

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de Gebruikershandleiding en Verantwoording (hierna: G&V) veelal letterlijk vermeld. Deze heeft enerzijds betrekking op de uitgangspunten van de toetsconstructie, de normen, de betrouwbaarheid en meetnauwkeurigheid en de validiteit, en anderzijds op het gebruik van de toets, communicatie over de toetsgegevens en de inhoudsverantwoording.

Algemeen

De Boom LVS-toets Rekenen-Wiskunde (2019) is een schoolvaardigheidstoets voor het primair onderwijs en maakt deel uit van een reeks LVS-toetsen. De toets is bedoeld om het niveau van de rekenvaardigheid van de leerling vast te stellen.

De toets is opgebouwd uit zes blokken, waarbij de opgaven van ieder blok de leerstof weerspiegelen van een leerjaar van het basisonderwijs (vanaf groep 3), zoals die in de reguliere reken-wiskunde-methoden in het basisonderwijs wordt onderwezen.

Meetpretentie

De toets meet de vaardigheid in Rekenen-Wiskunde van leerlingen, zoals deze wordt onderwezen in de meest gebruikte lesmethoden in het Nederlandse basisonderwijs. Daarbij gaat het met name om de nauwkeurigheid en het inzicht, die nodig zijn voor het oplossen van rekenopgaven, en niet zozeer om het tempo van het rekenen.

Doelgroep

De toets is bestemd voor groep 3 tot en met groep 8 van het Nederlandse basisonderwijs en het speciaal basisonderwijs.

Gebruiksdoel en functie

De toets kan op de volgende manieren worden ingezet:

- Niveaubepaling van de leerlingen, zowel individueel als per groep.
- Voortgangsbepaling van de leerlingen. Wanneer de toets minstens jaarlijks wordt afgenomen, kan deze worden ingezet t.b.v. het LVS.
- Signalering van de 'zwakke plekken' in de rekenvaardigheid van de leerlingen.
- In het kader van opbrengstgericht werken, leent de toets zich ook voor de evaluatie van oefenprogramma's voor het rekenen.

Inhoudelijke theoretische inkadering:

De toets sluit aan bij de kerndoelen van het primair onderwijs zoals die in 2006 door het Ministerie van OCW zijn vastgesteld. In totaal zijn er 58 kerndoelen, waarvan 11 kerndoelen vallen in het vak Rekenen-Wiskunde. Deze 11 kerndoelen zijn verdeeld over de categorieën 'wiskundig inzicht en handelen', 'getallen en bewerkingen' en 'meten en meetkunde'. Voor de concrete uitwerking van de kerndoelen is gebruik gemaakt van de inhoud van het referentiekader taal en rekenen (Meijerink, 2008; NoteBoom, Van Os en Spek, 2011) en de Toetswijzer bij de Centrale Eindtoets PO (CvTE, 2014).

Inhoud van het toetspakket

Het toetspakket bestaat uit:

- Gebruikershandleiding en verantwoording

Beoordeling van Boom LVS-toets Rekenen-Wiskunde

- Instructiekaart voor de leerkracht
- Toetsboekje A
- Toetsboekje B
- Antwoordenboekje A
- Antwoordenboekje B
- Informatieblad voor ouders/verzorgers
- Voorbeeld van een individueel rapport
- Voorbeeld van een individueel overzicht

2. Beoordeling van de kwaliteitsaspecten

De beoordeling vindt plaats volgens het 'Beoordelingskader voor (reeksen van) toetsen uit 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), Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en mevrouw Jennifer Roubiës MSc (secretaris).

De kwaliteit van de dataverzameling

S1 Is de steekproef representatief?

Bevindingen:

De normering van de Boom LVS-toets Rekenen-Wiskunde vond plaats op twee momenten. Dit betekent dat de normering is gebaseerd op twee steekproeven: steekproef I en steekproef II. De eerste normeringsronde vond plaats in november 2018 en de tweede normeringsronde in april 2019. De voor de normering geselecteerde scholen namen in beide normeringsrondes met alle groepen, dat wil zeggen groep 3 tot en met 8, deel aan het onderzoek.

