

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de wetenschappelijke verantwoording en de Handleiding veelal letterlijk vermeld. De wetenschappelijke verantwoording heeft betrekking op de uitgangspunten van de toetsconstructie, de normen, de betrouwbaarheid en meetnauwkeurigheid en de validiteit. De Handleiding heeft betrekking op het gebruik van de toets, communicatie over de toetsgegevens en de inhoudsverantwoording.

Algemeen

Het Cito Volgstelsel primair en speciaal onderwijs beoogt de vorderingen van individuele leerlingen, groepen leerlingen en het onderwijs op school van groep 1 tot en met groep 8 te volgen en te evalueren. De toetsen Spelling 3.0 groep 5 zijn een onderdeel van het Cito Volgstelsel primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 5 van het primair onderwijs. Onderstaande beschrijving is gebaseerd op de Handleiding.

Meetpretentie

De toetsen in de toets pakketten Spelling 3.0 van het Cito Volgstelsel primair en speciaal onderwijs zijn bedoeld om vast te stellen hoe goed een leerling kan spellen en hoe de spellingvaardigheid van de leerling zich in de loop van de jaren ontwikkelt.

Het vaststellen van de spellingvaardigheid gebeurt door de leerling woorden te laten opschrijven (actieve spelling). De spellingregels zelf worden niet expliciet bevestigd. De leerling laat indirect zien dat hij of zij de spellingregels beheerst door de gevraagde woorden correct te schrijven.

Doelgroep

De toetsen Spelling 3.0 groep 5 zijn bedoeld voor leerlingen in groep 5 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 5 en voor leerlingen met een ontwikkelingsachterstand en/of extra onderwijsbehoeften. Voor deze groepen speciale leerlingen zijn geen afzonderlijke normen vastgesteld en de toetsresultaten van deze leerlingen worden geïnterpreteerd met behulp van de gemiddelde vaardigheidsscores voor leerlingen uit het regulier onderwijs.

Gebruiksdoel en functie

Doel van de toetsen Spelling 3.0 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van de spelling, voor groep 5 specifiek het schrijven van niet-klankzuivere woorden met twee of meer lettergrepen. De toetsen maken het mogelijk om:

- De vaardigheid spelling van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) te beoordelen via een vergelijking van de behaalde scores met de scores van een landelijke referentiegroep oftewel niveaubepaling.
- De ontwikkeling van de vaardigheid spelling van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) door de leerjaren heen te volgen oftewel progressiebepaling.

Inhoudelijke theoretische inkadering:

De inhoud van de toetsen Spelling 3.0 is gebaseerd op het domein Begrippenlijst en Taalverzorging, en dan het onderdeel Taalverzorging, beschreven in het Referentiekader

Taal en Rekenen. De toetsen sluiten aan bij de indeling die is gehanteerd in het Referentiekader Taal. In de publicatie 'Leerstoflijnen begrippenlijst en taalverzorging beschreven' van de SLO, is aangegeven hoe de opbouw van de leerstoflijnen eruit kan zien voor de verschillende groepen. Voor de inhoud van de toetsen zijn deze uitwerkingen bepalend geweest voor de theoretische basis als voor de indeling van het categorieënoverzicht. Verfijning van de categorieën is mede gebaseerd op basis van analyse van methoden (taal) die veel gebruikt worden in het basisonderwijs. De koppeling die gemaakt wordt met het Referentiekader levert voor scholen een meerwaarde. Zeker wanneer de toetsen voor de bovenbouw gereed zijn en een directe koppeling met de referentieniveaus gemaakt kan worden, zodat inzicht wordt verkregen over wat de niveaus 1F en 1S voor spelling inhouden en hoe leerlingen zich met hun toetsresultaten verhouden tot de referentieniveaus.

