

## 1. Algemene informatie

### Algemeen en meetpretentie

De volgtoetsen van Diacijfer maken onderdeel uit van het Dia-LVS PO, een digitaal leerlingvolgsysteem met volgtoetsen voor taal en rekenen in lijn met het Referentiekader (Expertgroep Doorlopende Leerlijnen Taal en Rekenen, 2009). De Dia-volgtoetsen vormen een samenhangend geheel met de Dia-eindtoets die in 2016 officieel is toegelaten op basis van de Verantwoording Dia-eindtoets (Diataal BV, 2016). De volgtoetsen Diacijfer zijn adaptief en meten de algemene rekenvaardigheid van leerlingen in relatie tot de referentieniveaus. In deze beoordeling ligt Diacijfer 678 voor.

De volgtoetsen Diacijfer meten de rekenvaardigheid in lijn met de referentieniveaus rekenen zoals beschreven door de Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2009) van leerlingen vanaf eind groep 5 tot en met midden groep 8 van het basisonderwijs, en de ontwikkeling daarvan in deze onderwijsperiode. Er zijn zes adaptieve volgtoetsen, bedoeld om bij de leerlingen uit deze groepen het niveau en de groei van de rekenvaardigheid te kunnen vaststellen.

### Doelgroep

De doelgroep voor de toetsen die hier worden verantwoord bestaat uit de leerlingen in de bovenbouw van het reguliere en speciaal basisonderwijs vanaf eind groep 5 tot en met midden groep 8. De toetsen Diacijfer zijn adaptief en sluiten daardoor zo goed mogelijk aan bij het niveau van iedere leerling. Daardoor zijn ze ook af te nemen in het speciaal basisonderwijs.

### Gebruiksdoel en functie

De volgtoetsen Diacijfer zijn bedoeld om per meetmoment het niveau van de rekenvaardigheid van leerlingen te bepalen. Doordat de vaardigheidsscore (RN, rekenniveau) door middel van IRT-analyses op één schaal is gebracht, is het bovendien mogelijk om groei van leerlingen te volgen. Het volgen van de ontwikkeling van de rekenvaardigheid is naast de niveaubepaling het tweede primaire gebruiksdoel van de toetsen Diacijfer.

### Inhoudelijke theoretische inkadering:

De volgtoetsen Diacijfer zijn genormeerd aan de referentiesets rekenen (College voor Toetsen en Examens, 2016).

### Inhoud van het toetspakket

Het toetspakket Diacijfer 678 bestaat uit de volgende documenten:

- Wetenschappelijke verantwoording, deze bevat informatie over:
  - Uitgangspunten (hoofdstuk 2);
  - Inhoudsverantwoording (hoofdstuk 3);
  - Kalibratie en normering (hoofdstuk 4);
  - Betrouwbaarheid en validiteit (hoofdstuk 5);
  - 13 bijlagen, waaronder items en antwoorden.
- Handleiding Diacijfer 678
- Toetsreglement Dia-LVS PO
- Wegwijs in Dia-groeiwijzer
- Inzage digitale toetsitems diacijfer

## 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 (WV) en handleiding veelal letterlijk vermeld.*

### **De kwaliteit van de dataverzameling**

#### S1 Is de steekproef representatief?

##### *Bevindingen:*

Het ontwikkelde toetsmateriaal is in twee achtereenvolgende schooljaren bij de doelgroepen getest. In het schooljaar 2018-2019, in de maanden februari en maart is het materiaal onderzocht in een kalibratieonderzoek dat plaatsgevonden heeft in de groepen 5 tot en met 8. In het schooljaar 2019-2020 vonden op drie verschillende meetmomenten normeringsonderzoeken plaats. De beginmeting in de maanden september en oktober (voor de groepen 6, 7 en 8), de middenmeting in de maand januari (voor de groepen 6, 7 en 8) en de eindmeting in de maand juni (voor de groepen 5, 6 en 7).

Aan de hand van de kalibratieonderzoeken in het schooljaar 2018-2019 zijn de toetsen voor de normeringsonderzoeken samengesteld. Na afloop van het normeringsjaar is alle data samengenomen om de kalibratie opnieuw uit te voeren en de samenstelling van de toetsen definitief te maken. Bedoelde toetsen worden in de wetenschappelijke verantwoording (WV) besproken.

