

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 verantwoording van de inhoud en de psychometrische kwaliteit van de toets..

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 7 zijn een onderdeel van het Cito Volgstelsel primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 7 van het primair en speciaal onderwijs.

De toetsen voor groep 3 tot en met groep 8 worden vanaf het schooljaar 2014/2015 geleidelijk vervangen door de nieuwe toetsen Spelling 3.0. De toetsen voor groep 7 vormen dan samen met de LVS toetsen Spelling 3.0 groepen 3, 4, 5, 6 en groep 8 een systeem om vast te stellen hoe goed leerlingen kunnen spellen en hoe hun vaardigheid zich in de basisschoolperiode ontwikkelt. 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 bevraagd. 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 7 zijn bedoeld voor leerlingen in groep 7 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 7 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 7 specifiek het schrijven van complexe niet-klankzuivere woorden en het spellen van werkwoordsvormen. 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.

Inhoud van het toetspakket

Het toetspakket Spelling 3.0 groep 7 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),
 - algemene aandachtspunten voor het schoolplan (hfdst 5),
 - inhoudsverantwoording (hfdst 6),
 - communiceren over toetsresultaten met leerling en ouders (hfdst 7),
 - achtergrondinformatie en veel gestelde vragen (hfdst 8) en
 - enkele bijlagen
- Vier toetsen:
 - Toets M7 (niet-werkwoorden)
 - Toets E7 (niet-werkwoorden)
 - Toets M7 (werkwoorden)
 - Toets E7 (werkwoorden)
- 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 K. 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 leerlingvolg-systemen toegepast. Een belangrijk aspect van die werkwijze is om de normeringssteekproef (ongeveer 50 procent) te baseren op gegevens verkregen via een embedded field onderzoek en (ongeveer 50 procent) op gegevens die afkomstig zijn van de tweede generatie LVS-toetsen via 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 52-54 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 van gegevens uit DUO te vergelijken met de steekproefverdelingen van M7 en E7. In de tabellen 4.5 tot en met 4.8b (pag. 56-58) worden de resultaten van de representativiteitsanalyses gepresenteerd. De effectgroottes zijn klein tot zeer klein (ver onder of rond de .10) en de conclusie is dan ook dat de normeringssteekproeven een zeer goede afspiegeling van de populatieverdelingen zijn.

In tabel 4.4a van de wetenschappelijke verantwoording worden de aantallen leerlingen gegeven die meegenomen zijn bij de normering Spelling niet-werkwoorden. Voor M7 betreft dat 2.342 leerlingen (1.024 leerlingen uit de steekproef en 1.318 leerlingen via dataretour) afkomstig van 123 scholen. Voor E7 betreft dat 2.579 leerlingen (1.202 leerlingen uit de steekproef en 1.377 leerlingen via dataretour) afkomstig van 137 scholen.

In tabel 4.4b van de wetenschappelijke verantwoording worden de aantallen leerlingen gegeven die meegenomen zijn bij de normering Spelling werkwoorden. Voor M7 betreft dat 645 leerlingen uit de steekproef afkomstig van 32 scholen. Voor deze toets en dit afnamemoment M7 kon geen beroep worden gedaan op dataretour. Uit een eerste normeringssteekproef bleken de achtergrondvariabelen regio en stratum niet representatief te zijn. Hierom is in januari 2018 een aanvullend onderzoek uitgevoerd, waarmee regio en stratum alsnog representatief bleken te zijn. Voor E7 betreft het 2.418 leerlingen (1.111 leerlingen uit de steekproef en 1.307 leerlingen via dataretour) afkomstig van 119 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 7 het oordeel 'voldoende' toegekend.

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

Bevindingen:

Bij de ontwikkeling van de toetsen Spelling 3.0 groep 7 is geprobeerd om door middel van embedded field designs bias in de normen te vermijden door de afnamesituatie waarin de toets wordt afgenomen tijdens het normeringsonderzoek zoveel mogelijk te laten lijken op de afnamesituatie na uitgave van de toets. Het normeringsonderzoek gaat uit van een onvolledig dataverzamelingsdesign. 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, met uitzondering van Spelling werkwoorden M7. Voor dit afnamemoment bestond namelijk geen tweede generatie LVS-toets (zie figuur 4.6) en kon er dus ook geen sprake zijn van een embedded field onderzoek. Voor dit afnamemoment maakten leerlingen dan ook alleen twee taken met nieuw materiaal voor de derde generatie.

