

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de wetenschappelijke verantwoordingen van 2019 en 2020 veelal letterlijk vermeld.

Algemeen

Het Cito Volgsysteem 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 maken deel uit van de derde generatie toetsen van het Cito Volgsysteem primair en speciaal onderwijs en zijn bestemd voor leerlingen in de groepen 3 t/m 8 in het basisonderwijs. Ze zijn ook geschikt voor toepassing in het speciaal basisonderwijs en het speciaal onderwijs cluster 1, cluster 2 (TOS-leerlingen) en cluster 4. De toetsen Spelling 3.0 betreffen papieren en digitale toetsen voor alle leerjaren. Ter beoordeling liggen nu voor: toetsen Spelling 3.0 groep 8 digitaal.

Meetpretentie

De toetsen Spelling 3.0 groep 8 digitaal 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 8 digitaal zijn bestemd voor leerlingen in groep 8 van het basisonderwijs. De toetsen zijn ook geschikt voor leerlingen in het speciaal basisonderwijs en het speciaal onderwijs cluster 1, cluster 2 (TOS-leerlingen) en cluster 4. Voor deze groepen speciale leerlingen zijn geen afzonderlijke normen vastgesteld. De toetsresultaten van deze leerlingen worden geïnterpreteerd met behulp van de gemiddelde vaardigheidsscores voor leerlingen uit het reguliere onderwijs.

Gebruiksdoel en functie

Doel van de toetsen Spelling 3.0 groep 8 digitaal is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van de spelling, voor groep 8 specifiek het spellen van niet-werkwoorden, onveranderlijke woorden en het spellen van werkwoorden. 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 digitaal 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 8 digitaal bestaat uit de volgende documenten:

- Digitale toetspakket, met daarin vier toetsen:
 - Toets B8 (niet-werkwoorden)
 - Toets M8 (niet-werkwoorden)
 - Toets B8 (werkwoorden)
 - Toets M8 (werkwoorden)
- Digitaal toetspakket, met daarin handleiding met informatie over:
 - de afname van de toets
 - nakijken en verwerken van toetsgegevens
 - interpretatie van de toetsresultaten op leerling- en groepsniveaus
 - algemene aandachtspunten voor het schoolplan
 - communiceren over toetsresultaten met leerling en ouders
 - achtergrondinformatie en veel gestelde vragen
 - enkele bijlagen
- Wetenschappelijke verantwoording, deze bevat informatie over
 - Uitgangspunten van de toetsconstructie
 - Beschrijving van de toetsen
 - Kalibratie en normering
 - Betrouwbaarheid en meetnauwkeurigheid
 - Validiteit
 - Enkele bijlagen.

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 samenstelling van de Expertgroep Toetsen PO wordt weergegeven op de website, www.expertgroepoetsenpo.nl.

De kwaliteit van de dataverzameling

S1 Is de steekproef representatief?

Bevindingen:

Sinds 2013/2014 wordt door Cito een nieuwe werkwijze voor het normeren van leerlingvolgsystemen toegepast. Bij deze werkwijze wordt de normeringssteekproef gebaseerd op gegevens verkregen via een embedded field onderzoek en op gegevens verkregen via Cito dataretour. Door de bij de normering gevolgde werkwijze is de representativiteit van de normeringssteekproeven in principe gegarandeerd. Niettemin werd er naast die garantie een controle uitgevoerd op de representativiteit door de populatieverdelingen van gegevens uit DUO te vergelijken met de steekproefverdelingen van B8 en M8. De representativiteit van de steekproeven is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. In de beoordeling door de expertgroep van de papieren versie van de Spelling toets wordt geconstateerd dat er voor M8 sprake is van een representatieve steekproef, maar dat de steekproef voor B8 als onvoldoende moet worden beoordeeld. Het voorgaande betekent dat de wetenschappelijke verantwoording van zowel de papieren als de digitale versie alleen betrekking heeft op de normering van M8. T.z.t. zal een addendum over de papieren en de digitale B8 verschijnen (zie pagina 5 van de aanvulling op de wetenschappelijke verantwoording, 2020).

Conclusie:

Bij de beoordeling van de papieren versie door de expertgroep werd geconcludeerd dat voor B8 de steekproef niet representatief was en dat de beoordeling alleen M8 betrof. Voor M8 was de steekproef wel representatief. Aan aspect S1 wordt het oordeel '**voldoende**' toegekend voor de werkwoorden en de niet-werkwoorden op afnamemoment M8.

