

## **1. Algemene informatie**

### Algemeen en meetpretentie

Het IEP LVS is een methodeonafhankelijk volgsysteem waarin leerlingen vanaf leerjaar 3 tot aan de eindtoets gevolgd kunnen worden in hun ontwikkeling. Het IEP-LVS is een leer- en criteriumgericht volgsysteem. De toetsen Rekenen maken onderdeel uit van dit volgsysteem. De ter beoordeling voorliggende toetsen zijn de toetsen voor leerjaar 6, 7 en 8: versie 2 van de toetsen <1F-1F en <1F-1F-1S (<1F-1F\_v2 en <1F-1F-1S\_v2), die in 2019 ter beoordeling bij de Expertgroep ingediend zijn.

De toetsen Rekenen meten de rekenvaardigheid van leerlingen. De toetsen worden digitaal gemaakt. Het is mogelijk om de teksten uit de toets als boekje te downloaden en te printen voor de leerling. Op die manier kan de leerling de teksten op papier lezen.

### Doelgroep

Het IEP LVS bevat toetsen voor leerlingen in leerjaar 3 tot en met 8. De ter beoordeling voorliggende toetsen zijn bedoeld voor leerlingen in leerjaar 6, 7 en 8. Het IEP LVS is echter zo ingericht dat alle toetsen voor elke leerling toegankelijk zijn. Zo kan een leerling in leerjaar 5 die moeite heeft met rekenen, ook de toets op het niveau van leerjaar 4 maken. Maar het is bijvoorbeeld ook mogelijk om een leerling in leerjaar 6 die relatief goed is in rekenen al een toets op 2F-niveau te geven. Het voorgaande maakt dat het IEP LVS ook geschikt is voor leerlingen uit het SBO.

### Inhoudelijke theoretische inkadering:

Het Referentiekader taal en rekenen (Meijerink et al., 2009) ligt ten grondslag aan de rekentoetsen van leerjaar 6 tot en met 8. In dit referentiekader is beschreven welke kennis en vaardigheden van leerlingen worden verwacht. Binnen het rekenonderwijs wordt onderscheid gemaakt tussen drie fundamentele niveaus (1F, 2F en 3F) en drie streefniveaus (1S, 2S en 3S). Verder heeft het Referentiekader een cumulatief karakter. Zo beheerst een leerling op niveau 2F ook de inhoud van niveau 1F. Uitzondering hierop zijn de streefniveaus (S-niveaus), omdat die een abstracte en meer wiskundige rekenvaardigheid van leerlingen vragen. Zo omvat niveau 1S wel de inhoud van niveau 1F, maar omvat niveau 2F niet altijd de abstracte inhoud van niveau 1S. De leerling hoeft daardoor op niveau 2F niet de gehele inhoud van niveau 1S te beheersen.

In het IEP LVS zijn ook vragen opgenomen op het niveau 'op weg naar 1F' (<1F). Dit zijn vragen die eenvoudiger zijn dan het niveau 1F. Dit niveau is niet bij de wet vastgesteld. Vragen met het niveau 'op weg naar 1F' zijn in het IEP LVS opgenomen zodat gekeken kan worden in hoeverre een leerling op weg is naar 1F.

### Inhoud van het toetspakket:

Het toetspakket Rekenen voor leerjaar 6, 7 en 8 bestaat uit de volgende documenten:

- Verantwoording LVS-toetsen Rekenen <1F-1F versie 2 en <1F-1F-1S versie 2, deze bevat informatie over:
  - o De uitgangspunten van de toetsen (hfdst. 2),
  - o De inhoud van de toetsen (hfdst. 3),
  - o De steekproef (hfdst. 4),
  - o Het design van de dataverzameling (hfdst. 5),
  - o De kalibratie en kwaliteit van de items (hfdst. 6),

- IJking van de referentieniveaus (hfdst. 7),
- De constructvaliditeit (hfdst. 8),
- Toetswijzer (bijlage 1)
- Itemoverzicht (bijlage 2)
- Itemparameters (bijlage 3)
- TIA's (bijlage 4)
- Algemene toelichting methode (bijlage 5)
- Omzettingstabel (bijlage 6)
- Langrange Multiplier Tracelines (bijlage 7)
- Toelichtingen (bijlage 8)

## **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), 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:*

Uitgangspunt is een steekproef die resulteert in tenminste 400 observaties per item. Hoewel voor de kalibratie van de toetsen met het hier gehanteerde 1PL-model (Rasch model) het minimale aantal observaties 200 is, is voor het evalueren van de modelpassing van een redelijk alternatief model, hier het 2PL-model, minimaal 400 observaties de norm. Uit bijlage 3 (zie kolom 'Aantal afnames') blijkt dat het aantal observaties van de IEP LVS items ruimschoots voldoet aan de eisen die worden beschreven in het document 'Aanvulling COTAN Beoordelingssysteem' m.b.t. het aspect normering referentieniveaus (d.d. 16-06-2016).

De representativiteit op achtergrondgegevens van de normeringspopulatie (groep leerlingen die aan het normeringsonderzoek heeft deelgenomen) ten opzichte van de doelpopulatie (alle leerlingen in de bovenbouw van het regulier basisonderwijs) is naast aantallen afnames uiteraard ook van belang om een oordeel te kunnen toekennen aan dit aspect. Omdat voor de normering van de referentieniveaus (volgens het Referentiekader Taal en Rekenen van de commissie Meijerink, 2009) een extern criterium wordt gebruikt, het anker 2019 van de Eindtoets dat door de Expertgroep Toetsen PO is vastgesteld, wordt door de auteurs van de Wetenschappelijke Verantwoording (WV) gesteld dat het niet noodzakelijk is dat de normeringspopulatie ook een normpopulatie is en dientengevolge de representativiteitseisen aan een normpopulatie hier dus niet van toepassing zijn. Het hoofddoel van onderhavig onderzoek is volgens de auteurs namelijk het vaststellen van absolute cesuren op de referentieniveaus, waarvoor de samenstelling van de normeringspopulatie volgens de auteurs van ondergeschikt belang is. Anderzijds stellen de auteurs echter ook vast dat het voor de bepaling van de kwaliteit van de items wel van belang is om na te gaan dat de normeringspopulatie geen specifiek selecte groep leerlingen is, maar een reguliere doorsnede van de bovenbouw van het basisonderwijs.