Voor het kalibratieonderzoek voor de volgoetsen 6, 7 en 8 zijn scholen op verschillende manieren geworden. In 2017 heeft Diataal BV een partnerscholenconcept opgezet waarbij het doel was om 60 scholen te werven. Deze scholen verspreid over heel Nederland verbinden zich voor die of vier leerjaren aan het project. Naast de partnerscholen zijn extra scholen geworden om tot een representatieve steekproef te komen wat betreft regio en de mate van verstedelijking.

Aan het kalibratieonderzoek deden in totaal 89 scholen mee uit de groepen 5, 6, 7 en 8. Leerlingen maakten twee toetsdelen met nieuw materiaal ten behoeve van de ontwikkeling van de toetsen en daarnaast een selectie vragen uit de openbare referentiesets rekenen (College voor Toetsen en Examens, 2014), een selectie vragen uit zowel de Dia-eindtoets 2017 als 2018 en een selectie vragen met (nieuw) materiaal dat ook in het nieuwe te ontwikkelen LVS voor het voortgezet onderwijs wordt opgenomen. Bij de kalibratie zijn de data van zowel het oorspronkelijke kalibratieonderzoek als van de normeringsonderzoeken samengenomen.

In tabel 11 van de WV wordt de steekproef beschreven naar de achtergrondvariabelen geslacht (jongen - 9564 leerlingen, meisje - 9395 leerlingen), regio (noord - 2660 leerlingen, oost - 3371 leerlingen, zuid - 3856 leerlingen, west - 9072 leerlingen), denominatie (openbaar - 3170 leerlingen, overig - 15789 leerlingen), urbanisatiegraad (G4 - 1246 leerlingen, G5-G37 - 3485 leerlingen, overig - 14218 leerlingen).

Om na te gaan of de steekproef representatief is, is getoetst op verschillen tussen de populatie en de steekproef. Omdat door de grote aantallen de verschillen al snel significant zijn, keek men naar de effectgrootte door coëfficiënt phi te berekenen. Tabel 11 van de WVV laat zien dat, behalve voor geslacht, voor de drie andere achtergrondvariabelen de aantallen in de steekproef teveel afwijken van die in de populatie en de steekproef derhalve niet representatief is. Bij het kalibratieonderzoek zal er door middel van DIF-analyses nagegaan worden of de itemparameterschattingen als gelijk kunnen worden beschouwd in de verschillende subgroepen op basis van de achtergrondvariabelen die onderscheiden kunnen worden. Wanneer dat het geval is, zullen de items in de subgroepen gelijkwaardig functioneren als indicatoren voor de vaardigheid. Aangezien subgroepen kunnen verschillen in gemiddelde en variantie van het vaardigheidsniveau en qua aantallen niet in dezelfde mate zijn gerepresenteerd als in de populatie, worden bij het vaststellen van de relatieve normen steekproefgewichten toegepast. Op deze manier wordt er gecorrigeerd voor het feit dat de steekproef niet geheel representatief is.

*Conclusie:*  
**voldoende**

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

### *Bevindingen*

Het materiaal dat uitgezet is in het kalibratieonderzoek bevat, naast nieuw materiaal, een selectie vragen uit de openbare referentiesets, een selectie vragen uit de Dia-eindtoets 2017 als 2018 en een selectie vragen met materiaal dat ook in het nieuwe Dia LVS voor het VO voorkomt.

Het nieuwe materiaal is verdeeld over 78 itemblokken van 11 vragen en 4 losse vragen. Het materiaal afkomstig uit de openbare referentiesets, Dia-eindtoets en Diacijfer VO is verdeeld over 12 itemblokken met 11 vragen. Zie tabel 13 van de WVV voor een overzicht.

In tabel 56 tot en met tabel 59 in bijlage 9 staan de designs van de dataverzameling van het kalibratieonderzoek voor de groepen 5 tot en met 8 grafisch weergegeven. De stippen in de tabellen geven aan welke itemblokken in welke toetsen opgenomen zijn. Ook wordt aangegeven hoeveel vragen er in iedere set voorkomen. Leerlingen zijn aselect toegewezen aan een toets, zodat leerlingen binnen eenzelfde klas niet allemaal dezelfde toets maken.

Tabel 14 geeft per groep een overzicht van de verdeling van opgaven in de toets uit verschillende itemblokken. Zo maken leerlingen uit groepen 6,7 en 8 naast 4 blokken bedoeld voor hun eigen niveau ook 1 blok met items uit een leerjaar lager en 1 blok uit de zogenaamde RVE-blokken. Leerlingen uit groep 5 maken opgaven uit 5 itemblokken bedoeld voor groep 5, aangevuld met 1 van de 4 losse items. In groep 5 (tabel 56, bijlage 9) zijn er zo 18 verschillende toetsen samengesteld waarbij er overlap tussen de verschillende toetsen bestaat. Voor de groepen 6, 7 en 8 zijn er steeds 24 verschillende toetsen samengesteld met ook hier weer overlap tussen de verschillende toetsen.