Inhoud van het toetspakket

Het toetspakket Spelling 3.0 groep 5 bestaat uit de volgende documenten:

- Handleiding, deze bevat informatie over:
 - de afname van de toets (hfdst. 2),
 - nakijken en verwerken van toetsgegevens (hfdst. 3),
 - interpretatie van de toetsresultaten op leerling- en groepsniveaus (hfdst 4),
 - interpretatie van toetsresultaten op schoolniveau (hfdst 5),
 - theoretisch kader en achtergronden van de toets (hfdst 6),
 - communiceren over toetsresultaten met leerling en ouders (hfdst 7),
 - achtergrondinformatie en veel gestelde vragen (hfdst 8) en
 - enkele bijlagen
- Vier toetsen:
 - Toets E4M5 (makkelijke variant van de toets M5)
 - Toets M5 (Medio groep 5)
 - Toets E5 (Eind groep 5)
 - Toets M5E5 (makkelijke variant van de toets E5)
- Afnamekaarten met aanwijzingen voor de papieren afname van de toetsen
- Nakijkaarten
- Antwoordbladen
- Tabellen voor de vier toetsen voor het bepalen van de vaardigheidsscore en – niveau
- Wetenschappelijke verantwoording.

2. Beoordeling van de kwaliteitsaspecten

De beoordeling vindt plaats volgens het 'Beoordelingskader voor de psychometrische aspecten van (reeksen van) toetsen uit leerlingvolgsystemen (LOVS)', zoals opgesteld door de Expertgroep Toetsen PO. De Expertgroep Toetsen PO wordt gevormd door Prof. Dr. Cees Van der Vleuten (voorzitter), Prof. dr. Cees Glas (psychometrisch expert), Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en mevrouw Pauly N. Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

Sinds 2013/2014 wordt door Cito een nieuwe werkwijze voor het normeren van leerlingvolgsystemen toegepast. Een belangrijk aspect van die werkwijze is om de normeringssteekproef (ongeveer 50 procent) te baseren op gegevens uit het zogenaamde embedded field onderzoek en (ongeveer 50 procent) op gegevens uit Cito dataretour. Bij de normering van de derde generatie LVS-toetsen is rekening gehouden met de variabelen regio, urbanisatiegraad, schooltype (stratum) en sekse, waarbij de eerste drie variabelen op schoolniveau zijn gedefinieerd en de laatste variabele op leerlingniveau. Op pagina 46-47 van de wetenschappelijke verantwoording wordt een algoritme beschreven dat een representatieve steekproef moet garanderen. Niettemin werd er naast die garantie een controle op de representativiteit uitgevoerd door de populatieverdelingen uit gegevens van DUO en/of CBS te vergelijken met de steekproefverdelingen. De controle bevestigde dat de normeringssteekproeven een zeer goede afspiegeling van de populatie zijn met effectgroottes ver onder de 0.10. Ze zijn met 0.088 nog het hoogst voor de variabele schooltype bij het afnamemoment E5. Om deze reden is statistische weging van de resultaten van de normeringssteekproef niet nodig. Ook bleek de gemiddelde score op de Cito Eindtoets Basisonderwijs, en specifiek de score voor het toetsonderdeel spelling, voor de scholen in het normeringsonderzoek niet significant af te wijken van het populatiegemiddelde voor de Eindtoets. Dit is voor meerdere jaren onderzocht.

Voor de deelnemende scholen aan het normeringsonderzoek is nagegaan of zij als groep afwijken van wat men voor de totale populatie van scholen zou mogen verwachten. Niet alleen de gemiddelde score op de Cito Eindtoets Basisonderwijs, maar ook specifiek de score voor spelling bleek voor de scholen in het normeringsonderzoek niet af te wijken van het populatiegemiddelde voor de Eindtoets. Dit is voor meerdere jaren onderzocht.

In tabel 4.4 van de wetenschappelijke verantwoording worden de aantallen leerlingen gegeven die meegenomen zijn bij de normering. Voor M5 betreft dat 3602 leerlingen (1609 leerlingen uit de steekproef en 1993 leerlingen via dataretour) afkomstig van 160 scholen en voor E5 3285 leerlingen (1558 leerlingen uit de steekproef en 1727 leerlingen via dataretour) afkomstig van 158 scholen.