Voor bovengenoemd doel worden de volgende schoolachtergrondgegevens gebruikt: denominatie (openbaar en bijzonder), urbanisatiegraad (G4, G5-37 en >G37), schoolgrootte (<100, 100-300 en >300), regio (Noord, Oost, West en Zuid) en schoolgewicht (percentage gewichtenleerlingen, t.w. <23%, 23%-26.99%, 27%-32.99%, 33%-36.99% en  $\geq 37\%$ ), welke per school bij DUO openbaar beschikbaar zijn. Er worden geen persoonlijke achtergrondgegevens gebruikt omwille van de privacy. Het digitale platform IEP LVS is gebruikt om de data voor het normeringsonderzoek te verzamelen. De betrokken scholen hebben de toetsen <1F-1F\_v2 (aangeraden om in het tweede deel van leerjaar 6 af te nemen) en <1F-1F-1S\_v2 (aangeraden om in leerjaar 7 af te nemen) op eigen initiatief tijdens het reguliere onderwijsproces in schooljaar 2019-2020 afgenomen.

Er is hier dus sprake van 'purposeful sampling' (een doelsteekproef), een niet-probabilistische steekproeftechniek. Door deze werkwijze is gegarandeerd dat de data in het normeringsonderzoek is verzameld onder gelijke afnamecondities en afnamemomenten als waaronder de IEP LVS-toetsen ook in de komende jaren afgenomen gaan worden. Alleen data van leerlingen die minstens 85% van de vragen beantwoord hebben is meegenomen in het onderzoek.

Hoe de normeringspopulatie zich op achtergrondgegevens verhoudt tot de doelpopulatie na de selectie van de data op bruikbaarheid is weergegeven in tabel 4.1. Deze tabel laat zien dat niet alle categorieën van de achtergrondgegevens van de normeringspopulatie als een representatieve doelsteekproef van de bovenbouw van het reguliere basisonderwijs beschouwd kunnen worden. Zoals reeds eerder gememoreerd is de samenstelling van de normeringspopulatie volgens de auteurs echter van ondergeschikt belang voor het onderhavige onderzoek (dus niet noodzakelijkerwijs ook een normpopulatie), maar moet wel een reguliere doorsnede van de leerlingen in de bovenbouw van het basisonderwijs zijn en dus geen specifiek selecte groep leerlingen. Dit laatste blijkt uit het feit dat alle achtergrondcategorieën zijn vertegenwoordigd in de normeringspopulatie en er blijkens tabel 4.2 weinig grote effectgroottes  $\phi$  ( $\phi = \sqrt{\text{Chi-kwadraat}/N}$ ) zijn. Met grote effectgroottes worden waarden van  $\phi \geq 0.50$  bedoeld (Cohen, 1988). De auteurs concluderen derhalve dat representativiteit, in de betekenis dat op alle steekproeven die in de schaling gebruikt zijn een-en-hetzelfde IRT-model van toepassing is, aannemelijk is (zie paragraaf 7.2) en daarmee deze normeringspopulatie voor dit onderzoek bruikbaar is.

#### *Conclusie:*

De procedure voor het samenstellen van de steekproeven is onderbouwd en de omstandigheden en momenten waaronder data is verzameld, is vergelijkbaar met de afnamecondities en afnamemomenten waaronder de toetsen worden afgenomen. Hoewel de normeringspopulatie niet op alle achtergrondcategorieën als representatieve doelsteekproef beschouwd kan worden, is deze echter wel bruikbaar voor het doel waarmee de data verzameld zijn (i.e., geen relatieve maar absolute normering van de referentieniveaus). Op aspect S1 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend. De bovenstaande opmerkingen zijn suggesties die de toetsverantwoording duidelijker zouden maken.

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

##### *Bevindingen:*

In paragraaf 5.1 ('Normeringsonderzoek voor de IEP LVS-toetsen') wordt beschreven welk design is gebruikt om de data te verzamelen ten behoeve van het vaststellen van de huidige cesuren op de referentieniveaus en hoe dit design gebruikt zal worden om de cesuren in de toekomst up to date te houden. Uitgangspunt bij het vaststellen van de cesuren voor de referentieniveaus zijn de cesuren op het gezamenlijk anker van de Eindtoetsen 2019 geweest, een extern criterium, welke door de Expertgroep Toetsen PO zijn bepaald na de afname van de eindtoetsen in 2019. Eerder is in de 'Verantwoording IEP LVS 2019' beschreven hoe de cesuren van het gezamenlijk anker op de eindtoetsen 2019 middels twee schalingen zijn overgebracht naar de IEP LVS toetsen versie 1 (figuur 5.1), waarbij gebruik werd gemaakt van de applicatie Lexter. De reeds bestaande linking tussen het gezamenlijk anker Eindtoetsen 2019 en de IEP Eindtoetsen 2017, 2018 en 2019

is hierbij gebruikt in de eerste schaling. Voor de linking tussen de IEP Eindtoetsen 2017, 2018 en 2019 en het IEP LVS versie 1 is hierbij in de tweede schaling gebruikgemaakt van een ankerdesign.

