

## 1. Algemene informatie

In september 2022 heeft Cito BV in overleg met de Expertgroep Toetsen PO een appendix aangeleverd bij de Wetenschappelijke Verantwoording 3.0 t.b.v beoordeling van de nieuwe inhoud van de *Leerling in Beeld*-toetsen alsook de tussentijdse check op het functioneren van de opgaven.

De oorspronkelijke beoordeling van Cito LiB 3.0 is gepubliceerd op de website van de Expertgroep; waar relevant wordt naar deze beoordeling verwezen.

### Algemeen en meetpretentie

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.

Het begrijpen van geschreven teksten is een vaardigheid die traditioneel wordt gemeten met leesbegriptoetsen. Ook de toets Begrijpend lezen 3 t/m 5 van het Cito Volgsysteem primair en speciaal onderwijs beoogt die vaardigheid te meten en de opgaven in deze toets zijn er dan ook operationalisering van. De toetsen Begrijpend lezen zijn bedoeld om vast te stellen hoe goed een leerling geschreven teksten kan begrijpen en hoe de vaardigheid in begrijpend lezen van de leerling zich in de loop van de jaren ontwikkelt.

### Inhoud van het toetspakket

- Toetsboekjes
- Afname-instructies
- Nakijkaarten
- Notitie normering toetsen Leerling in beeld
- Instructie Computerprogramma LOVS en evaluatie digitale LVS-toetsen
- 2 excelbestanden met psychometrische data

## 2. Beoordeling van de kwaliteitsaspecten

*De beoordeling vindt plaats volgens het 'Beoordelingskader voor instrumenten binnen leerlingvolgsystemen (LVS)', 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), Prof. Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en Liza Kozłowska MA (secretaris).*

*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 kwaliteit van de dataverzameling**

#### S1 Is de steekproef van leerlingen representatief?

##### *Bevindingen:*

In eerdere psychometrische beoordelingen van LVS 3.0 toetsen Begrijpend lezen voor groep 3 t/m 5 (in het vervolg 3.0-toetsen genoemd) is geconstateerd dat de steekproeven van leerlingen representatief waren (zie website Expertgroep Toetsen PO onder 'Kwaliteitsoordelen van LVS-instrumenten'). De representativiteit van de steekproeven voor de normeringsonderzoeken werden telkens onderzocht met betrekking tot de achtergrondvariabelen regio, urbanisatiegraad, schooltype en sekse. Bij regio werd uitgegaan van de vier landsdelen / regio's van de CBS-indeling (noord, zuid, oost, west). Bij urbanisatiegraad werd uitgegaan van een tweedeling in stad en platteland, afgeleid van de CBS-indeling naar vijf niveaus van verstedelijking. Bij schooltype werd uitgegaan van de formatiegewichten volgens OCW, waarin drie niveaus worden onderscheiden die gebaseerd zijn op het opleidingsniveau van de ouders. Bij sekse werd een tweedeling gemaakt naar jongens en meisjes. De steekproefverdelingen weken over het algemeen weinig af van de populatieverdelingen. De effectgroottes coëfficiënt phi ( $\phi = \sqrt{\text{Chi-kwadraat}/N}$ ) lagen grotendeels onder de 0.10 en waren daarmee zeer klein (Cohen, 1988).

##### *Conclusie:*

Omdat voor de nieuwe Leerling in beeld Begrijpend lezen groep 3 t/m 5 (in het vervolg Lib-toetsen genoemd) dezelfde steekproeven werden gebruikt als voor de steekproeven voor de bestaande LVS 3.0 toetsen Begrijpend lezen voor groep 3 t/m 5, kan worden geconcludeerd dat de steekproeven voor de Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 representatief zijn, adequaat gestratificeerd zijn naar sekse, regio, schooltype en urbanisatiegraad en informatie geven 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 het oordeel '**voldoende**' toegekend aan de Lib-toetsen Begrijpend lezen voor groep 3 t/m 5.

#### S2 In geval van een onvolledig dataverzamelingsdesign: is het design adequaat?

##### *Bevindingen:*

De normeringsonderzoeken zijn uitgevoerd in de periode januari 2019 t/m juni 2022. Tijdens het normeringsonderzoek van de Lib-toets Begrijpend lezen groep 3 t/m 5 is enerzijds dezelfde werkwijze gehandhaafd als bij de bestaande LVS 3.0 toetsen, waarin leerlingen toetsen maken tijdens de reguliere M- en E-afnamemomenten (E3,

M4, E4, M5 en E5). Ook is er bij deze toetsen vernieuwd materiaal toegevoegd aan de bestaande toetscyclus in de vorm van nieuwe items (in het vervolg Lib-items genoemd). Teneinde een reëel beeld te krijgen van leerlingprestaties tijdens reguliere afnames is er anderzijds gebruikgemaakt van dataretour van 3.0-afnames. In Tabel 1 op pag. 11 van het document 'Normering toetsen Leerling in beeld (september 2022)' is gedetailleerd weergegeven welke onderzoeken (E3, M4, E4, M5 en E5) wanneer hebben plaatsgevonden.

