

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 leerkrachtmap veelal letterlijk vermeld. De wetenschappelijke verantwoording heeft betrekking op de uitgangspunten van de toetsconstructie, de normen, de betrouwbaarheid en meetnauwkeurigheid en de validiteit. De leerkrachtmap 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 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.

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 7 zijn bedoeld voor leerlingen in groep 7 van het primair onderwijs, leerlingen van het speciaal basisonderwijs en leerlingen in het speciaal onderwijs, cluster 1, 2 en 4. De toetsen 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 Spelling3.0 dig. groep 7 bestaat uit de volgende documenten:

- Instructie om op de digitale omgeving van Cito te komen
- Wetenschappelijke verantwoording

2. Beoordeling van de kwaliteitsaspecten

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

De kwaliteit van de dataverzameling

S1 Is de steekproef representatief?

Bevindingen:

Aan het begin van het toetsontwikkelingsproces van de LVS-toetsen Spelling 3.0 groep 7, zijn in 2012 en 2013 opgaven geconstrueerd. Deze opgaven zijn allereerst in papieren versie onderzocht in kalibratie-onderzoeken in 2014 en 2015 en – na opname van de opgaven in de uit te geven 'papieren' toetsen – in normeringsonderzoeken in 2016 en 2017. In 2017 en 2018 hebben vervolgens kalibratieonderzoeken voor de digitale items plaatsgevonden. Dit gebeurde in de vorm van papier-digitaal vergelijkingsonderzoek. De digitale items werden daarmee gekalibreerd op dezelfde schaal als de papieren versie. De Expertgroep Toetsen PO had als oordeel dat voor de normering van de papieren versie van Spelling 3.0 groep 7 een representatieve steekproef is gebruikt. Ook zijn de gebruikte normgroepen groot genoeg en representatief voor de doelpopulatie, zowel op schoolniveau als leerlingniveau.

Conclusie:

Omdat de digitale versie via een gelinkt design gekalibreerd is op dezelfde schaal als de papieren versie en de Expertgroep Toetsen PO aan de steekproef voor het normeren van de papieren versie reeds de beoordeling voldoende heeft gegeven, geldt hetzelfde voor de digitale versie. Daarmee wordt aan aspect S1.1. het oordeel '**voldoende**' toegekend.

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

Bevindingen:

In een papier-digitaal onderzoek is het design onvolledig: in tegenstelling tot een volledig papieren onderzoek overlappen de boekjes slechts gedeeltelijk. De echte link om de boekjes op één schaal te krijgen, verloopt via de papieren uitgave. Alle nieuwe (digitale) taken zijn afgenomen bij leerlingen die ook de papieren Starttaak van LVS Spelling tweede generatie gemaakt hebben. Omdat de spellingvaardigheid van een leerling niet verandert tijdens een toets, kan de vaardigheid op het papieren gedeelte van de toets vergelijkbaar worden geacht met de vaardigheid op het digitale gedeelte van de toets. Hierdoor zijn de twee afnamemethoden verbonden.

De opzet van het papier-digitaal onderzoek voor het 'medio' (M) afnamemoment groep 7 week af, omdat er geen toets Spelling werkwoorden M7 van de tweede generatie was. De data van dit papier-digitaal onderzoek werden gekoppeld aan de data die verzameld werden in het normeringsonderzoek voor de papieren uitgave M7 werkwoorden uit 2017. Om de koppeling te realiseren werden de data die verzameld zijn in het papier-digitaal onderzoek toegevoegd aan de dataset van het genoemde normeringsonderzoek. Op deze manier werden de afnames verbonden.

Conclusie:

Het onvolledige maar 'verbonden' design van het papier-digitaal onderzoek is adequaat. Aan aspect S1.2 wordt het oordeel '**voldoende**' toegekend.

Normering

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

Bevindingen:

Criterium heeft betrekking op absoluut normeren en is hier niet van toepassing.

*Conclusie: **N.V.T.***

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

Bevindingen:

Criterium heeft betrekking op absoluut normeren en is hier niet van toepassing.

*Conclusie: **N.V.T.***

N1.3 Is er voldoende overeenstemming tussen de beoordelaars?

Bevindingen:

Criterium heeft betrekking op absoluut normeren en is hier niet van toepassing.

*Conclusie: **N.V.T.***

N2.1 Zijn de normgroepen groot genoeg?

Bevindingen:

De beoordeling van de papieren versie van de Spelling 3.0 toetsen voor groep 7 is door de Expertgroep als voldoende beoordeeld en de digitale versie is op dezelfde schaal gekalibreerd.

Conclusie:

De steekproeven zijn groot genoeg. Daarmee wordt aan aspect N1.2.1. voor de digitale toetsen Spelling 3.0 voor groep 7 het oordeel '**voldoende**' toegekend.