De tweede schaling uit figuur 5.1 is nu opnieuw uitgevoerd teneinde de cesuren van het gezamenlijk anker van de eindtoetsen 2019 ook over te zetten op de IEP LVS toetsen versie 2, waarbij in deze nieuwe schaling nu ook de IEP LVS toetsen versie 2 zijn meegenomen. Hierbij zijn de itemparameters van de ankers tussen de IEP Eindtoetsen en de IEP LVS toetsen, die bij de eerste schaling vorig jaar geschat waren, gefixeerd bij de tweede schaling (i.e., IEP LVS toetsen). Alle items van de IEP Eindtoetsen en de IEP LVS toetsen versie 1 en de nieuwe IEP LVS toetsen versie 2 zijn op deze manier op dezelfde parameterschaal gezet. In het ankerdesign van tabel 5.1 wordt de linking tussen de IEP LVS toetsen versie 2 en de reeds goedgekeurde IEP LVS toetsen versie 1 visueel weergegeven. De link tussen de toetsen was minimaal 15 ankeritems, of als meetellend item of als niet-meetellend item, waarmee dus wordt voldaan aan de eis op pag. 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.

Belangrijk is om na te gaan of de geselecteerde ankeritems uit reeds goedgekeurde IEP LVS toetsen een goede afspiegeling zijn van de IEP LVS-toetsen (en dus representatieve ankeritems zijn), zowel inhoudelijk als psychometrisch. De auteurs concluderen dat de ankeritems inhoudelijk een goede afspiegeling zijn, omdat alle referentieniveaus en inhoudelijke toetsdomeinen van Rekenen (Getallen (30%), Verhoudingen (30%), Meten en meetkunde (20%) en Verbanden (20%)) vertegenwoordigd zijn (zie tabel 5.1 en bijlage 2). De auteurs concluderen dat de ankeritems psychometrisch ook representatief zijn, omdat er sprake is van een goede spreiding van de b-parameters (moeilijkheidsparameters in het Rasch-model als gehanteerd IRT-model) van de ankeritems (zie document 'Itemparameters IEP LVS Rekenen' in bijlage 3).

De normering van de IEP LVS toetsen zal ieder schooljaar blijvend worden gecontroleerd aan het meest recent afgenomen gezamenlijk anker van de eindtoetsen, omdat het niet vaststaat dat de referentiecesuren op de latente schaal onveranderd blijven de komende jaren. De vastgestelde cesuren (zoals beschreven in de hoofdstukken 6 en 7 van de WV) gelden daarom tot en met het einde van het schooljaar 2020-2021. Elk jaar wordt het anker voor een deel ververs met minimaal 15 representatieve ankeritems per referentieniveau uit de laatste IEP Eindtoets.

Het Rasch-model (1PLM) met de Marginal Maximum Likelihood (MML) schattingsmethode is toegepast voor de IRT-analyse (d.w.z. kalibratie van alle itemparameters op dezelfde schaal) van de IEP LVS-toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 zijn uitgevoerd in Versie 0.0.7 van de applicatie Lexter (2019), waarmee door de Expertgroep Toetsen PO ook de analyses gedaan zijn voor de vaststelling van de referentiecesuren van de IEP Eindtoets 2019. Zodoende zijn de resultaten goed vergelijkbaar.

In bijlage 3 ('Itemparameters en DIF IEP LVS Rekenen') zijn het aantal afnames, de moeilijkheid ( $\beta$ ), de meetfout van de moeilijkheid ( $SE(\beta)$ ) en de iteminformatie (i.e., Fisher-informatie) bij de gemiddelde vaardigheid van de afnamegroepen voor de IEP Eindtoets 2019 en de IEP LVS toetsen versie 2 in een tabel weergegeven, waarbij de vaardigheidsschaal is genormeerd door het gemiddelde en de standaarddeviatie van de afnamegroep van de IEP Eindtoets 2019 op respectievelijk 0 en 1 te zetten.

In paragraaf 7.2 wordt een methode besproken om de nauwkeurigheid van de parameterschattingen te beoordelen (Evers et al., 2010). Deze methode bestaat eruit om de nauwkeurigheid van de parameterschattingen na te gaan aan de hand van de constante 'c', die de relatie weergeeft tussen de standaardfout van de moeilijkheidsparameter van een item ( $SE(\beta)$ ) en de standaarddeviatie van de vaardigheidsverdeling van de kalibratiepopulatie ( $\sigma_\theta$ ). Volgens het COTAN-beoordelingssysteem (Evers et al., 2010) worden waarden van c lager of gelijk aan 0.2 als 'goed' beoordeeld en waarden tussen 0.3 en 0.4 als 'voldoende'. De c-waarden van de items per kalibratie (zoals reeds eerder besproken zijn er twee kalibraties uitgevoerd, namelijk de kalibratie van de IEP eindtoetsen en de IEP LVS toetsen) zijn berekend om de nauwkeurigheid van de parameterschattingen te kunnen beoordelen. Tabel 7.2 laat zien dat de gemiddelde waarde van de constante c (i.e., 0.026 en 0.050), berekend over alle items in beide kalibraties, veel lager is dan de vereiste waarde van 0.2 en geen item heeft een c-waarde van boven de 0.2 (maximum-waarde voor de constante c is 0.099). Op basis van deze resultaten kan de nauwkeurigheid van de parameterschattingen als goed beoordeeld worden.