De nieuwe Lib-items en bestaande items voor Begrijpend lezen groep 3 t/m 5 die in een nieuwe lay-out zijn gestoken (in het vervolg Lay-out-3.0-items genoemd) zijn tegelijk afgenomen met reguliere afnames van bestaande items voor Begrijpend lezen groep 3 t/m 5 (in het vervolg 3.0-items genoemd) in de uitgaven van de toetsen van het nieuwe Leerling in beeld volgsysteem. Hierdoor verzekert Cito BV zich ervan dat de geschatte itemkarakteristieken zo passend mogelijk zijn. Omdat de nieuwe Lib-items onder reguliere afnamecondities zijn afgenomen, is een passende inschatting van de moeilijkheid en het discriminerend vermogen van deze nieuwe Lib-items mogelijk onder het gehanteerde IRT-model OPLM.

Hierbij dient echter wel aangetekend te worden dat de periode waarin de normeringsonderzoeken voor Lib-toetsen Begrijpend lezen groep 3 t/m 5 van het nieuwe volgsysteem Leerling in beeld werden uitgevoerd (januari 2019 t/m juni 2022), samenviel met de coronapandemie. Met name de onderzoeken die in de eerste helft van 2021 zijn uitgevoerd, hadden te lijden onder een tegenvallend aantal observaties per item. Sommige items zijn slechts 150 keer afgenomen. Hoewel dit voldoende is om te beoordelen of het een deugdelijk item is, is de onzekerheid rondom de schatting van de itemparameters relatief groot. Het valt daarom niet uit te sluiten dat op een later moment moet worden geconstateerd dat andere itemparameters een betere beschrijving geven van het functioneren van het item. Komend schooljaar (2022-2023) zullen aanvullende analyses gedaan worden op de Lib-toetsen Begrijpend lezen groep 3 t/m 5 in de uitgave en worden de itemparameters voor items die toch wat anders blijken te functioneren bijgesteld voor het schooljaar daarop. Cito BV zal dit dus blijven monitoren a.d.h.v. dataretour, ook in de verdere toekomst.

Steeds was het uitgangspunt bij de samenstelling van de nieuwe Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 dat in de nieuwe toetsen voldoende – minimaal 20 – LVS 3.0-items zijn opgenomen. Daardoor is voldoende geborgd dat de LVS 3.0-schaal met bijbehorende normering ook van toepassing is op deze nieuwe Leerling in beeld-toetsen. Deze LVS 3.0-items zijn in de normeringsonderzoeken afgenomen samen met nieuwe Lib-items en met Lay-out-3.0-items en worden aangeduid als ankeritems die de nieuwe Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 verbinden met de bestaande LVS 3.0-toetsen Begrijpend lezen voor groep 3 t/m 5. De link tussen de toetsen was minimaal 20 ankeritems, waarmee dus wordt voldaan aan de eis op pagina 5 uit het (op de homepage van de Expertgroep Toetsen PO te vinden) document 'Beoordelingskader voor (reeksen van) toetsen uit leerlingvolgsystemen (LVS)' dat het anker uit minstens 15 items moet bestaan. Op grond van het bovenstaande kan worden geconcludeerd dat er sprake is van een onvolledig doch 'verbonden' dataverzamelingsdesign, omdat er overlap bestaat tussen de nieuwe Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 en de bestaande LVS 3.0-toetsen Begrijpend lezen voor groep 3 t/m 5 door middel van ankeritems die de toetsen aan elkaar linken (i.e., ankerdesign).

*Conclusie:*

Het onvolledige maar 'verbonden' dataverzamelingsdesign is adequaat. Het afnamedesign werd zo ingericht dat de nieuw geconstrueerde Lib-items voor Begrijpend lezen groep 3 t/m 5 bij de kalibratie konden worden gekoppeld aan de items van de bestaande LVS 3.0 toetsen voor Begrijpend lezen groep 3 t/m 5. Op aspect S2 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 derhalve het oordeel '**voldoende**' toegekend.

Daarbij wordt er wel van uitgegaan dat het voornemen om in het schooljaar 2022-2023 aanvullende analyses op de Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 uit te voeren om na te gaan of itemparameters voor het schooljaar daarop moeten worden bijgesteld voor items die toch wat anders blijken te functioneren, ook daadwerkelijk wordt gerealiseerd.

S3 In het geval van een observatie-instrument: is er sprake van een adequate steekproef van observatoren en randvoorwaarden waaronder de observatie wordt uitgevoerd?

*Bevindingen:*

Dit criterium is niet van toepassing (n.v.t.), omdat er hier sprake is van Lib-volgtoetsen Begrijpend lezen groep 3, 4 en 5 en er dus geen sprake is van een observatie-instrument.

*Conclusie:*

**n.v.t.**

S4 Er is een handleiding met duidelijke instructies voor de leerkracht over het zo objectief mogelijk uitvoeren en weergeven van de observaties door de leerkracht.