Het toetsboekje van de Boom LVS-toets Rekenen-Wiskunde versie A of versie B bestaat uit 6 blokken. Elk blok bestaat uit 12 tot 21 opgaven. In de regel bestaat iedere opgave uit 4 subopgaven, en in enkele gevallen uit 8 subopgaven. Blok 1 en 2, respectievelijk bedoeld voor groep 3 en 4, bevatten uitsluitend opgaven uit het domein 'Getallen', terwijl blok 6, bedoeld voor groep 8, opgaven bevat uit 8 van de 14 onderscheiden domeinen. Op het eerste normeringsmoment maakten de leerlingen van iedere school één of meerdere blokken van versie A. Op het tweede normeringsmoment maakten de leerlingen beide versies.

In tabel 6.2 en 6.6 van de Gebruikershandleiding en Verantwoording staan de aantallen scholen, klassen en leerlingen van de normeringssteekproef. Steekproef I: 18 scholen, 120 klassen en 2193 leerlingen die versie A gemaakt hebben. Steekproef II: 21 scholen, 138 klassen en 2449 leerlingen die versie A en B gemaakt hebben.

De representativiteit van de steekproef is onderzocht op schoolniveau (schoolgrootte, statusscore, regionale spreiding en stedelijkheid) en op leerlingniveau (sekse, leeftijd, migratieachtergrond en leerlinggewicht). Om de representativiteit van de steekproef te beoordelen, wordt in tabel 6.7 en 6.8 van de G&V de effectmaat Cramers V opgenomen. Cramers V is een associatiemaat tussen variabelen op nominaal niveau. Voor de interpretatie van Cramers V worden tevens de richtlijnen van Cohen (1988) vermeld. In het geval dat de representativiteit wordt beschreven door middel van een variabele op intervalniveau, zoals de statusscore, wordt Cohen's (d) gerapporteerd om de effectgrootte te beoordelen.

In hoofdstuk 6 van de G&V wordt de steekproef nauwkeurig beschreven op zowel schoolniveau als leerlingniveau. De verdeling van de relevante variabelen laat zich, eventueel na weging, in het algemeen goed vergelijken met de verdeling van deze variabelen in de populatie.

Er is tevens onderzoek gedaan naar schoolafhankelijkheid. In tabel 6.27 van de G&V staan de intraklassecorrelatiecoëfficiënten (ICC) vermeld. Van schoolafhankelijkheid blijkt sprake te zijn bij de normering van blok 3 en 4 maar dit effect is niet groot.

Conclusie:

Gezien de representativiteit van de steekproef ten aanzien van relevante variabelen op school- en leerlingniveau wordt aan aspect S1 van de Boom LVS-toets Rekenen-Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

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

Bevindingen:

In tabel 6.1 van de G&V staat het afnameschema van de normeringssteekproef. Te zien is dat bij steekproef I van versie A bij groep 3 blok 1 wordt afgenomen, bij groep 4 blok 1 & 2, bij groep 5 blok 2 & 3, bij groep 6 blok 3 & 4, bij groep 7 blok 4 & 5 en de referentieset 1F en bij groep 8 blok 5 & 6 en de referentiesets 1S en 2F. Te zien is dat bij steekproef II bij groep 3 blok 1 van versie A en versie B wordt afgenomen, bij groep 4 blok 1 & 2 van versie A en blok 2 van versie B, bij groep 5 blok 2 & 3 van versie A en blok 3 van versie B, bij groep 6 blok 3 & 4 van versie A en blok 4 van versie B, bij groep 7 blok 4 & 5 van versie A en blok 5 van versie B en bij groep 8 blok 5 & 6 van versie A en blok 6 van versie B.

Het meetmodel dat voor de kalibratie gebruikt is, is het partial credit model (PCM), een variant van het Rasch-model voor opgaven die polytoom gescoord worden. Bij de onderhavige LVS-toets konden op bepaalde opgaven 4 of 8 scorepunten behaald worden. Een belangrijke aanname van het PCM-model is die van de eendimensionaliteit. Om deze aanname te onderzoeken is voor de gegevens van de normeringssteekproef voor ieder blok een principale componenten analyse (PCA) uitgevoerd. Uit de tabellen in bijlage IV van de G&V blijkt dat er steeds sprake is van één duidelijke factor en dat de tweede factor van aanzienlijk minder gewicht is.