Modelpassing is onderzocht via een Differential Item Functioning (DIF) analyse (uitgevoerd in de applicatie Lexter) op de 20 ankeritems tussen de versie 1 en 2 toetsen (zie tabel 5.1) en daarnaast is via statistische toetsing nagegaan of de Item Response Curven (IRC's) de responsies goed representeren (eveneens uitgevoerd in Lexter). In bijlage 3 is per item aangegeven in welke boekjes het item was opgenomen, de Lagrange Multiplier statistiek (LM), het aantal vrijheidsgraden (df), de overschrijdingskans (Prob) en het absolute verschil (Abs. Diff). Met name de absolute verschillen zijn informatief, omdat significantie altijd gevoelig is voor de steekproefgrootte (bij grote steekproeven is een chi-kwadraat toets bijna altijd significant). Voor de 10 ankers van de <1F-1F-1S\_v2 en de eerdere IEP LVS toetsen is het gemiddelde absolute verschil 0.03 (max = 0.08; min = 0.01). Voor de 10 ankers van de <1F-1F\_v2 en de eerdere IEP LVS toetsen is het gemiddelde absolute verschil eveneens 0.03 (max = 0.05; min = 0.01). Op basis van deze resultaten kan geconcludeerd worden dat de observaties op de items – inclusief de ankeritems tussen de nieuwe en oude toetsen – van de IEP LVS toetsen en de verwachting weinig van elkaar verschillen, hetgeen (ongeacht het boekje) een goede modelpassing aannemelijk maakt.

Door middel van de First order Statistics optie ('Lagrange multiplier tracelines for Rasch-Type Model') uit de applicatie Lexter is de mate waarin de Item Response Curven de responsies goed representeren statistisch getoetst. Uit bijlage 7 is af te lezen dat van de 840 berekende effectgroottes er 86 groter zijn dan 0.10. De auteurs vinden het echter toch verantwoord om deze items in de IEP LVS toetsen te behouden in het hier gehanteerde Rasch-model, omdat het om een relatief klein percentage gaat dat een (te) grote effectgrootte (i.e., groter dan 0.10) laat zien en deze items volgens de KTT (Klassieke Testtheorie) bovendien goed functioneren (zie bijlage 4).

In paragraaf 6.3 ('Uitgangspunten selectie items IEP LVS-toetsen Rekenen') worden de volgende uitgangspunten besproken voor de definitieve selectie van items voor de IEP LVS-toetsen Rekenen versie 2 gebaseerd ten eerste op het niveau van de items en ten tweede op de statistieken afkomstig uit de TIA-analyse (zie bijlage 4 voor de TIA-statistieken van het normeringsonderzoek):

- (i) Items van het referentieniveau <1F en 1F hebben een p-waarde (proportie correct) onder de 0.95 en items van het referentieniveau 1S hebben een p-waarde onder de 0.90;

- (ii) Items hebben een p-waarde boven de 0.30 (open items) of een p-waarde boven de gokkans (meerkeuze items). Items met een p-waarde onder 0.30 (open items) of een p-waarde onder de gokkans (meerkeuze items) worden alleen geselecteerd als het gaat om open items met een goede rit-waarde;
- (iii) Items hebben een rit-waarde (item-totaalcorrelatie) van groter of gelijk aan 0.20, zodat de items als voldoende beoordeeld kunnen worden volgens het COTAN-beoordelingssysteem (Evers et al., 2010).

Voor de toetsen <1F-1F-1S\_v2 en <1F-1F\_v2 is het gelukt om bovenstaande uitgangspunten te hanteren én daarbij een goede inhoudelijke dekking van het Referentiekader taal en rekenen (Meijerink et al., 2009) en variatie in domeinen te garanderen (i.e., inhoudsvaliditeit).

#### Opm.:

Pag. 18, laatste twee alinea's: De auteurs stellen dat ze het toch verantwoord vinden om een aantal items die niet in het Rasch-model blijken te passen vanwege berekende effectgroottes groter dan 0.10 toch te behouden. Er worden hier twee redenen voor gegeven, namelijk ten eerste omdat de betreffende items in de toetsen volgens de KTT goed functioneren (zie bijlage 4) en ten tweede gaat het hier volgens de auteurs om een relatief klein percentage items dat een (te) grote effectgrootte laat zien. Van de 840 berekende effectgroottes zijn er echter 86 groter dan 0.10, hetgeen neerkomt op 10.24%. Het is discutabel of hier dan nog wel gesproken kan worden over een relatief klein percentage.

De auteurs stellen verder dat de keuze voor het Rasch-model voortkomt uit het feit dat de IEP Eindtoets, inclusief het gezamenlijk anker, ook gebruik maakt van het Rasch-model. Omdat de IEP Eindtoets, inclusief het gezamenlijk anker, gebruikt wordt voor de normering van het IEP LVS (zie fig. 5.1) is er gekozen hier ook het Rasch-model te hanteren. De Expertgroep kan meegaan in deze redenering en het ligt daarom dus niet voor de hand om na te gaan of het 2PL-model wellicht een betere modelpassing geeft dan het Rasch-model (1PL-model). Wel is dan echter de vraag gerechtvaardigd of (bij handhaving van het Rasch-model) een aantal items met (erg) grote effectgroottes (die dus blijkbaar niet goed in het geassumeerde Rasch-model passen) niet beter verwijderd kunnen worden, zodat het percentage items dat een (te) grote effectgrootte laat zien substantieel kleiner wordt dan 10.24%. Een andere indicatie dat bepaalde items blijkbaar niet goed passen in het Rasch-model is om na te gaan welke items een erg afwijkende rit-waarde (i.e., discriminerend vermogen) volgens de KTT hebben (zie bijlage 4). Dit zou er namelijk op kunnen duiden dat deze items wellicht beter in een IRT-model met ongelijke discriminatieparameters (i.e., 2PL-model) zouden passen. Bij het eventueel verwijderen van items op basis van psychometrische gronden (i.e., (te) grote effectgroottes en/of afwijkende rit-waarden) is het uiteraard goed om ook inhoudelijke overwegingen in ogenschouw te nemen door na te gaan of de resterende items tezamen nog wel een goede afspiegeling zijn van de IEP LVS toetsen (i.e., inhoudsvaliditeit).

