Op basis van het voorgaande wordt op aspect B1.2 aan de toetsen Begrijpend lezen 3.0 groep 7 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:*

Met de toets bieden de toetsontwikkelaars een uitwerking voor groep 7 van het PO en SO van Begrijpend lezen. Omdat de groei vanaf groep 5 afneemt ten opzichte van de jaren ervoor, kiest de leraar één afnamemoment per jaar. Dit kan zijn halverwege groep 7 (M7) of aan het einde van het schooljaar (E7).

Het tekstboekje van de toets in groep 7 is net als in groep 6 apart van het opgavenboekje. Het opgavenboekje bestaat uit 3 taken waardoor het mogelijk is op 3 aparte momenten binnen bijvoorbeeld een week een taak af te nemen om te zorgen voor voldoende concentratie. De opgaven gericht op informatiebronnen staan niet in het tekstboekje maar in het opgavenboekje. Omdat dit korte teksten of schema's zijn is dit inderdaad voor kinderen overzichtelijker.

Het tekstboekje van toets 7M en 7E is sober, bijna saai, vormgegeven. Hoewel illustraties niet teveel mogen afleiden, is de uitwerking zoals die voorligt wel minimaal met illustraties. Er zit een groot verschil tussen de vormgeving van de toets en de beeldcultuur waarin kinderen op dit moment leven. Dit is wel een bewuste keuze geweest omdat het voor alle leerlingen dus ook kinderen met autisme of ADHD niet teveel mag afleiden.

De keuze van teksten en onderwerpen is divers en passend bij de belevingswereld van leerlingen in groep 7.

De toelichting op de inhoud van elk van de toetsonderdelen in de wetenschappelijke verantwoording en de handleiding is uitgebreid en helder beschreven. Begrijpend lezen wordt in het Cito LOVS getoetst vanaf groep 3, maar vanaf groep 6 voor het eerst op elk van de vier deelvaardigheden begrijpen, interpreteren, samenvatten en opzoeken. Deze deelvaardigheden worden ook in de toetsen 7M en 7E getoetst met behulp van verschillende soorten teksten en opgaven.

Samenvatten valt niet onder referentieniveau 1F. Voor de doorgaande lijn en omdat samenvatten in het basisonderwijs vanaf groep 5/6 wordt aangeboden, heeft het ook een plek in de toets van groep 7 wat een goede keuze is. Naast de opgaven begrijpen en interpreteren met accent op tekstopgaven, die het grootste deel van de toetsen 7M en 7E beslaan, is dus ook een aantal opgaven rond samenvatten en opzoeken opgenomen.

In de niveaubeschrijving van het lezen van zakelijke teksten (Referentieniveaus Taal) is op niveau 1F het zoeken van informatie via internet nog niet opgenomen. Door de verdergaande digitalisering is door de toetsmakers ervoor gekozen om in de toetsen wel opgaven, op papier weliswaar, op te nemen rond digitale informatiebronnen. Dit is een waardevolle toevoeging.

Het zou te overwegen zijn om bij de teksten in de tekstboekjes aan te geven wat voor genre het is naast de titel die wordt vermeld. Het aangeven van het genre richt de lezer bij de start van het lezen van de tekst, wat een belangrijke vaardigheid is bij begrijpend lezen om teksten snel te doorgronden en het leesdoel te bepalen. Bij enkele teksten is één van de vragen wat voor soort tekst het is. Je kunt dan bij deze teksten ervoor kiezen geen genre neer te zetten. De toetsmakers hebben er bewust voor gekozen dit niet te doen en voor een andere insteek te kiezen namelijk leerlingen dit zelf al gaande te tekst te laten herkennen. Dit is ook terug te zien in de vormgeving van de teksten. Deze zijn als platte tekst weergegeven en uit hun authentieke context gehaald.

In de toetsen 7M en 7E worden bij enkele opgaven (zie ook bijlage met opmerkingen bij items van de toetsen) op basis van een aangeboden tekst gevraagd wat voor type tekst het is (informatief, instructie, betogend, etc.). Het lijkt logischer om te vragen naar tekstgenre (herkenbare vormen van communicatie zoals nieuwsbericht, verslag etc.) gekoppeld aan het doel van de tekst. Het ligt meer voor de hand dat kinderen tekstgenres herkennen gekoppeld aan het doel ervan bijvoorbeeld dat een beschrijving van een recept een ander doel heeft dan bijvoorbeeld een verslag zonder dit in antwoorden op items alleen te moeten koppelen aan typering van teksten. Hierdoor richt je kinderen op de verschillende communicatieve functies van tekstgenres en maak je ze ook meer bewust ervan.

In de Handleiding is de organisatorische kant van de afname en verwerking en interpretatie van de toetsresultaten, met extra aandacht voor de afname-instructies duidelijk uitgewerkt. Er worden goede handvatten gegeven voor de analyse- en verdiepingsvragen vanuit meerdere invalshoeken om te komen tot een juiste interpretatie en uiteindelijk ook aanpak. Op pagina 38 van de handleiding wordt gesproken over het clusteren van leerlingen op basis van de vaardigheidsscores in niveaugroepen. Het woord 'niveaugroepen' heeft jarenlang een negatieve connotatie gehad (en nog steeds) met beelden erbij van kinderen die te lang in niveaugroepen blijven zitten, weinig leren van

elkaar als kinderen op een laag niveau presteren etc. De beschrijvingen bij 1E en 1F op pagina 38 dragen dit gevaar enigszins in zich. De voorbeelden die worden gegeven weer minder. Het is goed om nog eens kritisch naar deze teksten te kijken en na te gaan of ook het woord niveaugroep niet vraagt om een ander woord vanwege de historische verkeerde connotatie.