Voor de schatting van de parameters van het gehanteerde model dient er zowel verticale linking (tussen de blokken voor de verschillende leerjaren) als horizontale linking (tussen versie A en versie B van de LVS-toets) plaats te vinden. De linking is gerealiseerd door parameters in de schattingsprocedure te fixeren, een methode die in de literatuur bekend staat als 'concurrent calibration with fixed parameters'. Om verticale linking mogelijk te maken, hebben de leerlingen in de steekproef I en II twee blokken van versie A gemaakt: blok 1 wordt gemaakt door groep 3 en 4, blok 2 door groep 4 en 5, blok 3 door groep 5 en 6, blok 4 door groep 6 en 7, blok 5 door groep 7 en 8 en blok 6 door groep 8. Hierdoor is gegarandeerd dat over een groot aantal ankeropgaven wordt beschikt. Om horizontale linking mogelijk te maken, hebben de leerlingen in steekproef II zowel een blok van versie A als een blok van versie B gemaakt: groep 3 maakt blok 1, groep 4 blok 2, groep 5 blok 3, groep 6 blok 4, groep 7 blok 5 en groep 8 blok 6. Op basis van de geschatte parameters is vervolgens voor ieder blok een correspondentietabel met ruwe scores en vaardigheidsscores opgesteld. Om negatieve scores te vermijden, is een transformatie van +2 op de vaardigheidsscores toegepast. Om de passing van het PCM-model te beoordelen, is een PCA uitgevoerd op de inter-itemcorrelaties van de residuele waarden. De redenering hierachter is dat als op basis van een PCA een structuur onderkend kan worden, dit een aanwijzing is voor een inadequate passing van het model. De veronderstelling is dus dat er geen duidelijke factor in de data van de residuele waarden te vinden is. De PCA's in bijlage IV van de G&V bevestigen de veronderstelling.

Een andere manier om de passing van het model te beoordelen, is door gebruik te maken van zogenaamde Infit- en Outfit-statistieken die beide gebaseerd zijn op het

kwadraat van het gestandaardiseerde residu tussen de geobserveerde en de op het model gebaseerde verwachte waarden. Infit is minder gevoelig voor extreme itemresponsen dan Outfit. In de literatuur zijn er verschillende opvattingen over de criteria waaraan Infit en Outfit moeten voldoen. Het meest zorgwekkend zijn Infit en Outfit maten > 2.00 en < 0.50 . In bijlage V van de G&V worden de waarden van voornoemde itemstatistieken gepresenteerd. In termen van Infit- en Outfit-waarden zijn er maar enkele slecht passende opgaven.

Volgens de hiervoor beschreven procedure zijn zowel voor versie A als voor versie B de parameters van het PCM-model geschat. Dit betekent dat de volgens het model berekende vaardigheidsscores op één schaal staan en dat het in de praktijk niet uitmaakt welke versie men bij een leerling afneemt om zijn of haar ontwikkeling te volgen. De parallelliteit van de twee versies werd bevestigd door gelijke informatiefuncties, nagenoeg gelijke parameterschattingen en nagenoeg perfecte positieve correlaties tussen ruwe scores en vaardigheidsscores.

Conclusie:

Het incomplete dataverzamelingsdesign is adequaat en ook de kalibratie is adequaat uitgevoerd. Aan aspect S2 wordt aan de Boom LVS-toets Rekenen - Wiskunde, groep 3 - groep 8, het oordeel '**voldoende**' toegekend.

Normering

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

Bevindingen:

Dit aspect heeft betrekking op absoluut normeren en is dus niet van toepassing.

Conclusie:

n.v.t.

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

Bevindingen:

Dit aspect heeft betrekking op absoluut normeren en is dus niet van toepassing.

Conclusie:

n.v.t.

N1.3 Is er voldoende overeenstemming tussen de beoordelaars?

Dit aspect heeft betrekking op absoluut normeren en is dus niet van toepassing.

Conclusie:

n.v.t.

N2.1 Zijn de normgroepen groot genoeg?

Bevindingen:

In tabel 62 en 6.6 van de G&V staan de aantallen leerlingen vermeld – zie beschrijving aspect S1. Voor elk van de zes onderscheiden groepen/blokken zijn de toetsresultaten van 330 tot 450 leerlingen gebruikt.

Conclusie:

Aan aspect N2.1 wordt aan de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

N2.2 Zijn de normgroepen representatief?