Uit het nieuwe materiaal zijn items geselecteerd voor de samenstelling van de toetsen voor de normeringsonderzoeken. De data van deze normeringsonderzoeken is toegevoegd aan die van het kalibratieonderzoek om een nieuwe kalibratie uit te kunnen voeren. Elk item in de kalibratie is gemaakt door 315 tot 3101 leerlingen (gemiddeld 1168). De items

met de lagere leerlingaantallen betreffen items die niet geselecteerd zijn voor het normeringsonderzoek en dus alleen in het oorspronkelijke kalibratieonderzoek zijn afgenomen. De items die zijn geselecteerd voor de uiteindelijke toetsen zijn gemaakt door 423 tot 3101 leerlingen (gemiddeld 1473).

Uit het voorgaande blijkt dat we te maken hebben met een onvolledig design verbonden door voldoende grote itemankers en dat is afgenomen bij voldoende grote aantallen leerlingen. Dit design is geschikt om IRT-analyses mee uit te voeren. De resultaten van het kalibratieonderzoek met behulp van het 2PL model, worden besproken in paragraaf 4.1.4. De itemparameters werden geschat met behulp van de marginal maximum likelihood methode in het computerprogramma Lexter (Glas & van Buuren 2019). In bijlage 10 zijn in tabel 60 tot en met tabel 65 de geschatte itemparameters te vinden. Op basis van een procedure die door de COTAN voorgesteld wordt om de nauwkeurigheid van de parameterschattingen te berekenen, kan vastgesteld worden dat de schattingen voldoende nauwkeurig zijn. Met behulp van DIF-statistieken is de itemfit en de modelfit bepaald. Geconcludeerd kan worden dat het 2PL model goed past bij de data.

*Conclusie:*  
**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:*  
**n.v.t.**

*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:*  
**n.v.t.**

*Conclusie:*  
**n.v.t.**

## **Normering**

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

*Bevindingen:*  
Gelijktijdig met het kalibratieonderzoek in de groepen 5 tot en met 8 van het PO, startte ook een kalibratie- en normeringsonderzoek voor Diacijfer in het VO. Op basis van de uitkomsten en resultaten uit het PO- en VO-onderzoek samen is de RN-schaal ontwikkeld. Hierbij is in eerste instantie gekozen om itemparameters uit de Dia-eindtoets, die in beide

onderzoeken uitgezet zijn, te fixeren. Hierdoor kwamen de tempparameters op de schaal van het gezamenlijk anker van alle eindtoetsaanbieders te liggen, dat gebruikt wordt bij de verschillende eindtoetsen om tot een gelijke normering voor de referentieniveaus 1F en 1S te komen. Vervolgens is door middel van een lineaire transformatie de RN-schaal zo gemaakt dat het gemiddelde gelijk is aan 1000 en de standaarddeviatie gelijk aan 100. De cesuren voor de referentieniveaus (1F en 1S en in het VO ook 2F en 3F) zijn vervolgens met behulp van de openbare referentiesets overgezet op de RN-schaal. Scholen die meededen aan het normeringsonderzoek kregen voor hun leerlingen volledige resultaten, dat wil zeggen, een RN-score, een referentieniveau en een percentielscore.