De toetsaanbieders hebben toegezegd dit advies in de toekomst ter harte te nemen.

#### *Conclusie:*

Het onvolledige maar 'verbonden' dataverzamelingsdesign is adequaat. De normen worden jaarlijks geëvalueerd en eventueel aangepast. Op aspect S2 wordt aan de toetsen IEP LVS

Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het volgende oordeel toegekend:  
**'voldoende.'**

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 toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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 toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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:*

In paragraaf 5.1 'Normeringsonderzoek voor de IEP LVS toetsen' wordt het proces beschreven hoe de cesuren voor de IEP LVS toetsen versie 2 via een gelinked design middels twee schalingen zijn afgeleid van het gezamenlijk anker van de eindtoetsen 2019. Zoals reeds beschreven bij het gebruikte dataverzamelingsdesign onder aspect S2 zijn de cesuren op het gezamenlijk anker van de Eindtoetsen 2019 (bepaald door de Expertgroep Toetsen PO na de afname van de Eindtoetsen in 2019) het uitgangspunt geweest bij het vaststellen van de absolute cesuren voor de referentieniveaus IEP LVS toetsen versie 2. Door middel van twee schalingen/kalibraties (kalibratie van de items in de drie IEP Eindtoetsen 2017, 2018 en 2019 en kalibratie van de items in de IEP LVS-toetsen) zijn deze cesuren in het normeringsonderzoek voor het IEP LVS overgebracht naar de IEP LVS-toetsen versie 2 (zie figuur 5.1 en tabel 5.1). Evenals de kalibratie is de cesuurbepaling uitgevoerd in de applicatie Lexter. De cesuurpunten zijn berekend in hele scorepunten op de scoreschalen van de verschillende toetsen. Er is dus sprake van een criteriumgerichte interpretatie van de toetsscores, waarbij de toetsscores vergeleken worden met een absolute norm (i.e., het gezamenlijk anker van de Eindtoetsen 2019 als extern criterium).



Vervolgens wordt in paragraaf 6.2 'Referentieniveaus binnen toetsen' beschreven hoe de toetsen <1F-1F-1S\_v2 en <1F-1S\_v2 zijn opgebouwd op basis van de referentieniveaus van de IEP Eindtoets 2019. Op basis van tabel 6.2 kan dan worden geconcludeerd dat de items van verschillende referentieniveaus in de toetsen <1F-1F-1S\_v2 en <1F-1S\_v2 een representatieve deelverzameling zijn van alle items in de schaling op dat niveau voor zover het de hoeveelheid informatie betreft die zij geven in de afnamegroepen <1F-1F-1S\_v2 en <1F-1S\_v2.

In tabel 7.1 wordt, naast de cesuren voor de IEP LVS toetsen <1F-1F-1S\_v2 en <1F-1S\_v2 in hele scorepunten, voor iedere cesuur de bijbehorende vaardigheid (de vaardigheid van het naar boven afgeronde scorepunt van de cesuur), de lokale meetfout van de vaardigheid, de lokale betrouwbaarheid voor de IEP Eindtoets 2019 en de IEP LVS toetsen versie 2 en het percentage leerlingen dat het betreffende niveau ten onrechte wel of niet heeft gehaald (misclassificaties) gerapporteerd. Daarnaast zijn voor de IEP LVS toets <1F-1F-1S\_v2 per cesuur ook de 95%-betrouwbaarheidsintervallen op de bijbehorende vaardigheid en score berekend. Hoewel de 95%-betrouwbaarheidsintervallen van de 1F- en de 1S-cesuur in de toets <1F-1F-1S\_v2 elkaar net overlappen, is dit acceptabel gezien de functie van beide toetsen (i.e., het schatten van de vaardigheid in de context van leergroei en niet in de context van het nemen van belangrijke beslissingen). Wat betreft de IEP LVS toets <1F-1F\_v2 stellen de auteurs dat deze toets slechts één cesuur kent en er dus geen sprake is van overlappende betrouwbaarheidsintervallen. Tenslotte valt uit tabel 7.1 nog af te lezen dat de nauwkeurigheid waarmee de vaardigheid gemeten wordt groter is bij de 1F-cesuur dan bij de 1S-cesuur in beide IEP LVS toetsen.

*Conclusie:*

De standaardbepalingsmethode is gemotiveerd en op de juiste wijze uitgevoerd. Op aspect N1.1 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het volgende oordeel toegekend: **voldoende**.

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

*Bevindingen:*

n.v.t.

*Conclusie:*

**n.v.t.**

N1.3 Is er voldoende overeenstemming tussen de beoordelaars?

*Bevindingen:*

n.v.t.

*Conclusie:*

**n.v.t.**

### N2.1 Zijn de normgroepen groot genoeg?

#### *Bevindingen:*

Uit bijlage 4 ('TIA-analyses IEP LVS Rekenen') valt af te lezen dat de normgroepen voor de (in moeilijkheid oplopende) IEP LVS-toetsen <1F-1F\_v2 en <1F-1F-1S\_v2 in de normeringspopulatie respectievelijk 1077 en 2227 zijn, welke aantallen groot genoeg zijn om schattingen te kunnen maken voor de doelpopulatie. Op pag. 7, 3<sup>e</sup> regel van onderen, van het document 'Scenario's voor ijking van de eindtoetsen op de referentieniveaus' van Glas, Emons en Berding-Oldersma (december 2016) wordt namelijk gesteld dat voor een betrouwbare cesuur het aantal leerlingen in de steekproef minimaal 1000 moet zijn. Dit document is te vinden op de homepage van de Expertgroep Toetsen PO onder 'Overige Informatie – Onderzoeken', uitgevoerd door de Expertgroep.