Bevindingen:

Bij aspect S1 is de representativiteit van de steekproef ten aanzien van relevante variabelen op school- en leerlingniveau besproken en voldoende bevonden.

Conclusie:

Aan aspect N2.2 wordt aan de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen

In tabel 7.1 en tabel 7.2 van de G&V zien we dat alle blokken een meer dan voldoende betrouwbaarheid, Cronbach's alfa, hebben voor zowel versie A (.769 -.926) als versie B (.798 - .913).

Tabel 7.3 van de G&V bevat de parallele betrouwbaarheden, correlaties tussen versie A en versie B, van de onderscheiden groepen/blokken. Met uitzondering van groep/blok 3 met een betrouwbaarheid van .620, zijn de betrouwbaarheden hoog.

De methode van Rudner (2001, 2005) is gebruikt om de accuratesse te bepalen. Deze methode gaat ervan uit dat voor de berekening van de accuratesse de cesuren zijn bepaald voor een op IRT gebaseerde vaardigheidsschaal wat hier het geval is. In bijlage XI van de G&V wordt de accuratesse van de niveau-indeling voor de verschillende toetsvormen en de verschillende normeringsmomenten gegeven. De accuratesse-waarden staan ook in tabel 7.4 van de G&V. Bij de niveau-indeling I – V zijn twee accuratesse-waarden < .50 en bij de niveau-indeling A – E is één waarde < .50.

Conclusie:

De (parallele) betrouwbaarheden en accuratesse-waarden zijn correct berekend. Aan aspect B1 wordt aan de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

B2 Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

Bevindingen:

Bij aspect B1 zijn de betrouwbaarheden vermeld en die betrouwbaarheden voldoen aan de standaarden die door de COTAN aan deze toetsen gesteld worden. Bij aspect B1 zijn ook de accuratesse-waarden vermeld en deze accuratesse-waarden zijn bevredigend te noemen.

Conclusie:

Aan aspect B2 wordt aan de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

Validiteit

V1 Inhoudsvaliditeit: 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:

De opbouw in de blokken (in elk leerjaar een nieuw blok) is goed verantwoord. Bij 'schatten' maakt men gebruik van meerkeuze, al het andere zijn open vragen. Dit past bij een efficiënte afwikkeling van de toets, het zou wellicht goed zijn dat de leerkracht zich wel bewust is van deze beperking in opgave-vormen. De leerling zal rekenen ervaren als iets waar maar 1 antwoord op mogelijk is, en zal ervaren dat niet gevraagd wordt om aan te geven hoe de oplossing tot stand is gekomen. In het vervolgonderwijs (o.a. bij wiskunde) wordt duidelijk gemaakt dat er juist voor de oplossingsmanier (de weg richting het antwoord) ook waardering kan zijn, en aandacht moet zijn. Het zou goed zijn dit in de beschrijving van deze LVS op te nemen, het zou anders tot een te beperkt beeld kunnen leiden wat rekenen/wiskunde is.

Beoordelaar heeft waardering voor de vormgeving. Sober maar helder. Er is in veel gevallen gezocht of grafisch materiaal ondersteunend kan zijn bij de opgave, en dit is in essentie goed uitgevoerd.

Blok 6 zal voor de zwakke rekenaar niet eenvoudig zijn (A niet en B niet). Dit brengt het lastige met zich mee dat de angst voor rekenen juist bij de rekenzwakken nog aangewakkerd wordt (de leerling zal onvoldoende scoren, het zelfvertrouwen neemt af). Ook hier graag de hand van de leerkracht die er op gewezen kan worden dat deze bijwerkingen plaatsvinden als er wordt gewerkt met LVS.

Beoordelaar ziet hier, in aansluiting op deze observatie, mogelijkheden voor het formatief gebruik van toetsopgaven uit deze toets. Gebruik voorbeelden (letterlijk, of varianten) van opgaven uit deze toets om deze gezamenlijk te bespreken in de klas, laat leerlingen in tweetallen deze opgaven maken, laat leerlingen vertellen hoe ze tot de oplossing zijn gekomen. Op deze manier maak je een brug tussen het reguliere onderwijs en de 'toets-setting' (een uur, individueel, alleen een kladblaadje, e.d.).