Conclusie:

De representativiteit van de steekproef is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. Op aspect S1.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor de toetsen Spelling groep 5 heeft kalibratieonderzoek plaatsgevonden tijdens het proefonderzoek en het normeringsonderzoek. Tijdens het proefonderzoek zijn in 2013 halverwege en aan het einde van leerjaar 5 opgaven voorgelegd aan leerlingen. Daarbij zijn op het afnamemoment medio groep 225 nieuwe opgaven geproeftoetst en op het afnamemoment eind groep 5 250 opgaven. Elke deelnemende school maakte op het medio-moment één taak met 25 nieuwe opgaven en de LVS-toets Spelling van de tweede

generatie. Op het eind-moment maakten leerlingen naast de LVS-toets Spelling van de tweede generatie één nieuwe taak met 30 opgaven. In totaal waren de nieuwe opgaven op beide afnamemomenten verdeeld over 12 boekjes. Elk boekje werd door ongeveer 150 leerlingen gemaakt. Door de nieuwe opgaven in één of twee boekjes op te nemen, werd ervoor gezorgd dat elke opgave door minimaal 200 leerlingen werd gemaakt.

Het proefonderzoek had tot doel de kwaliteit van de items te onderzoeken. Resultaten over passing van de items en de passing van het gebruikte itemresponstheoriemodel worden uitvoerig beschreven bij de resultaten van het normeringsonderzoek in paragraaf 4.2.3 'Toetsing van het IRT-model' in de wetenschappelijke verantwoording. Bij de analyses van de antwoorden van de leerlingen op de opgaven is nagegaan of de verschillende onderdelen een beroep doen op hetzelfde complex aan vaardigheden. Dat bleek (voor de meeste opgaven) het geval te zijn; opgaven die niet voldeden vielen af. Op basis van inhoudelijke en psychometrische criteria uit de klassieke testtheorie (p-waarden tussen 0.40 en 0.90 en rir-waarden > 0.20) werden 50 opgaven voor de toets M5, 50 opgaven voor de toets E4M5, 50 opgaven voor de toets M5E5, en 50 opgaven voor de toets E5 geselecteerd.

Hoofdstuk 4 van de wetenschappelijke verantwoording betreft de kalibratie en normering en presenteert een zogenaamd macro-design (zie Figuur 4.1) dat begint met de afname van toets M3 en M4 in januari 2013 en eindigt met de afname van toets M8 die gepland staat voor januari 2017. Uit figuur 4.1 in de verantwoording kan men afleiden dat er sprake is van een ankeritemdesign, zodat de prestaties van leerlingen over alle leerjaren heen kunnen worden afgebeeld op een gemeenschappelijke vaardigheidsschaal. Ook de kalibratie van leerjaar 5 is gebaseerd op een ankeritemdesign en wel een zogenaamd 'embedded field design' dat bedoeld is om bias in de normen als gevolg van de low stakes afnamesituatie te vermijden door de afnamesituatie tijdens het normeringsonderzoek zoveel mogelijk te laten lijken op de afnamesituaties na uitgave. Dit laatste hield in dat nieuw ontwikkelde items voor de derde generatie van het Cito Volgsysteem meeliepen in de al bestaande en op scholen toegepaste toetscyclus. In totaal lieten scholen hun leerlingen drie taken maken: één taak uit de tweede generatie van de LVS-toetsen en twee taken met nieuw materiaal voor de derde generatie. Deze drie taken zaten samen in één boekje en telden alle drie mee voor het resultaat van de leerling. In figuur 4.3 en 4.4 in de verantwoording staan de embedded field designs voor M5 en E5. M5 bestaat uit zes toetsversies/boekjes en E5 eveneens uit zeven toetsversies/boekjes. Boekje 1 van M5 bestaat uit items M5 LVS 2e generatie, M5E5 deel 1 en M5E5 deel 2, boekje 2 bestaat uit items M5 LVS 2e generatie, E4M5 deel 1 en M5E5 deel 2, boekje 3 bestaat uit items van M5LVS 2e generatie, E4M5 deel 2 en M5E5 deel 1, boekje 4 bestaat uit items van M5 LVS 2e generatie, E4M5 deel 2 en M5 deel 1, boekje 5 bestaat uit items van M5LVS 2e generatie, E4M5 deel 1 en M5 deel 2 en boekje 6 bestaat uit M5 LVS 2e generatie, M5 deel 1 en M5 deel 2. Het voorgaande geeft nogmaals aan dat er sprake is van een ankertoetsdesign met grote ankers. Elk boekje bestond uit 85 opgaven die verdeeld waren over drie taken. De 205 verschillende items uit het normeringsonderzoek M5 zijn door 1791 leerlingen gemaakt. De 25 items uit de bestaande taak Spelling werden door alle 1791 leerlingen gemaakt en de overige items werden door gemiddeld 597 leerlingen gemaakt. De 200 verschillende items uit het normeringsonderzoek E5 zijn in 2014 door 1635 leerlingen gemaakt. De 25 items uit de bestaande taak Spelling van de tweede generatie werden allen door de 1635 leerlingen uit het 2014 onderzoek gemaakt en de overige items door gemiddeld 467 leerlingen.