Met de RN-score is het referentieniveau, het niveau dat beschrijft wat een leerling op een bepaald moment in zijn schoolloopbaan moet beheersen, vastgesteld. Om met de toets een referentieniveau te kunnen vaststellen (absolute normering), zijn er tijdens het onderzoek items opgenomen uit de referentiesets. Voor elk van deze referentiesets zijn er grensscores vastgesteld door Stichting Cito. Grensscores worden bepaald met een groep beoordelaars die bestaat uit stakeholders uit het veld; docenten, wetenschappelijke instituten, onderwijsadvies, SLO. Hierbij heeft de Stichting Cito een combinatie toegepast van een Angoff procedure en een bookmark procedure. In de Angoff procedure geven beoordelaars aan welk percentage van de grensleerlingen (leerlingen die het referentieniveau net hebben bereikt) de opgaven goed maakt. In de bookmark procedure krijgen de beoordelaars de opgaven voorgelegd op volgorde van moeilijkheid en wordt hen gevraagd vanaf welke opgaven minder dan 67% van de grensleerlingen de opgave goed maakt. Na overleg en vergelijking met reeds vastgestelde cesuren hebben beoordelaars een definitief oordeel gegeven. Middeling van de definitieve oordelen van alle beoordelaars heeft vervolgens geresulteerd in de geadviseerde grensscores, oftewel de referentiecesuren. De referentiecesuren die zijn bepaald voor de verschillende niveaus kunnen door toetsontwikkelaars worden overgebracht op hun eigen toetsen door opgaven uit de openbaar beschikbare referentiesets tegelijkertijd op te nemen met eigen materiaal. Er is een applicatie gebruikt waarmee het mogelijk is om de grensscores voor de referentieniveaus te bepalen voor selecties van de opgaven uit de openbare referentiesets. In het onderhavige geval is gekozen voor itemschattingen met het OPLM omdat het OPLM het meest lijkt op het 2PL. De thetas voor de referentiecesuren op de schaal voor Diacijfer zijn bepaald met behulp van ware score equivalering. Omgerekend naar de RN-schaal zijn deze respectievelijk 913 voor 1F en 1135 voor 1S. In tabellen 17 en 18 staan per toets en route de lokale standaardmeetfouten bij deze cesuren.

Leerlingen krijgen naast een referentieniveau ook een percentielscore (relatieve normering). Op deze manier kunnen de resultaten van de leerlingen vergeleken worden met leerlingen in hetzelfde leerjaar, op hetzelfde meetmoment. Om te corrigeren voor tekorten in de steekproef zoals eerder beschreven, zijn steekproefgewichten toegepast. De leerlingaantallen na het toepassen van de steekproefgewichten zijn weergegeven in tabel 47 tot en met tabel 55 in bijlage 8.

*Conclusie:*  
**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:*

In bijlage 8 worden de steekproeven per onderzoeksmoment weergegeven. In tabel 43 tot en met tabel 46 in bijlage 8 zijn de steekproefgegevens opgenomen van de leerlingen die aan het kalibratieonderzoek hebben deelgenomen in het schooljaar 2018-2019. Deze gegevens zijn uitgesplitst voor de groepen 5 tot en met 8. In bijlage 8 zijn de steekproefgegevens weergegeven van de leerlingen die aan de normeringsonderzoeken hebben meegedaan. Deze gegevens zijn uitgesplitst per groep en per leermoment. De steekproeven voor het kalibratie- en normeringsonderzoek zijn van voldoende grootte.

*Conclusie:*

**voldoende**

N2.2 Zijn de normgroepen representatief?

*Bevindingen:*

Bij de bespreking van de representativiteit bij aspect S1 werd geconstateerd dat de steekproeven alleen met betrekking tot geslacht representatief waren. Met dit gebrek aan representativiteit is rekening gehouden door de toepassing van steekproefgewichten.

*Conclusie:*

**Voldoende**

N2.3 Zijn de normen correct bepaald?

*Bevindingen:*

Met verwijzing naar aspect N.1.1 kan geconstateerd worden dat de absolute en de relatieve normen correct zijn bepaald.

*Conclusie:*

**Voldoende**

### ***Betrouwbaarheid***

B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

*Bevindingen:*

Bij het berekenen van betrouwbaarheidsmaten werd gebruik gemaakt van simulaties. Voor de onderhavige toetsen werd een zogenaamde globale betrouwbaarheid geschat. De globale betrouwbaarheid is de proportie door scoring verklaarde variantie en wordt gegeven in het softwarepakket Lexter. In tabel 20 staan voor de verschillende momenten de betrouwbaarheden weergegeven. De 18 betrouwbaarheden zijn alle groter dan of gelijk aan 0,93.

Leerlingen worden wat betreft referentieniveau en de percentielscores op een aantal manieren gecategoriseerd. De percentielscores worden gecategoriseerd in vijf groepen en voor de referentieniveaus in twee of drie groepen. Met behulp van de 'ware vaardigheid' en de geschatte vaardigheid zijn betrouwbaarheidstabellen bepaald die in twee maten worden samengevat: de marginal classification accuracy en de accuracy plus/minus 1. De betrouwbaarheidstabellen zijn voor ieder meetmoment te vinden in bijlage 13. Tabel 21.2 bevat voor ieder meetmoment de twee samenvattende maten. De marginal classification accuracy varieert tussen 72% en 81% en de accuracy plus/minus 1 tussen 99% en 100%. Voor beide geldt dat het aantal misclassificaties klein is. Ook voor de classificaties in referentieniveaus zijn in bijlage 13 betrouwbaarheidstabellen opgenomen. In geen enkel geval krijgt een leerling een niveau dat meer dan één afwijkt.