Op verschillende plekken in de handleiding wordt aangegeven dat bij analyse van de toetsresultaten ook de relatie moet worden gelegd met de toetsresultaten op technisch lezen, begrijpend luisteren en woordenschat om de juiste stappen te ondernemen. Het is belangrijk dat de toetsontwikkelaars dit doen omdat begrijpend lezen niet los kan worden gezien van technisch lezen, begrijpend luisteren en woordenschat.

De wetenschappelijke verantwoording geeft ook ten behoeve van de onderwijskundige beoordeling een compleet beeld, is gedegen en verhelderend. De tekst maakt duidelijk dat het Referentiekader Taal (Domein 2: Lezen), de kerndoelen, bestaande tussendoelen lezen, de daarop aansluitende SLO publicatie die de leerstoflijnen lezen beschrijft, wetenschappelijke publicaties rond begrijpend lezen en de analyse van methoden van begrijpend lezen de bronnen voor de indeling, opbouw en inhoud van de toetsen zijn geweest.

De beoordeling van validiteit beperkt zich hierna tot het statistisch/psychometrisch onderzoek dat verricht is inzake validiteit en dan specifiek begripsvaliditeit.

In de wetenschappelijke verantwoording wordt verslag gedaan van onderzoek naar verschillende aspecten van begripsvaliditeit: unidimensionaliteit, itemkwaliteit, itembias, convergente en divergente validiteit en verschillen tussen relevante subgroepen. Hiervoor is verwezen naar het kalibratieonderzoek van de toetsen waarvan de conclusie is dat het aannemelijk is dat er sprake is van unidimensionaliteit en dat de items één latente trek of construct meten die we, gezien ook de uitkomsten van het hierna genoemde onderzoek, kunnen aanduiden als 'vaardigheid begrijpend lezen'. Hiermee is dus voldaan aan de noodzakelijke voorwaarde voor begripsvaliditeit.

Uit het onderzoek naar itemkwaliteit bleek dat de gemiddelde moeilijkheidsgraad van de toets M7 op het vooraf gewenste niveau lag, namelijk tussen .50 en .89. De gemiddelde moeilijkheidsgraad van de toets E7 lag tussen .37 en .92. Daarmee wijkt de gemiddelde moeilijkheidsgraad licht af van het gewenste niveau tussen .40 en .90. Bij het opnemen van twee iets te gemakkelijke (tussen .37 en .40) en één iets te moeilijke (.92) opgave werden inhoudelijke overwegingen geprefereerd boven psychometrische. Met een gemiddelde moeilijkheid van .71 voor beide toetsen (M7 en E7) zijn de toetsen als geheel niet te moeilijk en wordt voorkomen dat de leerling gefrustreerd raakt tijdens de toetsafname. De gemiddelde Rit-waarde is voor M7 gelijk aan .35 en voor E7 gelijk aan .33, waarbij een Rit-waarde van .30 door COTAN als goed gekwalificeerd wordt.

Onderzoek naar de samenhang tussen de toetsen Begrijpend lezen met andere LVS toetsen zoals Technisch lezen M7, Spelling M7, Rekenen-Wiskunde M7, en Woordenschat M7 lieten correlaties zoals verwacht zien, d.w.z. de convergente en divergente validiteit was voldoende. De data geven aan dat er met de toetsen Begrijpend lezen 3.0 gemeten wordt wat men beoogt te meten, namelijk begrijpend lezen.

Onderzoek naar Differentieel Item Functioneren (DIF) met betrekking tot jongens en meisjes per toetsopgave liet zien dat er geen sprake was van itembias met betrekking tot



seks. Uit onderzoek naar verschillen tussen relevante subgroepen kwam naar voren dat meisjes conform verwachting iets hoger scoren dan jongens (-.25 voor M7 en -.16 voor E7). Uit aanvullend onderzoek naar verschillen tussen relevante subgroepen bleken de resultaten aan te sluiten bij de verwachtingen die op grond van theoretische inzichten en eerder onderzoek konden worden geformuleerd: jongere/versnelde leerlingen doen het aanzienlijk beter dan reguliere leerlingen en die doen het weer iets beter dan oudere / vertraagde leerlingen. Wel wordt opgemerkt dat voor dit onderzoek gebruik is gemaakt van slechts 6 leerlingen M7 en 10 leerlingen E7. Meisjes scoren op beide afnamemomenten iets hoger dan jongens. Leerlingen met een leerlinggewicht van .30 en 1.20 scoren op zowel M7 als E7 lager dan leerlingen met leerlinggewicht .00, terwijl leerlingen met leerlinggewicht .00 het hoogst scoren en leerlingen met leerlinggewicht .30 bij beide afnamemomenten hoger scoren dan leerlingen met leerlinggewicht 1.20.

*Conclusie:*

Op aspect V1.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 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:*

Uit het kalibratieonderzoek blijkt dat de items passen bij het gehanteerde IRT model en dat het model ook past voor de toetsen M7 en E7 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.

*Conclusie:*

Op aspect VA1.1 wordt aan de toetsen Begrijpend lezen 3.0 groep 7 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 voor de toetsen Begrijpend lezen 3.0 groep 7 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 leerkracht 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 leerkrachten 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 7 het oordeel '**voldoende**' toegekend.

### 3. Verzamelstaat

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

### 4. Literatuurlijst

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

- Tomesen M., Engelen R. en Hiddink L. (2018). *Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 7*. Arnhem: Cito B.V.
- Cito B.V. (2017). *Leerkrachtmap Begrijpend lezen 3.0 voor groep 7*. Arnhem: Cito B.V.