Uit het kalibratieonderzoek blijkt dat voor de vier toetsen zowel de items als de gehele toetsen passen bij het itemresponstheoriemodel OPLM. Het voorgaande betekent dat er sprake is van een eendimensionale vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Conclusie:

Het onvolledige dataverzamelingsdesign is adequaat. Op aspect S1.2 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

In tabel 4.9 van de wetenschappelijke verantwoording wordt voor M5 en E5 de normtabel met relatieve normen op leerlingniveau gepresenteerd. Voor aantallen leerlingen en scholen zie S1.1. Voor beide afnamemomenten worden vaardigheidsverdelingen gepresenteerd, d.w.z. gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75, P80, P90. Met behulp van deze percentielen kunnen de twee niveau-indelingen (A t/m E en I t/m V) opgesteld worden.

In tabel 4.11 van de wetenschappelijke verantwoording wordt voor M5 en E5 de normtabel met relatieve normen op schoolniveau gepresenteerd. Voor beide toetsen/afnamemomenten worden vaardigheidsverdelingen gepresenteerd, d.w.z. gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75, P80, P90.

Conclusie:

De normgroepen zijn groot genoeg (zie S1.1). Op aspect N1.2.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit van de steekproeven werd in S1.1 en 2.1 besproken en daar werd geconstateerd dat de steekproeven representatief waren voor regio, urbanisatiegraad, schooltype en sekse.

Voor Spelling 3.0 Groep 5 geldt dat de normen geldig zijn tot en met 2024. Daarnaast monitort Cito periodiek de normering. Jaarlijks wordt aan de hand van representatieve afnamedata nagegaan of er systematisch verschuivingen in het prestatieniveau plaatsvinden. Indien nodig wordt de normering aangepast.

Conclusie:

Op aspect N1.2.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

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

Bevindingen:

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

Conclusie:

De betrouwbaarheidsgegevens worden correct berekend. Op aspect B1.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor E4M5, M5, M5E5 en E5 zijn drie betrouwbaarheidsgegevens berekend: standaardmeetfout, MAcc en een gesimuleerde test-hertest betrouwbaarheidscoëfficiënt. Voor E4M5 zijn de gegevens gelijk aan 2,435, 0,92 en 0,92, voor M5 gelijk aan 2,790, 0,92 en 0,92, voor M5E5 gelijk aan 2,542, 0,93 en 0,93 en voor E5 gelijk aan 2,792, 0,92 en 0,92. De auteurs van de wetenschappelijke verantwoording verwijzen naar het beoordelingssysteem van de COTAN waar voor tests die geen zware consequenties voor leerlingen hebben, zoals de toetsen Spelling 3.0 groep 5, een betrouwbaarheidscoëfficiënt van meer dan 0,80 als 'goed' aangemerkt wordt.

Naast klassieke betrouwbaarheidscoëfficiënten is ook de lokale betrouwbaarheid en de meetnauwkeurigheid onderzocht. De betekenis van de meetnauwkeurigheid voor de beslissingen die met de toets genomen worden, is af te leiden uit de betrouwbaarheidstabellen van twee niveauverdelingen (I t/m V en A t/m E) voor E4M5, M5, M5E5 en E5. Tabel 5.6 (p. 58) van de wetenschappelijke verantwoording ~~laten~~ laat zien dat bij toets M5 de leerlingen betrouwbaarder geassocieerd worden dan bij de drie andere toetsen. In het algemeen worden de leerlingen met de hoogste en de laagste niveau-indicatie voldoende betrouwbaar geassocieerd. Overwogen zou moeten worden om minder dan vijf niveau-indicaties te gebruiken, zodat de kans op classificatiefouten wordt verlaagd.