*Bevindingen:*

Dit criterium is niet van toepassing (n.v.t.), omdat er hier sprake is van Lib-volgtoetsen Begrijpend lezen groep 3, 4 en 5 en er dus geen sprake is van een observatie-instrument.

*Conclusie:*

**n.v.t.**

## **Normering**

N1.1 Is de standaardbepalingsmethode gemotiveerd en op de juiste wijze uitgevoerd?

*Bevindingen:*

Dit criterium heeft betrekking op absoluut normeren en is dus niet van toepassing (n.v.t.) op de Lib-toets Begrijpend lezen groep 3, 4 en 5, omdat deze toets betrekking heeft op relatief normeren.

*Conclusie:*

**n.v.t.**

N1.2 Zijn de beoordelaars/vakdeskundigen/experts naar behoren geselecteerd en getraind?

*Bevindingen:*

Dit criterium heeft betrekking op absoluut normeren en is dus niet van toepassing (n.v.t.) op de Lib-toets Begrijpend lezen groep 3, 4 en 5, omdat deze toets betrekking heeft op relatief normeren.

*Conclusie:*

**n.v.t.**

N1.3 Is er voldoende overeenstemming tussen de beoordelaars?

*Bevindingen:*

Dit criterium heeft betrekking op absoluut normeren en is dus niet van toepassing (n.v.t.) op de Lib-toets Begrijpend lezen groep 3, 4 en 5, omdat deze toets betrekking heeft op relatief normeren.

*Conclusie:*

**n.v.t.**

N2.1 Zijn de normgroepen groot genoeg?

*Bevindingen:*

In eerdere psychometrische beoordelingen van LVS 3.0 toetsen Begrijpend lezen voor groep 3 t/m 5 is geconstateerd dat de normgroepen groot genoeg zijn (zie website Expertgroep Toetsen PO onder 'Kwaliteitsoordelen van LVS-instrumenten'). Omdat de nieuwe Lib-items tegelijk zijn afgenomen met reguliere 3.0-afnames tijdens de normeringsonderzoeken in de periode januari 2019 t/m juni 2022, mogen we er daarom van uitgaan dat de normgroepen voor de Lib-toetsen Begrijpend lezen voor groep 3 t/m 5 ook groot genoeg zijn.

*Conclusie:*

De normgroepen zijn groot genoeg. Op aspect N2.1 wordt aan de Lib-toetsen Begrijpend lezen groep 3, 4 en 5 het oordeel '**voldoende**' toegekend.

N2.2 Zijn de normgroepen representatief?

*Bevindingen:*

De representativiteit van de steekproeven werd hierboven onder aspect S1 besproken en daar werd geconstateerd dat de steekproeven representatief waren voor regio, urbanisatiegraad, schooltype en sekse.

*Conclusie:*

De normgroepen zijn representatief. Op aspect N2.2 wordt aan de Lib-toetsen Begrijpend lezen groep 3, 4 en 5 het oordeel '**voldoende**' toegekend.

### N2.3 Zijn de normen correct bepaald?

#### *Bevindingen:*

Omdat de items waaruit de Lib-toetsen Begrijpend lezen groep 3 t/m 5 zijn opgebouwd op dezelfde vaardigheidsschaal liggen als de 3.0-toetsen Begrijpend lezen groep 3 t/m 5, is ervoor gekozen om de huidige normering van de 3.0-toetsen toe te passen op de nieuwe Lib-toetsen. Bij de uitrol van LVS 3.0 zijn alle toetsen van het leerlingvolgsysteem op de relevante onderdelen genormeerd, welke in eerste instantie is gebaseerd op de normeringsonderzoeken voor LVS 3.0 in 2014 en vervolgens bijgesteld op basis van gegevens uit dataretour. Door 3.0-items te hergebruiken in de Lib-toetsen, is er een direct anker met de oorspronkelijke 3.0-schaal en is er dus een doorgaande lijn van de 3.0-toetsen naar de Lib-toetsen. Op deze manier kan de bestaande normering van 3.0-toetsen Begrijpend lezen groep 3 t/m 5 worden overgebracht op de nieuwe Lib-toetsen Begrijpend lezen groep 3 t/m 5. De bestaande normering van 3.0-toetsen Begrijpend lezen groep 3 t/m 5 maakt gebruik van relatieve normen op basis van percentielscores en wordt geïnterpreteerd op basis van de vaardigheidsverdeling in een adequate, landelijke, referentiegroep, waarbij de vaardigheidsscore wordt uitgedrukt in de symmetrische niveau-indeling I t/m V en in de asymmetrische niveau-indeling A t/m E.

De Lib-toetsen worden zorgvuldig gemonitord om te controleren of de vertaling van de behaalde (ruwe) score naar vaardigheidsscore en de normering op de vaardigheidsscores nog passend zijn. Om die reden zijn er twee keer per jaar controles ingebouwd om de kwaliteit van de normering te garanderen, welke bestaan uit het monitoren van vaardigheidsscores in relatie tot het gebruik van relatieve normeringen en het vergelijken van karakteristieken van items uit de normeringsonderzoeken en uit reguliere afnames van de Lib-toetsen.