#### *Conclusie:*

De normgroepen zijn groot genoeg. Op aspect N2.1 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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 representativiteit van de normeringspopulatie op de achtergrondvariabelen niet optimaal was om als normpopulatie voor het gehele reguliere basisonderwijs (doelpopulatie) te fungeren. Volgens de auteurs was het echter voor het bepalen van de cesuren ook niet per se noodzakelijk dat de verschillende categorieën leerlingen representatief vertegenwoordigd waren, mits er gebruikgemaakt werd van een gelinkt design met het gezamenlijke anker van de Eindtoetsen 2019 en mits alle categorieën van leerlingen in de steekproef vertegenwoordigd waren. Representativiteit dient in dit verband dan opgevat te worden in de betekenis dat het aannemelijk is dat op alle steekproeven die in de schaling gebruikt zijn een-en-hetzelfde IRT-model van toepassing is (zie ook opmerking 1 onder aspect S1).

#### *Conclusie:*

De normgroepen zijn representatief. Op aspect N2.2 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend.

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

#### *Bevindingen:*

De gevolgde standaardsettingsprocedure middels twee schalingen (zie figuur 5.1 en tabel 5.1) past bij de data van de normeringspopulatie en er zijn ook schattingsfouten van de cesuren berekend (zie tabel 7.1).

De normen voor de IEP LVS-toetsen <1F-1F\_v2 en <1F-1F-1S\_v2 zijn geldig tot en met het einde van het schooljaar 2020-2021.

*Conclusie:*

De normen zijn correct bepaald en er wordt duidelijk aangegeven tot wanneer (i.e., tot welk jaartal) de normen geldig zijn. Op aspect N2.3 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend.

**Betrouwbaarheid**

B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

*Bevindingen:*

Er is gebruikgemaakt van de programma's TiaPlus en Lexter om de betrouwbaarheidsgegevens te berekenen. In bijlage 4 worden de globale betrouwbaarheden (Cronbach's alpha) weergegeven voor de IEP LVS-toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2, welke zijn berekend met het programma TiaPlus. Het programma Lexter is gebruikt om de (conditionele) lokale betrouwbaarheid uit te rekenen bij de verschillende vaardigheidsniveaus behorend bij de cesuren in scorepunten, welke qua interpretatie grote overeenkomst met de globale betrouwbaarheidscoëfficiënt Cronbach's alpha uit de klassieke testtheorie (KTT). Deze coëfficiënt wordt ook in de psychometrische literatuur beschreven en als correct aangemerkt. Daarnaast worden in tabel 7.1 voor iedere cesuur ook nog het percentage leerlingen berekend dat het betreffende niveau ten onrechte wel of niet heeft gehaald (classificatiefouten). In bijlage 5 wordt op een correcte manier beschreven hoe deze classificatiefouten worden berekend. We kunnen ervan uitgaan dat de betrouwbaarheidsgegevens correct zijn berekend, omdat gebruikgemaakt is van bekende en algemeen beschikbare software.

*Conclusie:*

De betrouwbaarheidsgegevens zijn correct bepaald. Op aspect B1 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend.

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

*Bevindingen:*

In bijlage 4 kan afgelezen worden dat de globale betrouwbaarheden in termen van Cronbach's alpha (interne consistentie betrouwbaarheden) voor de IEP LVS toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 gelijk zijn aan respectievelijk 0.88 en 0.89. Omdat het behalen van een referentieniveau valt in de categorie minder belangrijke beslissingen op individueel niveau, is met deze waarden voor de globale betrouwbaarheden ruimschoots voldaan aan de eis van het COTAN beoordelingssysteem (Evers et al., 2010) dat de minimale betrouwbaarheidscoëfficiënt van toetsen voor minder belangrijke beslissingen tenminste 0.70 moet zijn. Tabel 7.1 laat zien dat aan deze minimale eis ook ruimschoots wordt voldaan voor de geschatte lokale betrouwbaarheid van alle referentiecesuren van de IEP LVS toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2, welke varieert van 0.817 – 0.887.

Verder laat tabel 7.1 nog zien dat de classificatiefouten (i.e., voor de twee IEP LVS-toetsen Rekenen het percentage leerlingen dat het betreffende niveau ten onrechte wel of niet heeft gehaald) loopt van 11% tot 13% (hoe hoger de lokale betrouwbaarheid, hoe lager de classificatiefout). Deze percentages hebben betrekking op scores dicht bij een cesuur

en er geldt dan ook dat het percentage misclassificaties bij een score verder van de cesuur af per definitie lager is. Omdat de berekende classificatiefouten in de context van de IEP LVS toetsen geen summatieve toetsen betreft waarop een leerling kan zakken of slagen, heeft een misclassificatie daarmee voor de leerling geen directe grote gevolgen. In combinatie met het feit dat de IEP LVS toetsen volgtoetsen zijn waar geen belangrijke beslissingen mee worden genomen, kan er geconcludeerd worden dat de classificatiefouten als acceptabel gezien kunnen worden.

*Conclusie:*

De betrouwbaarheid van de IEP LVS-toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft en ingestemd wordt met de eisen voor de betrouwbaarheid van het COTAN beoordelingssysteem (Evers et al., 2010). Op aspect B2 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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:*

De items toetsen de leerdoelen die ze beogen te toetsen. Speciaal is ook gekeken naar de items <1F, 1F, 1S. De items van het betreffende niveau toetsen ook daadwerkelijk de doelen behorende bij dat niveau. In zijn algemeenheid kan gesteld worden dat er geen misverstanden kunnen ontstaan over de juistheid van de gegeven antwoorden.