De informatie uit de betrouwbaarheidstabellen worden samengevat in twee indices: de plus/minus 1 niveau-index en de marginal classification accuracy. Uit de indices blijkt dat gegeven het gebruiksdoel van de toetsen de classificaties voldoende betrouwbaar zijn, d.w.z. dat gebruikers rekening dienen te houden dat 95% van de leerlingen in een scoregroep in werkelijkheid ook in die scoregroep moet scoren en met maximaal 1 niveau verschil (i.e., óf één scoregroep daarboven óf één scoregroep daaronder).

Conclusie:

De betrouwbaarheid van de toetsen Spelling 3.0 groep 5 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft.

Op basis van het voorgaande wordt op aspect B1.2 aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

Validiteit

V1. Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items)

Bevindingen:

De toets Spelling toetst zoals gezegd de actieve spellingsvaardigheid van leerlingen in groep 5. De standaard wijze van afname van de toetsen Spelling groep 5 is dat de leerkracht twee maal, op verschillende dagdelen, een zinsdictee (25 opgaven per keer, 50 opgaven in totaal) afneemt. Daarbij leest de leerkracht de zin voor, en geeft de instructie 'Schrijf op: < woord toetsitem>'. Er is ook een digitale toetsafname mogelijk, waarbij de pc de opgaven voorleest en leerlingen het antwoord intikken.

Een suggestie voor nog meer uniformiteit in afname kan zijn dat er een cd wordt geleverd waarop het zinsdictee ingesproken is, vergelijkbaar met de afname bij de Cito toetsen Luistervaardigheid. Ook al wordt de leerkracht in de handleiding gewezen op de noodzaak niet overdreven te articuleren of anderszins onbedoeld 'hints' te geven, een cd is écht voor iedereen hetzelfde. In een aanvullende reactie geeft Cito B.V. aan dat additioneel onderzoek gedaan is bij leerkrachten naar het gebruik van een cd voor de afname van de toets. Hieruit blijkt dat de leerkrachten geen voorstanders zijn voor het gebruik van een cd. Dit heeft te maken met de bekendheid van de stem van de leerkracht bij de leerlingen en de tongval van de leerkrachten. De vraag is of leerlingen woorden goed herkennen wanneer deze volgens ABN worden uitgesproken. Op grond van deze gegevens is er gekozen om af te zien van het gebruik van de cd.

Verder onderzoek naar het effect van het gebruik van de CD is nodig om de keuze verder te onderbouwen.

Er wordt uitgebreid aandacht besteed aan het onderbouwen van de relevantie van de gekozen items als vertegenwoordigers van spellingcategorieën en moeilijkheidsgraden, in de context van groep 5, en van de indeling die ten grondslag ligt aan het Referentiekader Taal.

De inhoud van de items geeft geen aanleiding tot kanttekeningen over relevantie of efficiëntie ervan.

Merk op dat de beoordeling van dit aspect zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat verricht is inzake validiteit en dan specifiek begripsvaliditeit.

In de wetenschappelijke verantwoording wordt verslag gedaan van onderzoek naar verschillende aspecten van begripsvaliditeit: unidimensionaliteit, itemkwaliteit, itembias, convergente en divergente validiteit en verschillen tussen relevante subgroepen. Hiervoor

is verwezen naar het kalibratieonderzoek van de toetsen waarvan de conclusie is dat het aannemelijk is dat er sprake is van unidimensionaliteit en dat de items één latente trek of construct meten die we, gezien ook de uitkomsten van het hierna genoemde onderzoek, kunnen aanduiden als 'spellingsvaardigheid'. Hiermee is dus voldaan aan de noodzakelijke voorwaarde voor begripsvaliditeit. Uit het onderzoek naar itemkwaliteit bleek verder dat de gemiddelde moeilijkheidsgraad van de toetsen op het vooraf gewenste niveau lag (tussen 0,36 en 0,95) en dat de moeilijkheidsgraad van de items een goede spreiding had.