Nieuwe items worden getest in de toetsen van het volgsysteem Leerling in beeld voor Begrijpend lezen groep 3 t/m 5 door de methode van zaaien en oogsten, waarbij nieuwe nog niet eerder geteste items (verversingsitems) worden opgenomen in de reguliere Lib-toetsen. Of de itemkarakteristieken van deze nieuwe items in lijn liggen met de itemkarakteristieken van items gemeten in het normeringsonderzoek tijdens de coronapandemie, zal moeten worden uitgewezen door analyses. Door het tegenvallende aantal observaties in de normeringsonderzoeken en de opgelopen achterstanden door de schoolsluitingen in die periode, moet rekening worden gehouden met de mogelijkheid dat de karakteristieken van de nieuwe items afwijken en dat als gevolg daarvan de normering van de Lib-toetsen in schooljaar 2022-2023 moet worden bijgesteld (evenals de itemparameters wellicht moeten worden bijgesteld op basis van aanvullende analyses in schooljaar 2022-2023, omdat ze toch wat anders blijken te functioneren).

#### *Conclusie:*

In eerdere psychometrische beoordelingen van LVS 3.0 toetsen Begrijpend lezen groep 3 t/m 5 is geconstateerd dat de normen correct zijn bepaald (zie website Expertgroep Toetsen PO onder 'Kwaliteitsoordelen van LVS-instrumenten') en omdat de huidige normering van de 3.0-toetsen wordt toegepast op de nieuwe Lib-toetsen, mogen we ervan uitgaan dat de normen op de Lib-toetsen Begrijpend lezen groep 3 t/m 5 ook correct zijn bepaald. Op aspect N2.3 wordt aan de Lib-toetsen Begrijpend lezen groep 3, 4 en 5 het oordeel '**voldoende**' toegekend.

Evenals bij aspect S2, wordt er daarbij wel weer van uitgegaan dat het voornemen om in het schooljaar 2022-2023 aanvullende analyses uit te voeren om na te gaan of de

normering wellicht moet worden bijgesteld vanwege het tegenvallend aantal observaties ten tijde van de coronapandemie, ook daadwerkelijk wordt gerealiseerd.

## **Betrouwbaarheid**

### B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

#### *Bevindingen:*

In eerdere psychometrische beoordelingen van LVS 3.0 toetsen Begrijpend lezen voor groep 3 t/m 5 is beschreven hoe, gebruikmakend van het feit dat de items uit de LVS 3.0 toetsen geschaald zijn met behulp van de OPLM software, een (globale) betrouwbaarheidscoëfficiënt, de MAcc ('Accuracy of Measurement'), berekend kan worden die qua interpretatie grote overeenkomst vertoont met de betrouwbaarheidscoëfficiënt uit de klassieke testtheorie (KTT) (zie website Expertgroep Toetsen PO onder 'Kwaliteitsoordelen van LVS-instrumenten'). Omdat voor deze berekeningen gebruik is gemaakt van bekende software, kunnen we ervan uitgaan dat de betrouwbaarheidsgegevens correct zijn berekend. Omdat de items waaruit de Lib-toetsen Begrijpend lezen groep 3 t/m 5 zijn opgebouwd op dezelfde vaardigheidsschaal liggen als de LVS 3.0-toetsen Begrijpend lezen groep 3 t/m 5, mogen we ervan uitgaan dat de betrouwbaarheidsgegevens voor de Lib-toetsen Begrijpend lezen groep 3 t/m 5 eveneens correct zijn berekend.

#### *Conclusie:*

Op aspect B1 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

### B2 Zijn de betrouwbaarheidsgegevens voldoende gezien de conclusies en eventuele beslissingen die met het instrument genomen worden?

#### *Bevindingen:*

Uit de Excel-tabel met psychometrische gegevens voor de Lib-toetsen Begrijpend lezen groep 3 t/m 5 kan worden afgelezen dat de MAcc-betrouwbaarheidscoëfficiënt voor de papieren versie voor de afnamemomenten E3 (eind 3), M4 (midden 4), E4 (eind 4), M5 (midden 5) en E5 (eind 5) gelijk zijn aan respectievelijk 0.89, 0.86, 0.86, 0.86 en 0.84. Bij de adaptieve afname volgens het Multistage Testing model (MST model) worden de volgende drie verschillende paden van de MST onderscheiden: het iets makkelijkere pad, het gemiddelde pad en het iets moeilijker pad. De MAcc-betrouwbaarheidscoëfficiënt voor de adaptieve afname is in de Excel-tabel met psychometrische gegevens vervolgens geschat over de drie paden gezamenlijk (MST\_totaal) en is voor de afnamemomenten E3, M4, E4, M5 en E5 gelijk aan respectievelijk 0.87, 0.86, 0.83, 0.85 en 0.83. Volgens het COTAN-beoordelingssysteem (Evers et al., 2010) kunnen tests voor minder belangrijke beslissingen op individueel niveau, bijvoorbeeld leerlingvolgtoetsen, met een betrouwbaarheid van 0.8 of hoger als 'goed' worden geclassificeerd. De hierboven genoemde betrouwbaarheden zitten daar allemaal ruim boven en de globale betrouwbaarheid van de Lib-toetsen Begrijpend lezen groep 3 t/m 5 is dus goed te noemen.