Misschien nog iets over de rekenmachine. Ook hier geldt dat er een brug geslagen kan worden tussen de toetspraktijk van deze toets en wat er verder in het onderwijs gebeurt. Ik weet overigens niet zeker of dit echt in het basisonderwijs zou moeten (groepen 7 en 8), maar ik zie wel een abrupte overgang tussen po en vo, en het zou wellicht goed zijn ook hier een vorm te vinden dat bijv. opgaven uit de toets ook eens met de rekenmachine worden uitgerekend. Het zal namelijk opvallen dat een rekenmachine vaak geen handige hulp biedt bij een dergelijke opgave (of er moet eerst 'vertaald' worden van de wijze waarop de vraag is gesteld naar de vaststelling welke bewerking moet worden ingetypt op

de rekenmachine). Juist deze didactische mogelijkheid is wellicht iets om een enkele keer toe te voegen (een les na de toets, e.d.).

Beoordelaar probeert te zoeken naar manieren waarop de toets meer ingebed raakt in het onderwijs, en niet te sterk is losgezongen (en dan ook waarde verliest).

3. Afname, scoring en normering

Dit is goed en helder beschreven.

Het is goed dat er gekeken is naar het verschil met de Cito LVS (had ook een andere mogen zijn). Hoewel vergelijken altijd lastig is, levert het informatie over hoe toetsaanbieder er zelf voor staat (met eigen toets en interpretatie van rekenen/wiskunde). Gebruik een dergelijke afweging ook voor de komende jaren, pas dan kan helder worden over wat gemeten wenst te worden, en hoe je als toetsaanbieder e.e.a. operationaliseert. Zoals bij de inleiding al aangegeven meet de toets vrij smal 'rekenen/wiskunde'. Het zou dus goed zijn het validiteitsvraagstuk aan te grijpen om hierover een meer weloverwogen keuze te maken (en daarna, na die keuze, helder te zijn naar de gebruiker wat die keuze is).

V2 Constructvaliditeit: Meet de toets in zijn geheel datgene wat hij beoogt te meten?

Bevindingen:

In hoofdstuk 8 van de G&V worden drie argumenten gepresenteerd die volgens de toets-constructeurs pleiten voor de begripsvaliditeit van de onderhavige LVS-toets.

Het eerste argument betreft de ontwikkeling van de gemiddelde rekenvaardigheid in het basisonderwijs zoals deze op basis van de normeringsgegevens kan worden vastgesteld.

In grafiek 8.1 van de G&V is de gemiddelde ontwikkeling van de rekenvaardigheid, gemeten met de rekenen-wiskundetoets, uitgezet gedurende het basisonderwijs. In de grafiek is naast de gemiddelde ontwikkelingslijn van Boom LVS-toets Rekenen-Wiskunde ter vergelijking ook de gemiddelde ontwikkelingslijn van de Cito-toets Rekenen-Wiskunde 3.0 opgenomen. Doel van deze vergelijking is na te gaan of de gemiddelde groei over de leerjaren van het basisonderwijs bij deze twee toetsen hetzelfde patroon laat zien. De grafiek geeft aan dat de patronen een opvallende gelijkenis vertonen.

Het tweede argument betreft groepsverschillen. Er is gekeken naar verschillen voor sekse, achtergrond van de leerlingen en naar leerlinggewicht. Jongens blijken iets beter te zijn in Rekenen-Wiskunde dan meisjes, iets wat bevestigd wordt door internationaal onderzoek. Leerlingen met een Nederlandse achtergrond doen het iets beter dan leerlingen met een niet-westerse migratieachtergrond terwijl de verschillen met leerlingen met een westerse migratieachtergrond niet eenduidig zijn. Er is een duidelijk verschil tussen leerlingen met en leerlingen zonder een gewicht.

Het derde argument betreft de cesuren van de referentieniveaus. Om te kunnen bepalen of de leerlingen het gevraagde referentieniveau beheersen, is een cesuur voor de Boom LVS-toets Rekenen-Wiskunde vastgesteld met behulp van referentiesets. Tabel 8.5 van de G&V laat zien dat de referentiesets voldoende betrouwbaar zijn. Tabel 8.6 van de G&V laat zien dat de correlaties tussen de referentiesets en de relevante blokken aan de verwachting voldoen, namelijk > 0.70 . Bij het vaststellen van de cesuren blijken de cesuren voor het 1S-niveau en het 2F-niveau nagenoeg identiek. Dit betekent dat volgens de Boom LVS-toets Rekenen-Wiskunde leerlingen die het 1S-niveau halen ook het 2F-niveau halen. Op basis van de vastgestelde cesuren en de normeringsgegevens

kunnen de verwachte percentages leerlingen die het 1F-niveau en het 1S/2F-niveau halen, worden vastgesteld. Deze percentages, respectievelijk 87% en 40%, sluiten aan bij de verwachting.