N2.2 Zijn de normgroepen representatief?

Bevindingen:

De beoordeling van de papieren versie van de Spelling 3.0 toetsen voor groep 7 is door de Expertgroep als voldoende beoordeeld en de digitale versie is op dezelfde schaal gekalibreerd.

Conclusie:

Aan aspect N1.2.2. wordt voor de digitale toetsen Spelling 3.0 voor groep 7 het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen:

Vanwege het onvolledige design is de betrouwbaarheid van de toetsen in klassieke zin niet rechtstreeks te bepalen. Het is echter wel mogelijk om de betrouwbaarheid van iedere toets te schatten door gebruik te maken van het feit dat alle items die zijn opgenomen in de toetsen OPLM-geschaald zijn. Hierbij wordt gebruik gemaakt van het programma OPLAT. In plaats van de klassieke coëfficiënt alpha wordt gebruik gemaakt van de Accuracy of Measurement (MAcc). Deze coëfficiënt is in de psychometrische literatuur beschreven en als correct aangemerkt.

Omdat alle items OPLM-gekalibreerd zijn, is het mogelijk een hertest te simuleren. Daarvoor is een dubbele afname gesimuleerd van 1.000.000 leerlingen, waarbij enerzijds de vaardigheidsverdeling van alle leerlingen, anderzijds alle itemparameters als uitgangspunt is genomen. Steeds is een bepaalde vaardigheid aselekt uit de verdeling genomen en zijn twee bij deze vaardigheid horende afnames gesimuleerd. Uiteindelijk is de correlatie tussen deze 1.000.000 dubbele (virtuele) afnames berekend. Men kan deze simulatie beschouwen als een test-hertestonderzoek. Deze procedure wordt ook in de psychometrische literatuur beschreven en als correct aangemerkt.

Omdat gebruik gemaakt is van bekende en algemeen beschikbare software, kan ervan uitgegaan worden dat de betrouwbaarheidsgegevens correct berekend zijn.

Conclusie:

Aan aspect B1.1 wordt aan de digitale toetsen Spelling 3.0 groep 7 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Er wordt verwezen naar de COTAN criteria voor toetsen voor minder belangrijke beslissingen. De interne consistentie betrouwbaarheid is, volgens deze criteria, voldoende bij een betrouwbaarheidscoëfficiënt tussen 0,70 en 0,80. De verschillende betrouwbaarheidsmaten variëren van 0,89 tot 0,98. Daarmee zijn de betrouwbaarheden van ruim voldoende niveau. Aanvullend hierop is gekeken naar de nauwkeurigheid van de verschillende toetsen. Voor de verschillende scoregroepen is de accuracy 54,3% of hoger. Om de nauwkeurigheidstabellen samen te vatten, wordt gebruik gemaakt van de plus/minus 1 niveau-index en de Marginal Classification Accuracy. Gemiddeld gezien heeft de plus/minus 1 niveau-index, afhankelijk van de toets en de gekozen indeling in scoregroepen, de waarde 0,986 tot 0,999. De Marginal Classification Accuracy loopt uiteen van 65,6 tot 75,1 procent. Daarmee is zowel de overall betrouwbaarheid als nauwkeurigheid van de verschillende scoregroepen voldoende.

Conclusie:

De betrouwbaarheid van de digitale toetsen Spelling 3.0 groep 7 is ruim 'voldoende' volgens de beoordelingscriteria voor de betrouwbaarheid van de COTAN.

Op aspect B.1.2. wordt aan digitale toetsen Spelling 3.0 voor groep 7 het oordeel '**voldoende**' toegekend.

Validiteit

V1 Inhoudvaliditeit: 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 te plaatsen in en te herleiden tot de inhoudelijke theoretische inkadering zoals die door de aanbieder is opgesteld. Met de toetsen wordt actieve beheersing van de spelling van zelfstandig naamwoorden en werkwoorden getoetst. De gekozen doelwoorden en de zinnen waarin zij gepresenteerd worden, zijn bijna zonder uitzondering goed.

De keuze voor en gebruiksmogelijkheden van de digitale toetsversies voor de doelgroepen en meetmomenten in groep 7 wordt grondig toegelicht. Daarnaast is aandacht voor de mogelijkheid tot foutenanalyse, interpreteren en analyseren van resultaten leerling- en groepsniveau, en communicatie met leerling en ouders over de resultaten.

De toets Spelling toetst uiteraard ook in de digitale afname zoals gezegd de actieve spellingsvaardigheid van leerlingen in groep 7. Hierbij leest de pc de opgaven voor, waarna de leerlingen het antwoord intikken.