*Conclusie:*

De items van de toets dragen bij aan de validiteit van de toets.

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

*Bevindingen:*

In paragraaf 7.2 ('Passing van het meetmodel en nauwkeurigheid van de parameterschattingen') was al aannemelijk gemaakt dat er sprake was van een passing van het geassumeerde meetmodel (i.e., het Rasch model) en er dus mag worden uitgegaan van unidimensionaliteit, hetgeen impliceert dat aan de noodzakelijke (maar niet voldoende) voorwaarde van constructvaliditeit wordt voldaan. In hoofdstuk 8 wordt aanvullend onderzoek verricht naar de onderstaande andere aspecten, welke kunnen worden opgevat als enkele argumenten die pleiten voor de constructvaliditeit van de IEP LVS-toetsen Rekenen: (1) soortgenotenonderzoek in de vorm van correlatieel onderzoek tussen zowel toetsen versie 1 en 2 per vaardigheid als tussen toetsen van verschillende vaardigheden en (2) itemkwaliteit (i.e., psychometrische kwaliteit van de items).

1. Er is op de IEP LVS toetsen een soortgenotenonderzoek uitgevoerd in de vorm van onderzoek naar convergente validiteit tussen twee binnen één leerjaar opeenvolgende IEP LVS toetsen van dezelfde vaardigheid en tevens is er onderzoek gedaan naar divergente validiteit tussen IEP LVS toetsen van verschillende vaardigheden (Lezen, Taalverzorging en Rekenen) binnen één en hetzelfde leerjaar. Met andere woorden, er is gebruikgemaakt van een Multi-Trait Multi-Method matrix (MMTM), waarbij scores op IEP LVS rekentoetsen groep 6, 7, 8 hoog zouden moeten correleren met toetsen die hetzelfde construct meten en laag met toetsen die een ander construct meten. De resultaten van deze beide onderzoeken worden gerapporteerd in tabel 8.1. Conform de verwachting is hieruit af te lezen dat de correlatie tussen de twee toetsen Rekenen (i.e., versie 1 en versie 2) structureel hoger is dan tussen twee toetsen van verschillende vaardigheden binnen hetzelfde leerjaar, hetgeen de aannemelijkheid van de constructvaliditeit ondersteunt.
2. In tabel 8.2 zijn de gemiddelden en ranges van de p- en rit-waarden IEP LVS toetsen Rekenen weergegeven (gebaseerd op de uiteindelijke selectie van items per toets). Uit deze tabel blijkt dat met name de gemiddelden voldoen als criterium voor de itemkwaliteit per toets met p-waarden tussen de 0.59 en 0.62. De grootte van de ranges p-waarden in tabel 8.2 (variërend tussen 0.24 en 0.61) wordt sterk bepaald door enkele outliers (zie hiervoor ook bijlage 4).

Opm.:

Pag. 19: Naast onderzoek naar (1) soortgenotenonderzoek in de vorm van correlatieel onderzoek tussen zowel toetsen versie 1 en 2 per vaardigheid als tussen toetsen van verschillende vaardigheden en (2) itemkwaliteit (i.e., psychometrische kwaliteit van de items), zou het aanbeveling verdienen om ook nog aanvullend onderzoek te doen naar de intercorrelaties tussen de inhoudelijke domeinen Rekenen (Getallen, Verhoudingen, Meten en meetkunde en Verbanden) binnen de IEP LVS rekentoetsen als verdere empirische evidentie voor constructvaliditeit. Hoge intercorrelationele samenhangen duiden er dan op dat al deze vier inhoudelijke domeinen hetzelfde onderliggende construct (i.e., rekenvaardigheid) meten en kan dus opgevat worden als verdere empirische evidentie voor constructvaliditeit. Dit punt is meer een suggestie dan een strikte eis.

De toetsaanbieders hebben toegezegd dit advies in de toekomst ter harte te nemen.

*Conclusie:*

De gerapporteerde resultaten in Hoofdstuk 8 ('Constructvaliditeit') van de WV vormen een psychometrische ondersteuning voor de constructvaliditeit van de toetsen IEP LVS Rekenen versie 2 en er wordt dus gemeten wat men beoogt te meten, namelijk rekenvaardigheid in de leerjaren 6, 7, 8. Op aspect V2 wordt derhalve aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 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 correcte manier gemeten?

#### *Bevindingen:*

De twee IEP LVS-toetsen Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 zijn gekalibreerd op eenzelfde schaal. De vaardigheidsscores op deze twee toetsen zijn hierdoor onderling vergelijkbaar en de vaardigheidsontwikkeling van de leerlingen kan worden gevolgd door hun scores op de verschillende opeenvolgende momenten met elkaar te vergelijken.

Omdat deze vaardigheidsschaal moeilijk te interpreteren is voor leerkrachten, is de vaardigheidsscore getransformeerd naar een lineaire schaal die ook wel de ontwikkelscoreschaal wordt genoemd. Dit is een schaal waar leerkrachten en leerlingen aan gewend zijn en die, helder en eenduidig voor leerkrachten en leerlingen, verfijnd weergeeft welk vaardigheidsniveau een leerling heeft. In bijlage 6 wordt een omzettingsschaal gegeven voor verschillende ruwe scores (lopend van 0 t/m 40), vaardigheidsscores en ontwikkelscores. Hierbij dient nog wel opgemerkt worden dat de ontwikkelscore 'geen resultaat' wordt gegeven voor scores onderaan de scoreschaal waar de vaardigheidsschatters een lokale betrouwbaarheid hebben die lager is dan de vereiste 0.70, die COTAN hanteert. Aan de scorepunten aan de bovenkant van de scoreschaal met een lokale betrouwbaarheid onder de 0.70 wordt de ontwikkelscore van het hoogste scorepunt met een lokale betrouwbaarheid van  $\geq 0.70$  toegekend.