*Conclusie:*

De (globale) betrouwbaarheidsgegevens zijn voldoende gezien de conclusies en eventuele beslissingen die met de toetsen worden genomen. Op aspect B2 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

**Validiteit**

V1 Inhoudsvaliditeit: Dragen de items in het instrument bij aan de validiteit van het instrument (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items)?

*Bevindingen:*

Tussen de toetsen van groep 3 tot en met groep 5 neemt de complexiteit van de tekst toe. De teksten in toets E3 bestaan nog uit korte zinnen van ongeveer 6 woorden en bestaan vooral uit klankzuivere woorden. In de toetsen van de groepen 4 en 5 neemt de lengte van de zinnen toe en komen er bijvoorbeeld meer samengestelde woorden in voor, niet klankzuivere woorden, meer personages etc. Hier zit een heldere opbouw in complexiteit tussen de toetsen.

De teksten zijn afwisselend in alle toetsen. Zowel informatieve teksten (verslag, brief/kaart, uitleg, advertentie etc.) als fictieve teksten (o.a. ook gedichten/liedjes) komen aan de orde. De teksten zijn vlot geschreven en sluiten aan bij de beleavingswereld van kinderen. De illustraties zijn passend bij de geschreven teksten en illustreren de kern waar de tekst over gaat. De teksten zijn zoveel mogelijk authentiek weergegeven.

De vragen bij de teksten zijn meerkeuzevragen en een enkele ordeningsvraag (zet de plaatjes in goede volgorde). De antwoordcategorieën zijn goed onderscheidend van elkaar. De vragen lopen op in de toetsen van op een meer beschrijvend niveau de vragen te stellen dus specifieke informatie herkennen en het begrijpen van zinnen naar een meer structurerend niveau. Dit zijn vragen naar het leggen van verbanden.

In de teksten wordt vooral begrijpen en interpreteren getoetst wat een goede keuze is gezien het belang van deze kenmerken voor het lezen. Evalueren en samenvatten, andere kenmerken binnen de leesvaardigheid in het Referentiekader voor Taal, zullen bij een toenemende leesvaardigheid steeds meer onderdeel uitmaken van de toetsen. In de toetsen van groep 5 zie je dat ook nadrukkelijk terug.

*Conclusie:*

**'voldoende'**.

V2 Constructvaliditeit: Meet het instrument in zijn geheel datgene wat het beoogt te meten?

*Bevindingen:*

Onder de veronderstelling dat het aannemelijk is dat de Lib-toets Begrijpend lezen voor groep 3 t/m 5 in zijn geheel past onder het gehanteerde IRT-model OPLM (hetgeen overigens in de toekomst nog wel statistisch zou moeten worden getoetst met de globale R1c-toetsen; zie ook de opmerking onder aspect S2) mag er worden uitgegaan van



unidimensionaliteit voor het gehanteerde IRT-model OPLM als meetmodel. Met behulp van confirmatieve factoranalyse zou eventueel nog nader onderzocht kunnen worden of er inderdaad sprake is van een één-factor model. Deze beide bevindingen tesamen zouden dan impliceren dat het redelijk is om ervan uit te gaan dat de Lib-items één latente trek of construct meten, welke we kunnen aanduiden als begrijpende leesvaardigheid. Hiermee is voldaan aan de noodzakelijke (maar niet voldoende) voorwaarde voor (psychometrische) constructvaliditeit.

Evenals voor de bestaande LVS 3.0-toetsen Begrijpend lezen groep 3 t/m 5, zou ook voor de Lib-toetsen Begrijpend lezen groep 3 t/m 5 nog onderzoek gedaan kunnen worden naar de volgende aspecten als aanvullende onderbouwing van psychometrische constructvaliditeit: soortgenootvaliditeit in de vorm van convergente validiteit (i.e., toetsen die hetzelfde construct meten moeten hoog correleren) en divergente validiteit (i.e., toetsen die verschillende constructen meten moeten juist laag correleren), intercorrelationele verbanden tussen de verschillende inhoudelijke subvaardigheden bij Begrijpend lezen groep 3 t/m 5, itembias, verschillen tussen relevante subgroepen (bijvoorbeeld op basis van geslacht en leerlinggewicht) en itemkwaliteit. We kunnen er namelijk niet zonder meer van uitgaan dat op al deze aspecten empirische ondersteuning zal worden gevonden voor de constructvaliditeit van de Lib-toetsen Begrijpend lezen groep 3 t/m 5. Alleen het aspect itemkwaliteit kan hier op basis van de psychometrische gegevens uit de Excel-tabel in de vorm van de beide itemparameters moeilijkheden (p-waarden) en item-testcorrelaties (rit-waarden) uit de klassieke testtheorie (KTT) nader onderzocht worden als aanvullende onderbouwing van (psychometrische) constructvaliditeit voor de Lib-toetsen Begrijpend lezen groep 3 t/m 5.