*Conclusie:*

**voldoende**

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

*Bevindingen:*

Met verwijzing naar aspect B1 kan geconstateerd worden dat de globale betrouwbaarheden en de categoriseringenvoldoende zijn.

*Conclusie:*

**Voldoende**

### **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 items in de toets toetsen de doelen die ze beogen te toetsen. Er kunnen geen misverstanden ontstaan over de juistheid van de gegeven antwoorden.

#### *Conclusie:*

De items van de toets dragen bij aan de validiteit van de toets, derhalve '**voldoende**'.

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

#### *Bevindingen:*

Met verwijzing naar het kalibratieonderzoek bij S2 kan geconstateerd worden dat er sprake is van een eendimensionale schaal. Daarnaast is de soortgenoot validiteit onderzocht door de correlatie tussen de vaardigheidsscores op Diacijfer en Cito LVS te berekenen. De correlatie is 0,85.

#### *Conclusie:*

**voldoende**

### **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:*

De volgtoetsen Diacijfer zijn met behulp van het 2PL model op één schaal gebracht. Daardoor zijn hun vaardigheidsscores onderling vergelijkbaar en kunnen op verschillende momenten vergeleken worden. De vaardigheidsschaal loopt van RN-score 400 tot 1145 voor de leerlingen van eind groep 5 tot middengroep 8. Om voor leraren, leerlingen en ouders de doorlopende vaardigheidsontwikkeling inzichtelijk te maken en te helpen bij de interpretatie daarvan zijn de vaardigheidsscores omgezet naar een kleur op de kleurenliniaal. De achterliggende cesuren voor 1F, 2F en 3F op de vaardigheidsschaal zijn als ondergrens genomen voor de kleuren groen, paars en bruin. In de grafieken van de leerlingen wordt naast de RN-score per toetsmoment ook een 'streefscoregebied' aangegeven. Zo wordt zichtbaar gemaakt hoe een leerling presteert ten opzichte van het 'streefniveau' dat voor deze leerling haalbaar zou zijn.

#### *Conclusie:*

**Voldoende**



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

*Bevindingen:*

In de WVV en de Handleiding werd de betrouwbaarheid van de groei op de vaardigheidsschaal in eerste instantie niet (adequaat) weergegeven. In navolging van de initiële verantwoording van Diatekst, is nu een soortgelijke paragraaf 5.3 (Volgaspect) toegevoegd. In deze paragraaf wordt geïllustreerd dat het in principe mogelijk is voor iedere vaardigheidsschatting standaard meetfouten te bepalen. Vervolgens wordt een mogelijke toepassing van deze gegevens beschreven, namelijk het bepalen van de Reliable Change Index (Jacobsen & Truax, 1991; Jabrayilov, Emons, & Sijtsma, 2016). Momenteel wordt voor gebruikers geen inschatting van de onderwijskundige en statistische significantie van de groei van individuele leerlingen gemaakt; in de verantwoording wordt aangetoond dat dit in principe wel mogelijk is.

*Conclusie:*

**Voldoende**

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

*Bevindingen:*

Met verwijzing naar Va1 en de Handleiding kan geconstateerd worden dat er aan de gebruiker adequate informatie over de interpretatie van groei verstrekt wordt.

Vanuit onderwijskundig perspectief wordt opgemerkt dat de handleiding van Diacijfer een duidelijke en uitgebreide beschrijving bevat van hoe de leerkracht de resultaten van de leerlingen dient te interpreteren. De score van een leerling wordt uitgedrukt in een rekenniveau (RN). In een groeigrafiek wordt de RN-score gegeven met bijbehorende streefscore. Hiermee kan de leerkracht zien hoe een leerling ten opzichte van de streefscore en van zichzelf groeit. Streefscores zijn gekoppeld aan het verwachte uitstroomprofiel. Met de percentielscore kan de leerkracht de leerling vergelijken met het landelijk gemiddelde.