In figuur 4.4 en 4.5 worden de embedded field designs weergegeven voor respectievelijk M7 en E7 niet-werkwoorden. In het design zien we dat er voor niet-werkwoorden vijf toetsversies zijn afgenomen voor M7 en zes voor E7. Toetsversie 1 van M7 bestaat uit een taak M7 LVS 2e generatie, een taak M7 deel 1 en een taak M7 deel 2, etc. Behalve de toetsen M7 en E7 worden ook een taak van de tweede generatie, een reservetaak, ankertaken (naar lager en hoger niveau) en bij design E7 ook een taak passieve spelling uit de toets Taalverzorging afgenomen. Een enkele toetsversie werd bij tussen de 96 en 276 leerlingen afgenomen.

In figuur 4.6 en 4.7 worden de (embedded field) designs weergegeven voor respectievelijk M7 en E7 werkwoorden. In het design zien we dat er voor werkwoorden drie toetsversies zijn afgenomen voor M7 en vijf voor E7. Toetsversie 1 van M7 bestaat uit een taak M7 deel 1 en een taak M7 deel 2, etc. Behalve de toetsen M7 en E7 worden ook een reservetaak afgenomen en bij E7 tevens een taak van de tweede generatie, een ankertaak MS en een taak passieve spelling uit de toets Taalverzorging. Een enkele toetsversie werd bij tussen de 153 en 271 leerlingen afgenomen.

In het normeringsonderzoek M7 niet-werkwoorden in 2016 zijn 175 verschillende items voorgelegd aan 1.180 leerlingen van groep 7. Hiervan hebben 1.170 leerlingen de volledige taken gemaakt en die leerlingen zijn verdeeld over vijf boekjes. Elk boekje bestond uit 85 – 90 opgaven, verdeeld over drie taken. De 30 opgaven uit de bestaande taak Spelling 2e generatie werd door alle 1.170 leerlingen uit het normeringsonderzoek gemaakt. De overige opgaven kwamen steeds in twee boekjes voor en werden gemiddeld door 468 leerlingen gemaakt.

In het normeringsonderzoek E7 niet-werkwoorden in 2016 zijn 191 verschillende items voorgelegd aan 1.311 leerlingen van groep 7. Hiervan hebben 1.286 leerlingen de volledige taken gemaakt en die leerlingen zijn verdeeld over zes boekjes. Elk boekje bestond uit 75 – 90 opgaven, verdeeld over drie taken. De 30 opgaven uit de bestaande taak Spelling 2^e generatie werd door alle 1.286 leerlingen uit het normeringsonderzoek gemaakt. De overige opgaven kwamen steeds in twee boekjes voor en werden gemiddeld door 429 leerlingen gemaakt.

In het normeringsonderzoek M7 werkwoorden in 2017 zijn 60 verschillende items voorgelegd aan 544 leerlingen van groep 7. Hiervan hebben 540 leerlingen de volledige taken gemaakt en die leerlingen zijn verdeeld over drie boekjes. Elk boekje bestond uit 40 opgaven, verdeeld over twee taken. De opgaven kwamen steeds in twee boekjes voor en werden gemiddeld door 360 leerlingen gemaakt.

In het normeringsonderzoek E7 werkwoorden in 2016 zijn 148 verschillende items voorgelegd aan 1.317 leerlingen van groep 7. Hiervan hebben 1.305 leerlingen de volledige taken gemaakt en die leerlingen zijn verdeeld over vijf boekjes. Elk boekje bestond uit 70 – 78 opgaven, verdeeld over drie taken. De 25 opgaven uit de bestaande taak Spelling 2^e generatie werd door alle 1.305 leerlingen uit het normeringsonderzoek gemaakt. De overige opgaven kwamen steeds in twee boekjes voor en werden gemiddeld door 518 leerlingen gemaakt.

Uit het kalibratieonderzoek blijkt dat voor de vier toetsen zowel de items, gegeven de M- en S-toetsen, als de gehele toetsen, gegeven de R1c-toetsen, passen bij het itemresponsstheoriemodel OPLM. Dit betekent dat er sprake is van een eendimensionale vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden. Op grond van de gegevens uit de kalibratie zijn de definitieve selecties van items gemaakt voor de uitgaven van de toetsen Spelling 3.0 M7 / E7 niet-werkwoorden en Spelling 3.0 M7 / E7 werkwoorden.