Conclusie:

Op basis van de gepresenteerde argumenten wordt aan aspect V2 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, het oordeel '**voldoende**' toegekend.

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 adequate manier gemeten?

Bevindingen:

Uit de resultaten van het in hoofdstuk 5 van de G&V gepresenteerde kalibratieonderzoek blijkt dat de opgaven van de parallelle versies A en B van de Boom LVS-toets Rekenen-Wiskunde op een ééndimensionele vaardigheidsschaal afgebeeld kunnen worden waardoor zowel het niveau als de groei van de rekenvaardigheid adequaat vastgesteld kunnen worden. Grafiek 8.1 van de G&V laat de (lineaire) ontwikkeling zien van de gemiddelde rekenvaardigheid in het basisonderwijs.

Voor het antwoord op de vraag of groei van individuele leerlingen op een adequate manier gemeten wordt, wordt verwezen naar de vier casusbeschrijvingen in paragraaf 4.7 van het G&V. De casusbeschrijvingen laten voorbeelden van individuele rapporten en individuele overzichten zien van de groei van leerlingen.

Conclusie:

Aan aspect Va1 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, wordt het oordeel '**voldoende**' toegekend.

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

Bevindingen:

De ruwe toetsscores zijn op een vaardigheidsschaal geplaatst waarmee de groei van de leerlingen gevolgd kan worden. Om te bepalen of er sprake is van werkelijke groei worden 90% betrouwbaarheidsintervallen gebruikt.

Conclusie:

Aan aspect Va2 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, wordt het oordeel '**voldoende**' toegekend.

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

Bevindingen:

In de vier casusbeschrijvingen in paragraaf 4.7 van G&V wordt adequaat toegelicht hoe groei geïnterpreteerd dient te worden. Bijlage 3 van de G&V bevat de verwachte vaardigheidsgroei na een bepaald aantal onderwijsmaanden.

Conclusie:

Aan aspect Va3 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, 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:

Dit blijft, voor een gemiddelde lezer, een lastig hoofdstuk, en dat is jammer, want hier liggen de kansen voor de leerkracht voor 'data-driven teaching' (evidence-informed). Het zou goed zijn hier nog eens met enkelen over door te praten wat hier nu exact nodig is voor een helder beeld. De casusbeschrijvingen zijn daarbij handig, maar vragen nog steeds om meer toelichting.

Op pagina 37 wordt aangegeven dat scores bijgehouden kunnen worden in het scoreboekje. Dit boekje is niet bij de materialen aanwezig. De zin 'door de ruwe goedscore te vertalen naar de normscore krijgt u een beeld waar de betreffende leerling wat betreft rekenvaardigheid staat in vergelijking met landelijk niveau' is onduidelijk. Hoe vindt die vertaling plaats? Op de instructiekaart wordt over deze vertaling niet gesproken, maar wordt alleen verwezen naar de productpagina van Boom. Dit is voor leerkrachten te beperkt.

Door Boom is een adequate reactie op bovenstaande bevindingen gegeven.

Conclusie:

Aan aspect I1 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, wordt 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:

Toets sluit aan op de kennis en vaardigheden zoals omschreven in de referentieniveaus van het betreffende domein.

Conclusie:

Aan aspect R1 van de Boom LVS-toets Rekenen – Wiskunde, groep 3 – groep 8, wordt het oordeel **'voldoende'** toegekend.

3. Verzamelstaat

Kwaliteitsaspect	Code	Oordeel
------------------	------	---------

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

4. Literatuurlijst

- Drenth, P.J.D., & Sijtsma, K. (2006), *Testtheorie*. Bohn, Stafleu, van Loghum.
- Sijtsma, K. (2009) On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120.
- Van Vugt, J., de Vos, T., Milikowski, M. & Milikowski, R. (2019). *Rekenen-Wiskunde. Gebruikershandleiding en Verantwoording*. Amsterdam: Boom Uitgevers.