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 7, en van de indeling die ten grondslag ligt aan het Referentiekader Taal.

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

Conclusie:

Aan aspect V1. wordt voor de digitale toetsen Spelling 3.0 voor groep 7 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor de verantwoording van de validiteit wordt verwezen naar de Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 voor groep 7. Voor dit aspect werd aan de toetsen Spelling 3.0 groep 7 op dit aspect het oordeel 'voldoende' toegekend.

Voor de digitale items zijn eigen moeilijkheids- en discriminatieparameters geschat. Een voorwaarde voor constructvaliditeit is dat het meetmodel past bij de data. Er zijn geen opgaven met DIF ten opzichte van de papieren toetsen opgenomen. Met M-toetsen, S-toetsen en de R1c toets is gekeken of dit model paste. Dat was het geval. Ook werd er geconstateerd dat er nauwelijks tot geen sprake was van differentieel itemfunctioneren (DIF).

Conclusie:

Op aspect V.2. wordt aan de digitale toetsen Spelling 3.0 groep 7 het oordeel '**voldoende**' toegekend.

Het volg-aspect

Va1 Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Bevindingen:

Uit de resultaten van het kalibratieonderzoek blijkt dat de items van de verschillende digitale 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:

Aan aspect VA1. wordt aan de digitale toetsen Spelling 3.0 groep 7 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

De scores van de leerlingen worden op een eendimensionale vaardigheidsschaal weergegeven. De groei kan uitgerekend worden door opeenvolgende scores van elkaar af te trekken. In de wetenschappelijke verantwoording wordt bovendien de standaard meetfout voor de verschillende schaalscores weergegeven. Alhoewel de betrouwbaarheid van de groei niet expliciet wordt gerapporteerd, kan daarmee wel een goede indruk van de betrouwbaarheid gekregen worden.

Conclusie:

Op aspect VA2 wordt aan de digitale toetsen Spelling 3.0 groep 7 het volgende oordeel toegekend: '**voldoende**'.

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

Bevindingen:

In de Handleiding wordt aan de hand van een leerlingrapport de interpretatie van groei op een duidelijke manier beschreven. De 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:

Aan aspect VA3. wordt aan de digitale 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 rapportage- en registratieformulieren voor een leerlingrapport, groepsrapport, groepsoverzicht en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken).

Er wordt aandacht besteed aan hoe met ouders over de toetsresultaten gecommuniceerd kan/moet worden. Met name wordt daarbij gewezen op het leerlingrapport waarin zowel het niveau van de leerling als de progressie van de leerling gepresenteerd worden. Hiervoor is ook een folder ouderinformatie beschikbaar die men via de website van het Cito kan downloaden.

Daarnaast worden docenten 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 methode-gebonden toetsen duidelijk maken en erop wijzen dat deze toetsen leerlingen anders (kunnen) beoordelen. 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.

In de Handleiding 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 digitale toetsen Spelling 3.0 groep 7 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:

De toetsaanbieder geeft gedetailleerde en overzichtelijk informatie over de totstandkoming van de toets als geheel en voor de keuze voor de diverse items afzonderlijk. Op basis van passende bronnen (Onderdeel 'Taalverzorging' uit Referentiekader Taal; Van der Beek & Paus (2011) *Leerstoflijnen begrippenlijst en taalverzorging beschreven: uitwerking van het referentiekader Nederlandse taal voor het domein begrippenlijst en taalverzorging op de basisschool*; inventarisatie van inhoud en relevante taalmethodes) is een valide set set toetsitems geselecteerd.

Conclusie:

Op aspect R1 wordt aan de digitale toetsen Spelling 3.0 groep 7 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

Kwaliteitsaspect	Code	Oordeel
De kwaliteit van de steekproef	S1	voldoende
	S2	voldoende
Normering	N1.1	n.v.t.
	N1.2	n.v.t.
	N1.3	n.v.t.
	N2.1	voldoende
	N2.2	voldoende
Betrouwbaarheid	B1	voldoende
	B2	voldoende
Validiteit	V1	voldoende
	V2	voldoende
Volg-aspect	Va1	voldoende
	Va2	voldoende
	Va3	voldoende
Inzicht in leervorderingen	I1	voldoende
Referentieniveaus	R1	voldoende

4. Literatuurlijst

- Tomesen, M., Wouda, J., Krämer, I. en Horsels, L. (2018). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 groep 7*. Arnhem: Cito B.V.
- Tomesen, M., Wouda, J. en Horsels, L. (2019). *Aanvulling bij de wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 groep 7*. Arnhem: Cito B.V.
- Cito (2017). *Cito Volgstelsel primair en speciaal onderwijs. Spelling 3.0 groep 7*. Arnhem: Cito B.V.