Uit de Excel-tabel met psychometrische gegevens valt uit de p-waarden per item af te leiden dat voor de afnamemomenten E3 (N = 57), M4 (N = 57), E4 (N = 62), M5 (N = 54) en E5 (N = 66) de gemiddelde p-waarden gelijk zijn aan respectievelijk 0.68, 0.69, 0.68, 0.66 en 0.66 en de ranges van de p-waarden zijn gelijk aan respectievelijk (0.38-0.91), (0.41-0.90), (0.23-0.92), (0.29-0.93) en (0.36-0.93). De gemiddelde moeilijkheid ligt tussen 0.66 en 0.69. Er wordt in het algemeen gestreefd naar een gemiddelde p-waarde tussen de 0.60 en 0.75, waardoor de toetsen niet te makkelijk en niet te moeilijk zijn en wordt voorkomen dat de leerling gefrustreerd raakt tijdens de toetsafname. De gemiddelde moeilijkheidsgraad ligt dus op het vooraf gewenste niveau. Uit de ranges blijkt dat de p-waarden tussen 0.23 en 0.93 liggen, hetgeen betekent dat er gezorgd is voor een goede spreiding van moeilijkheid over de items. In enkele gevallen is een item met een (extreem) hoge p-waarde (i.e., erg makkelijk item) en (extreem) lage p-waarde (i.e., erg moeilijk item) opgenomen, maar dit zijn echter zonder uitzondering items met een 'voldoende tot goede' rit-waarde (i.e, discrimineerden 'goed tot voldoende' tussen hoogscorende en laagscorende leerlingen).

Verder valt uit de Excel-tabel met psychometrische gegevens uit de rit-waarden per item af te leiden dat de gemiddelde rit-waarden voor de verschillende afnamemomenten gelijk zijn aan (waarbij rit\_pap, rit\_mak, rit\_gem en rit\_moe staan voor respectievelijk de rit-waarde van de papieren versie, de rit-waarde van het iets makkelijkere pad, de rit-waarde van het gemiddelde pad en de rit-waarde van het iets moeilijker pad bij de adaptieve afname volgens het MST):

E3 (N = 37):

rit\_pap = 0.46; rit\_mak = 0.47; rit\_gem = 0.48; rit\_moe = 0.46.

M4 (N = 36):

rit\_pap = 0.41; rit\_mak = 0.41; rit\_gem = 0.43; rit\_moe = 0.42.

E4 (N = 37):

rit\_pap = 0.41; rit\_mak = 0.40; rit\_gem = 0.41; rit\_moe = 0.40.

M5 (N = 36):

rit\_pap = 0.41; rit\_mak = 0.43; rit\_gem = 0.42; rit\_moe = 0.41.

E5 (N = 37):

rit\_pap = 0.39; rit\_mak = 0.39; rit\_gem = 0.40; rit\_moe = 0.39.

Ook valt uit de Excel-tabel met psychometrische gegevens uit de rit-waarden per item af te leiden dat de ranges voor deze rit-waarden voor de verschillende afnamemomenten gelijk zijn aan:

E3 (N = 37):

rit\_pap = (0.29-0.58); rit\_mak = (0.33-0.58);  
rit\_gem = (0.33-0.58); rit\_moe = (0.29-0.57).

M4 (N = 36):

rit\_pap = (0.27-0.57); rit\_mak = (0.27-0.58);  
rit\_gem = (0.27-0.57); rit\_moe = (0.27-0.55).

E4 (N = 37):

rit\_pap = (0.29-0.50); rit\_mak = (0.29-0.51);  
rit\_gem = (0.28-0.53); rit\_moe = (0.28-0.52).

M5 (N = 36):

rit\_pap = (0.24-0.52); rit\_mak = (0.25-0.52);  
rit\_gem = (0.26-0.52); rit\_moe = (0.26-0.52).

E5 (N = 37):

rit\_pap = (0.27-0.50); rit\_mak = (0.26-0.51);  
rit\_gem = (0.28-0.50); rit\_moe = (0.27-0.50).

Door de COTAN wordt een rit-waarde van boven de 0.30 gekwalificeerd als goed en boven de 0.40 als uitstekend. Uit de gemiddelde rit-waarden valt dus af te lezen dat de gemiddelde rit-waarde bij alle vijf afnamemomenten goed tot uitstekend is. Verder wordt er naar gestreefd om alleen items in de toetsen op te nemen met een rit-waarde van 0.20 of hoger, de door de COTAN genoemde ondergrens voor de beoordeling 'voldoende'. Uit de ranges van de rit-waarden blijkt dat bij alle afnamemomenten er geen enkel item is met een rit van lager dan 0.20.