De samenhang tussen de afzonderlijke items met de totaalscore bleek hoog te zijn, nl. voor alle vier de toetsen een gemiddelde rit-waarde van 0,42 of hoger (voor de toets E4M5 lag de rit-waarde van één item onder 0,20). Onderzoek naar de samenhang tussen de toetsen Spelling met andere LVS toetsen zoals Rekenen-Wiskunde M5, Begrijpend lezen M5, Technisch lezen M5 en Woordenschat M5 lieten correlaties zoals verwacht zien, d.w.z. de convergente en divergente validiteit was voldoende.

Onderzoek naar Differentieel Item Functioneren (DIF) met betrekking tot jongens en meisjes liet zien dat er nauwelijks sprake was van DIF naar sekse. Slechts bij één item (in de toets E5) was er sprake van differentieel functioneren met betrekking tot sekse.

Uit onderzoek naar verschillen tussen relevante subgroepen bleken resultaten aan te sluiten bij de verwachtingen die op grond van theoretische inzichten en eerder onderzoek konden worden geformuleerd: jongere/versnelde leerlingen het iets beter doen dan reguliere leerlingen en dat die het weer iets beter doen dan oudere/vertraagde leerlingen; meisjes iets hoger scoren dan jongens; leerlingen met een leerlinggewicht van 0,3 en 1,2 op moment M5 en E5 lager scoren dan leerlingen met leerlinggewicht 0.

Conclusie:

Op aspect V1.1 wordt aan de toetsen Spelling 3.0 groep 5 op dit aspect het oordeel '**voldoende**' toegekend.

Het volg-aspect

VA1.1. Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Bevindingen:

De resultaten van het kalibratieonderzoek (zie ook S1.2) laten zien dat de items van de verschillende toetsen Spelling 3.0 groep 5 op een eendimensionale vaardigheidsschaal afgebeeld kunnen worden en dat aan de hand van de door de leerling behaalde vaardigheidsscores op de onderscheiden toetsen diens groei adequaat gemeten kan worden.

Conclusie:

Op aspect VA1.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

VA1.2. Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden? Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Bevindingen:

In de Handleiding wordt aan de hand van een leerlingrapport de interpretatie van groei op een duidelijke manier beschreven. De 67 procent betrouwbaarheidsintervallen van de vaardigheidsscores worden zowel op het leerlingrapport als in afzonderlijke tabellen vermeld.

Groei wordt zowel met vergelijkingen tussen verschillende leerlingen als met vergelijkingen op andere tijdstippen van dezelfde leerlingen weergegeven. Zowel in de wetenschappelijke verantwoording en de handleiding wordt het volgen van groei van leerlingen adequaat toegelicht.

Conclusie:

Op aspect VA1.2 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

Inzicht in leervorderingen

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

Bevindingen:

Via de portal van Cito B.V. kan gebruik worden gemaakt van rapportage-/registratieformulieren voor een leerlingrapport, groepsrapport, groepsoverzicht (overzicht van één groep leerlingen tijdens hun schoolperiode) en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken).

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

Daarnaast wordt de docent gewezen op misverstanden die zich bij de interpretatie van de niveau-indelingen bij de ouders kunnen voordoen. Ook moeten zij aan ouders het verschil tussen methode-onafhankelijke en methodegebonden toetsen duidelijk maken en erop wijzen dat deze toetsen leerlingen anders (kunnen) beoordelen. In hoofdstuk 8 worden ook veel gestelde vragen behandeld die weliswaar voor de docenten bestemd zijn maar zij kunnen met die informatie bijvoorbeeld via tienminutengesprekken ook de ouders beter voorlichten.

De wijze waarop de registratieformulieren zijn vormgegeven en de uitleg die voor docenten beschikbaar is, geven het niveau en de groei van de individuele leerling weer.

Conclusie:

Op aspect I1.1 wordt aan de toetsen Spelling 3.0 groep 5 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

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

4. Literatuurlijst

- Tomesen, M., Wouda, J. & Horsels, L. (2016). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 groep 5*. Arnhem: Cito B.V.
- Cito (2015). *Cito Volgsysteem primair en speciaal onderwijs. Spelling 3.0 groep 5*. Arnhem: Cito