Conclusie:

Het onvolledige dataverzamelingsdesign is adequaat. Op aspect S1.2 wordt aan de toetsen Spelling 3.0 groep 7 het oordeel 'voldoende' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

In de tabellen 4.9a en 4.9b van de wetenschappelijke verantwoording wordt voor M7 en E7 niet-werkwoorden respectievelijk werkwoorden de normtabel met relatieve normen op leerlingniveau gepresenteerd. Voor aantallen leerlingen en scholen zie S1.1. Voor alle vier de toetsen/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 de tabellen 4.11a en 4.11b van de wetenschappelijke verantwoording wordt voor M7 en E7 niet-werkwoorden respectievelijk werkwoorden de normtabel met relatieve normen op schoolniveau gepresenteerd. Voor alle vier de 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 N.1.2.1. wordt aan de toetsen Spelling 3.0 groep 7 het oordeel 'voldoende' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit van de steekproeven werd in S1.1 en S1.2 besproken en daar werd geconstateerd dat de steekproeven representatief waren voor regio, urbanisatiegraad, schooltype en sekse. Voor Spelling 3.0 groep 7 (niet-werkwoorden en werkwoorden) geldt dat de normen geldig zijn tot en met 2026. 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 N1.2.2. wordt aan de toetsen Spelling 3.0 groep 7 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 succesvol geschaald zijn met OPLM, een betrouwbaarheidscoëfficiënt, de MAC ('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 7 het oordeel 'voldoende' toegekend.

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

Bevindingen:

Voor elk van de vier toetsen M7 en E7 niet-werkwoorden en werkwoorden zijn drie betrouwbaarheidsgegevens berekend: standaardmeetfout, M_{Acc} en een gesimuleerde test-hertest betrouwbaarheidscoëfficiënt. Voor M7 niet-werkwoorden zijn de gegevens gelijk aan 2.684, 0.93 en 0.93, voor E7 niet-werkwoorden gelijk aan 2.790, 0.92 en 0.92, Voor M7 werkwoorden gelijk aan 2.540, 0.85 en 0.85 en voor E7 werkwoorden gelijk aan 2.822, 0.89 en 0.89. 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 7, 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 M7 en E7. In het algemeen worden de leerlingen met de hoogste en de laagste niveau-indicatie voldoende betrouwbaar geclassificeerd.

De informatie uit de betrouwbaarheidstabellen kan worden samengevat met 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, maar dat er wel rekening mee moet worden gehouden dat er altijd sprake zal zijn van misclassificaties ter grootte van maximaal 1 niveau.

Conclusie:

De betrouwbaarheid van de toetsen Spelling 3.0 groep 7 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 7 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 items in de toetsen zijn geheel te plaatsen in hetgeen hierboven onder *Inhoudelijke theoretische inkadering* genoemd wordt. Met de toetsen wordt inderdaad actieve beheersing van de spelling van zelfstandig naamwoorden en werkwoorden getoetst. De toevoeging van werkwoordspelling op meetmoment M7 is conform wensen in het veld. De

gekozen doelwoorden en de zinnen waarin zij gepresenteerd worden, zijn bijna zonder uitzondering goed.

Een aantal kleine kanttekeningen:

- Er wordt opgemerkt dat de toetsen beginnen en eindigen met een relatief eenvoudig item, om voor de leerlingen een goede start en prettige afsluiting van de toets te bewerkstelligen. Dat lijkt niet altijd geslaagd in het materiaal voor de niet-werkwoorden: zeerob, apotheek, viaduct, nectar. Echter uit de p-waarden (0.84, 0.82, 0.83 en 0.88) blijkt dat deze woorden als relatief eenvoudig kunnen worden aangemerkt.
- Misschien kan de zinvolle opmerking wat meer herhaald worden (misschien ook op de afnamekaart voor de leerkracht?) dat eigennamen in de toetszinnen, die ook in de klas voorkomen, vervangen mogen worden door de leerkracht.
- Bij het materiaal van de toetsen *niet_werkwoorden* kan van een aantal items gesteld worden dat ze, ondanks *voldoende* vanuit de psychometrische data gezien, toch inhoudelijk wat geforceerd overkomen voor leerlingen in groep 7:
 - M7, T1, item 11 "de kandidaat voor de baan wacht nerveus op de uitslag."
 - M7, T2, item 6 "Myra betreft een klein woninkje."
 - M7, T2, item 21: "In de verte zie ik een vliegend object."
- Bij het materiaal van de toetsen *werkwoorden*: geen detailopmerkingen.