#### *Conclusie:*

Uit de Excel-tabel met psychometrische gegevens valt alleen iets te zeggen over het aspect itemkwaliteit als aanvullende ondersteuning voor de psychometrische constructvaliditeit. Uit de gerapporteerde resultaten van de klassieke itemparameters in deze Excel-tabel valt af te leiden dat de kwaliteit van de items hoog is: zowel de p-waarden als de rit-waarden zijn goed tot uitstekend te noemen. Wat betreft het aspect itemkwaliteit kan dus geconcludeerd worden dat dit aspect aanvullende ondersteuning geeft voor psychometrische constructvaliditeit.

Andere aspecten (soortgenootvaliditeit in de vorm van convergente en divergente validiteit, intercorrelationele verbanden tussen de verschillende inhoudelijke subvaardigheden bij Begrijpend lezen groep 3 t/m 5, itembias en verschillen tussen relevante subgroepen) als aanvullende ondersteuning voor psychometrische constructvaliditeit, dienen echter in de toekomst nog wel nader onderzocht te worden. Ook zal in de toekomst nog aannemelijk gemaakt moeten worden dat de Lib-toets Begrijpend lezen voor groep 3 t/m 5 in zijn geheel past onder het gehanteerde IRT-model OPLM (zie ook de opmerking onder aspect S2) en er dus mag er worden uitgegaan van unidimensionaliteit voor het gehanteerde IRT-model OPLM als noodzakelijke (doch niet voldoende) voorwaarde voor psychometrische constructvaliditeit.

Het instrument in zijn geheel meet datgene wat het beoogt te meten. Op aspect V2 wordt aan de Lib-toetsen Begrijpend lezen 3 t/m 5 derhalve het oordeel '**voldoende**' toegekend, waarbij er wel van uitgegaan wordt dat in de toekomst nader onderzoek wordt uitgevoerd naar het hierboven genoemde in de Conclusie.

### **Het volg-aspect**

Va1 Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een correcte manier gemeten?

#### *Bevindingen:*

Door de verankering met de bestaande LVS 3.0-toetsen Begrijpend lezen groep 3 t/m 5, kunnen de items van de Lib-toetsen Begrijpend lezen groep 3 t/m 5 op een eendimensionale vaardigheidsschaal afgebeeld worden. Aan de hand van de door de leerling behaalde vaardigheidsscores op de onderscheiden toetsen kan diens groei correct gemeten worden. In de verschillende leerjaren wordt met verschillende toetsen de groei gemeten die met behulp van IRT op één en dezelfde vaardigheidsschaal zijn gebracht.

#### *Conclusie:*

Er is voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt en de groei wordt op een correcte manier gemeten. Op aspect Va1 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

Va2 Wordt de betrouwbaarheid van de groei op die schaal correct weergegeven?

#### *Bevindingen:*

In de Handleiding van bestaande LVS 3.0-toetsen Begrijpend lezen groep 3 t/m 5, met name in hoofdstuk 4, wordt aan de hand van een leerlingrapport de interpretatie van groei op een duidelijke manier beschreven. De 67%-betrouwbaarheidsintervallen van de vaardigheidsscores worden zowel op het leerlingrapport als in afzonderlijke tabellen vermeld. We mogen ervan uitgaan dat dit voor de nieuwe Lib-toetsen Begrijpend lezen groep 3 t/m 5 ook gebeurt.

#### *Conclusie:*

De betrouwbaarheid van de groei op die schaal wordt correct weergegeven. Op aspect Va2 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

Va3 Worden er voldoende gegevens verstrekt (aan de gebruiker) over hoe groei geïnterpreteerd dient te worden?

*Bevindingen:*

Met behulp van de relatieve normeringen kunnen de Lib-toetsen Begrijpend lezen groep 3 t/m 5 ingezet worden om de ontwikkeling van leerlingen te volgen en te evalueren 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. In de Handleiding is meer informatie te vinden over de wijze waarop de gebruiker dit kan doen. Geconcludeerd kan worden dat het volgen en evalueren van groei van leerlingen adequaat wordt toegelicht. Voor alle vergelijkingen geldt dat uitspraken over de voortgang van leerlingen gerelativeerd moeten worden vanwege de (on)betrouwbaarheid van de toetsen, hetgeen ook wordt vermeld op het leerlingrapport.

*Conclusie:*

Er worden voldoende gegevens verstrekt (aan de gebruiker) hoe groei geïnterpreteerd dient te worden. Op aspect Va3 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

### **Inzicht in leervorderingen**

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

*Bevindingen:*

Voor bestaande LVS 3.0-toetsen Begrijpend lezen groep 3 t/m 5 kan via de portal van Cito B.V. gebruik worden gemaakt van rapportageformulieren 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) en we mogen ervan uitgaan dat dit ook geldt voor de nieuwe Lib-toetsen Begrijpend lezen groep 3 t/m 5.