Vanuit psychometrisch perspectief wordt aangevuld dat de RN-score van een leerling is gekoppeld aan een kleur op een liniaal (zie Figuur 1: De kleurschaal van Diacijfer), waardoor groei van de leerling inzichtelijk wordt gemaakt voor leerkrachten, leerlingen en ouders. Hierdoor kunnen de volgtoetsen Diacijfer niet alleen summatief, maar vooral ook formatief gebruikt worden. De kleurenliniaal is gebaseerd op een vaardigheidsschaal die doorloopt van het basisonderwijs tot en met het voortgezet onderwijs, waardoor leerlingen gevolgd kunnen worden in een doorlopende leerlijn. De kleuren van lichtgrijs t/m roze worden gegeven in het basisonderwijs en de kleuren die erna komen worden pas gegeven in het voortgezet onderwijs. De referentieniveaus 1F en 2F worden gehaald in het kleurgebied groen, respectievelijk paars. Vervolgens zijn er op deze schaal tussenniveaus bepaald (de groei van 1F naar 2F loopt via oranje en rood, etc.), waardoor de ontwikkeling van leerlingen fijnmaziger is af te lezen dan de referentieniveaus (kleuren kunnen worden gezien als 'ontwikkelingszones').

Naast de RN-score van de leerling is in de groeigrafieken ook een gebied gemarkeerd dat het streefscoregebied aangeeft, waardoor in de groeigrafiek in een oogopslag zichtbaar is

hoe een leerling presteert t.o.v. zijn leeftijdsgenoten uit de landelijke steekproef. De leerkracht kan deze gebieden per leerling instellen op laag (rond het 20<sup>e</sup> percentiel), midden (rond het 50<sup>e</sup> percentiel, de standaard) of hoog (rond het 80<sup>e</sup> percentiel). Een leerkracht kan, eventueel in gesprek met ouders en leerlingen, bepalen of gebruikgemaakt wordt van de mogelijkheid om het streefscoregebied naar beneden of juist naar boven bij te stellen. In de 'Handleiding Diacijfer' wordt het bovenstaande gedetailleerd toegelicht voor de leerkracht, waarin ook voorbeeldresultaten worden gedemonstreerd op school-, cohort-, groeps- en leerlingniveau op basis van vaardigheids- en streefscores.

*Conclusie:*

**Voldoende**

### ***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:*

In de Handleiding worden begrippen zoals referentieniveaus, vaardigheidsscores, streefscores, percentielscores, (sub)domeinscores begrijpelijk uitgelegd. De handleiding Diacijfer 678 bevat een bijlage waarin *algemene* informatie voor ouders over de Dia-toetsen is opgenomen.

*Conclusie:*

De toetsaanbieder (i.e., Diataal) levert een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/verzorgers/voogden/docenten begrijpelijk is. Het oordeel is derhalve '**Voldoende**'.

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

*Bevindingen:*

De Handleiding bevat diverse aanknopingspunten om de leervorderingen te evalueren. Zo wordt in afbeelding 6 per leerling het behaalde rekenniveau, referentieniveau en percentielscore weergegeven. Op basis hiervan kan de leraar bepaalde vervolgstappen formuleren.

De resultaten worden op allerlei manieren inzichtelijk gemaakt. O.a. op cohort, groeps- en leerlingniveau.

Tevens vermeldt de handleiding: "*Gedurende schooljaar 2020-2021 wordt de resultaatmodule aangepast en aangevuld met nieuwe functionaliteiten. Als er een aanpassing is geweest ontvang je hier bericht over en zal deze handleiding worden aangepast.*" Hieruit kan geconcludeerd worden dat de toetsenmaker voornemens is nog extra functionaliteiten toe te voegen.

Op leerlingniveau zijn o.a. de vaardigheidsscore, een vergelijking van de vaardigheidsscore met de score op de NSCCT en de percentielgroep waar de leerling in

scoort gegeven. Ook zijn de domeinscores per leerling gegeven zodat te zien is welke goed-percentages een leerling op de verschillende rekendomeinen heeft behaald.

*Conclusie:*

De leerlingresultaten zijn op allerlei manieren te analyseren en bieden zo aanknopingspunten om mogelijke vervolgstappen te formuleren. Het oordeel is derhalve **'voldoende'**.

### **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:*

De inhoud van de toets sluit voldoende aan op de kennis en vaardigheden in de referentieniveaus van de verschillende domeinen.

*Conclusie:*

**'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

- Hacquebord, H., Albers, M., Biesebeek, N. te., Hol, M., Schilstra, G. Schippers, G. & Kok, M. (2021). Wetenschappelijke verantwoording Diacijfer 678. Diataal BV Groningen.