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.

Uit het kalibratieonderzoek van de toetsen blijkt 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 'spellingvaardigheid'. Hiermee is voldaan aan de noodzakelijke voorwaarde voor begripsvaliditeit.

Uit het onderzoek naar itemkwaliteit bleek dat de moeilijkheidsgraad van de items op het vooraf gewenste niveau lag, namelijk tussen 0.35 en 0.90 voor niet-werkwoorden en tussen 0.41 en 0.92 voor werkwoorden. De gemiddelde Rit- waarde is voor alle vier de toetsen 0.38 of hoger waarbij een Rit-waarde van hoger dan 0.30 door de Cotaan als goed wordt gekwalificeerd. De kalibratieresultaten op itemniveau en de analyses met betrekking tot de nauwkeurigheid van de itemparameterschattingen bevestigen voorgaande positieve bevindingen.

Onderzoek naar de samenhang tussen de toetsen Spelling met andere LVS toetsen zoals Technisch lezen M7, Begrijpend lezen M7, Woordenschat M7 en Rekenen-Wiskunde M7 liet correlaties zoals verwacht zien, d.w.z. de convergente en divergente validiteit was voldoende. De resultaten geven aan dat er gemeten wordt wat men met de toetsen Spelling 3.0 groep 7 beoogt te meten, namelijk spellingvaardigheid.

Onderzoek naar Differentieel Item Functioneren (DIF) met betrekking tot jongens en meisjes liet zien dat er geen of nauwelijks sprake was van DIF naar sekse. Slechts bij twee

items bij de toetsen M7 en E7 werkwoorden en bij één item bij de toets M7 niet-werkwoorden was er sprake van differentieel functioneren met betrekking tot sekse.

Uit onderzoek naar verschillen tussen relevante subgroepen bleken de resultaten aan te sluiten bij de verwachtingen die op grond van theoretische inzichten en eerder onderzoek konden worden geformuleerd: jongere/versnelde leerlingen doen het iets beter dan reguliere leerlingen en die doen het weer iets beter dan oudere/vertraagde leerlingen; meisjes scoren iets hoger dan jongens, alleen bij M7 niet-werkwoorden is de gemiddelde score van meisjes gelijk aan die van jongens; leerlingen met een leerlinggewicht van 0,3 en 1,2 scoren op M7 en E7 bij zowel werkwoorden als niet-werkwoorden lager dan leerlingen met leerlinggewicht 0.

Conclusie:

Op aspect V1. wordt aan de toetsen Spelling 3.0 groep 7 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:

Uit de resultaten van het kalibratieonderzoek blijkt dat de items van de verschillende toetsen Spelling 3.0 groep 7 (werkwoorden en niet-werkwoorden) 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 7 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, met name in hoofdstuk 4, 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.

In de Handleiding wordt zowel met vergelijkingen tussen verschillende leerlingen als met vergelijkingen op andere tijdstippen van dezelfde leerlingen het volgen van groei van leerlingen adequaat toegelicht.

Conclusie:

Op aspect VA1.2. wordt aan de toetsen Spelling 3.0 groep 7 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 rapportageformulieren voor een leerlingrapport, groepsrapport, groepsoverzicht (overzicht van één groep leerlingen tijdens hun schoolperiode) en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken).

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.

Conclusie:

Op aspect I1. wordt aan de toetsen Spelling 3.0 groep 7 het oordeel 'voldoende' toegekend.

3. Verzamelstaat

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

4. Literatuurlijst

- Tomesen, M., Wouda, J., Kramer, I. en Horsels, L. (2018). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 groep 7*. Arnhem: Cito B.V.
- Cito (2017). *Cito Volgsysteem primair en speciaal onderwijs. Spelling 3.0 groep 7*. Arnhem: Cito