#### *Conclusie:*

Er is voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt en de groei wordt op een adequate manier gemeten. Op aspect Va1 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend.

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

#### *Bevindingen:*

De (conditionele) lokale betrouwbaarheid voor ieder scorepunt wordt op dezelfde manier geschat als de (conditionele) lokale betrouwbaarheid van de cesuurpunten (op de thetaschaal) voor de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 (zie bijlage 5 voor uitleg over de gedetailleerde berekening). Deze (conditionele) lokale betrouwbaarheid vertoont qua interpretatie grote overeenkomst met de globale betrouwbaarheidscoëfficiënt Cronbach's alpha uit de klassieke testtheorie (KTT) en wordt in de psychometrische literatuur beschreven en als correct aangemerkt. Verder kan nog vermeld worden dat de lokale betrouwbaarheden van de thetaschaal en de ontwikkelscoreschaal gelijk zijn, omdat de ontwikkelscoreschaal een lineaire transformatie is van de thetaschaal.

#### *Conclusie:*

De betrouwbaarheid van de groei op de ontwikkelscoreschaal wordt correct weergegeven. Op aspect Va2 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel '**voldoende**' toegekend.

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

*Bevindingen:*

De interpretatie van de vaardigheidsontwikkeling van de leerling wordt voor de leerkracht ondersteund door de twee grafische weergaven in de hoofdstukken 9 en 10. De voortgangsgrafiek en de leergroeimeter, zijn gecombineerd te gebruiken. Voor leerkrachten zijn de leervorderingen van een leerling ook digitaal beschikbaar in het IEP LVS, welke door de leerkrachten ook geprint kunnen worden in de vorm van de IEP LVS Talentenkaart. Voor de interpretatie van de leervorderingen worden aan leerkrachten handvatten gegeven in de handleiding van het IEP LVS <https://handleiding.toets.nl/snel-op-weg-met-het-iep-lvs-245>.

*Conclusie:*

Er worden voldoende gegevens verstrekt (aan de gebruiker) over hoe groei geïnterpreteerd dient te worden. Op aspect Va3 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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:*

Leraren kunnen de leervorderingen van de leerlingen per domein inzien (in percentage goed-scores). Ook kunnen de leraren de antwoorden van de leerlingen inzien wat het formatief gebruik van de toetsen bevordert. Aan leerkrachten worden handvatten gegeven voor de interpretatie van de leervorderingen. Op de IEP portal is ook nog extra informatie te vinden voor leraren die zich nog verder in deze materie willen verdiepen.

Speciaal voor ouders/verzorgers is er een leesbare Leeswijzer die hen handvatten geeft voor de interpretatie van de leervorderingen die op de Talentenkaart zijn weergegeven. De Talentenkaart is zo gemaakt dat deze voor de meeste leerlingen te begrijpen zal zijn.

*Conclusie:*

De aanbieder (i.e., Bureau ICE) 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 toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 het oordeel **voldoende** toegekend.

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

*Bevindingen:*

De opmerkingen die onder de paragraaf Va3 werden gemaakt zijn ook hier van toepassing: alle relevante informatie dient in de verantwoording aanwezig te zijn.

Na enig zoekwerk werd het volgende vastgesteld. De leerkracht wordt ondersteund bij de interpretatie van de vaardigheidsontwikkeling van de leerling door het gecombineerd gebruik van de twee grafische weergaven, de voortgangsgrafiek en de leergroeimeter. De leerkracht kan hiermee evalueren in welke mate de leerling ten opzichte van zijn/haar verwachting en/of ten opzichte van de verwachte groeifactor zich ontwikkelt en kan hij/zij inschatten hoe waarschijnlijk het is dat de leerling het beoogde streefniveau zal gaan bereiken. In de 'Handreiking interpreteren toetsresultaten' worden leerkrachten geholpen bij de interpretatie van de ontwikkelscores en krijgen zij advies over het bepalen of een toets 'passend' was qua niveau voor de leerling.

De resultaten kunnen gebruikt worden om op zowel leerlingniveau, als op groeps- en schoolniveau te evalueren en analyseren. In de bovenbouw zie je welk referentieniveau de leerling beheerst en welke ontwikkelscore daarbij hoort. Ook kun je de resultaten vergelijken met het landelijk gemiddelde zodat je weet hoe een leerling of groep ervoor staat. Ook is het mogelijk om te sorteren op categorie of domein, zodat je kunt zien voor welke leerlingen nog extra vervolgstappen noodzakelijk zijn.

*Conclusie:*

Er is een evaluatie van de leervorderingen en op basis van deze evaluatie worden vervolgstappen geformuleerd. Op aspect I2 wordt aan de toetsen IEP LVS Rekenen <1F-1F\_v2 en <1F-1F-1S\_v2 voor groep 6, 7, 8 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:*

We hebben alle items <1F, 1F, 1S bekeken. De items van het betreffende niveau sluiten goed aan bij het niveau dat bij de items is aangegeven en ze toetsen ook daadwerkelijk de doelen behorende bij dat niveau. Ook sluiten de items goed aan op de betreffende domeinen.

*Conclusie:*

De items van het betreffende niveau sluiten goed aan bij het niveau dat bij de items is aangegeven en ze toetsen ook daadwerkelijk de doelen behorende bij dat niveau. Ook sluiten de items goed aan op de betreffende domeinen. Het oordeel is '**voldoende**'.



### 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>Voldoende</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>Voldoende</b>

### 4. Literatuurlijst

- Bezdán, E., Binsbergen, M., Haitjema, T., Helsloot, J. & Laan, J. (2020). Verantwoording IEP LVS-toetsen Rekenen. Culemborg: Bureau ICE.