In de Handleiding wordt in hoofdstuk 7 aandacht besteed aan hoe met ouders over de toetsresultaten gecommuniceerd kan/moet worden. In het bijzonder wordt daarbij gewezen op het leerlingrapport waarin zowel het niveau van de leerling als de progressie van de leerling numeriek en grafisch gepresenteerd worden. Hiervoor is ook een folder ouderinformatie beschikbaar die men via de website van het Cito kan downloaden. Daarnaast wordt de docent 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.

In hoofdstuk 8 worden ook veel gestelde vragen behandeld die weliswaar voor de docenten bestemd zijn, maar zij kunnen met die informatie bijvoorbeeld via tien-minutengesprekken ook de ouders beter voorlichten.

*Conclusie:*

De aanbieder levert een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/verzorgers/voogden/docenten begrijpelijk is. Op aspect I1 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

I2 Is er een evaluatie van de leervorderingen en worden op basis van deze evaluatie vervolgstappen geformuleerd?

*Bevindingen:*

Het volgen van leerlingen in hun groei, ook wel aangeduid als progressiebepaling, is een van de belangrijkste functies van het Cito Volgsysteem primair en speciaal onderwijs (LVS). De toetsen van het LVS (en dus ook de nieuwe Lib-toetsen Begrijpend lezen groep 3 t/m 5) geven de leerkracht (en ouders en leerlingen zelf) informatie over de ontwikkeling van de vaardigheden van de leerlingen, zowel 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?

Er zijn zowel rapportages op leerling-, groeps- als schoolniveau. In de rapportagematerialen zijn twee niveau-indelingen opgenomen, waarmee de leerkracht de scores van een leerling kan vergelijken met die van een grote groep leerlingen. De leerkracht kan een keuze maken uit een indeling in de niveaus, I tot en met V en A tot en met E. Daarnaast heeft de leerkracht de mogelijkheid om functioneringsniveaus op te vragen. De functioneringsniveaus geven aan met welke gemiddeld scorende leerling de vaardigheidsscore van de getoetste leerling vergelijkbaar is. Een functioneringsniveau M5 betekent bijvoorbeeld dat de vaardigheidsscore van de leerling overeenkomt met de score van de gemiddeld scorende leerling medio groep 5.

Op basis van deze belangrijke informatie uit de evaluatie van de leervorderingen wat betreft lacunes in de ontwikkeling van de leerling in het onderwijs kunnen vervolgstappen worden geformuleerd door het onderwijs beter op het kind af te stemmen. In de handleiding worden hiervoor in hoofdstuk 4 'Interpreteren en analyseren op leerling- en groepsniveau' tips gegeven in de vorm van het interpreteren en analyseren van de rapportages, aanvullende signalering bij problemen en hoe de leerkracht de gegevens kan gebruiken voor de onderbouwing van het handelen.

*Conclusie:*

Er is een evaluatie van de leervorderingen en op basis van deze evaluatie worden vervolgstappen geformuleerd. Op aspect I2 wordt aan de Lib-toetsen Begrijpend lezen groep 3 t/m 5 het oordeel '**voldoende**' toegekend.

**Referentieniveaus**

R1 Sluit de inhoud van de toets aan op de kennis en vaardigheden zoals omschreven in de referentieniveaus van het betreffende domein (voor toetsen vanaf groep 6)?

*Bevindingen:*

n.v.t.

*Conclusie:*

**n.v.t.**

### 3. Verzamelstaat

<b>Kwaliteitsaspect</b>	<b>Code</b>	<b>Oordeel</b>
De kwaliteit van de steekproef	<i>S1</i>	<b>Voldoende</b>
	<i>S2</i>	<b>Voldoende</b>
	<i>S3</i>	<b>n.v.t.</b>
	<i>S4</i>	<b>n.v.t.</b>
Normering	<i>N1.1</i>	<b>n.v.t.</b>
	<i>N1.2</i>	<b>n.v.t.</b>
	<i>N1.3</i>	<b>n.v.t.</b>
	<i>N2.1</i>	<b>Voldoende</b>
	<i>N2.2</i>	<b>Voldoende</b>
	<i>N2.3</i>	<b>Voldoende</b>
Betrouwbaarheid	<i>B1</i>	<b>Voldoende</b>
	<i>B2</i>	<b>Voldoende</b>
Validiteit	<i>V1</i>	<b>voldoende</b>
	<i>V2</i>	<b>Voldoende</b>
Volg-aspect	<i>Va1</i>	<b>Voldoende</b>
	<i>Va2</i>	<b>Voldoende</b>
	<i>Va3</i>	<b>Voldoende</b>
Inzicht in leervorderingen	<i>I1</i>	<b>Voldoende</b>
	<i>I2</i>	<b>Voldoende</b>
Referentieniveaus	<i>R1</i>	<b>n.v.t.</b>

### 4. Literatuurlijst

- Alberts, N., Engelen, R., Groenen, I., Kamphuis, F. (2021). Wetenschappelijke verantwoording Woordenschat 3.0, voor groep 3 tot en met 8. Cito.