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

Bevindingen:

In 2012 en 2013 zijn de opgaven voor de LVS toetsen Spelling 3.0 groep 8 geconstrueerd. Deze opgaven zijn allereerst in papieren versie onderzocht in kalibratieonderzoeken in 2015 en 2016 en in normeringsonderzoeken in 2017. Zie voor het kalibratieonderzoek van de papieren items voor groep 8 paragraaf 4.2 van de wetenschappelijke verantwoording van de LVS toetsen Spelling 3.0 voor groep 8. In januari 2018 vond het papier-digitaal kalibratieonderzoek M8 niet-werkwoorden en M8 werkwoorden plaats. De hoofdvraag in deze kalibratieonderzoeken was: meten de papieren items en de digitale items dezelfde vaardigheid? Alle opgaven van de papieren zijn omgezet naar digitale versies, aangevuld met reserve-opgaven. In een papier-digitaal onderzoek is het design onvolledig: in tegenstelling tot een volledig papieren onderzoek overlappen de zogeheten boekjes slechts

gedeeltelijk. De link om de boekjes op eenzelfde schaal te krijgen, verloopt via de papieren uitgave. De afnamedesigns in de tabellen 4.1 en 4.2 van de wetenschappelijke verantwoording laten zien dat de overlap via de papieren uitgave ervoor zorgt dat de afnamedesigns verbonden zijn.

Conclusie:

Het onvolledige dataverzamelingsdesign is adequaat om de papieren en digitale items op eenzelfde schaal te kunnen afbeelden en dus de vraag te beantwoorden of papieren en digitale items dezelfde vaardigheid meten. Aan aspect S2 wordt dan ook het oordeel 'voldoende' toegekend.

Normering

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

Bevindingen:

Aspect N1.1 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:

Aspect N1.2 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:

Aspect N1.3 heeft betrekking op absoluut normeren en is hier niet van toepassing.

Conclusie:

N.V.T.

N2.1 Zijn de normgroepen groot genoeg?

Bevindingen:

Voor de normering werden de gegevens uit het normeringsonderzoek aangevuld met de gegevens Cito dataretour. Het aantal leerlingen dat meegenomen is in de normering Spelling niet-werkwoorden op afnamemoment M8 was 1914 waarvan 945 afkomstig uit de steekproef en 969 uit dataretour. Aantal scholen was gelijk aan 114. Het aantal leerlingen dat meegenomen is in de normering Spelling werkwoorden op afnamemoment

Beoordeling van LVS toets Spelling 3.0 groep 8 dig. – Cito B.V.

M8 was 2560, waarvan 654 uit de steekproef en 906 uit dataretour. Aantal scholen was gelijk aan 87.

Voor het afnamemoment M8 zijn twee normtabellen op leerlingniveau voor Spelling niet-werkwoorden en werkwoorden opgesteld. Die tabellen geven het gemiddelde, de standaardafwijking, kurtosis, scheefheid en percentielen (P10, P20, P25, P40, P50, P60, P70, P75, P80, P90) van de verdelingen van vaardigheidsscores.

Conclusie:

De steekproeven zijn groot genoeg en dus wordt aan aspect N2.1 het oordeel '**voldoende**' toegekend.

N2.2 Zijn de normgroepen representatief?

Bevindingen:

Voor de beoordeling van dit aspect verwijzen we naar de bespreking van aspect S1 en de beoordeling van de papieren versie door de expertgroep.

Conclusie:

Omdat de normeringssteekproef voor B8 niet representatief was voor regio en schooltype zijn de normen voor afnamemoment voorlopig vastgesteld en geldig tot en met 2019.

De normeringssteekproef voor M8 is representatief voor de variabelen regio, urbanisatiegraad, schooltype en sekse. Aan aspect N2.2 wordt voor M8 het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1 Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen:

Gebruikmakend van het feit dat de items succesvol geschaald zijn met itemresponsmodel OPLM kan de MAcc (Accuracy of Measurement) berekend worden, die qua interpretatie grote overeenkomst vertoont met de betrouwbaarheid uit de klassieke testtheorie.

Conclusie:

De betrouwbaarheidsgegevens worden correct berekend. Op aspect B1 wordt aan de toets voor afnamemoment M8 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor de toets M8 niet-werkwoorden en de toets M8 werkwoorden is, uitgaande van 50 items, de MAcc gelijk aan 0,91. Ook de gesimuleerde test-hertest betrouwbaarheid is voor beide toetsen gelijk aan 0,91. Voor dit soort toetsen wordt een betrouwbaarheid van 0,80 en hoger door de COTAN als 'goed' aangemerkt. De figuren 5.1a en 5.1b geven grafisch weer hoe het met de lokale meetnauwkeurigheid van de twee toetsen gesteld is. De figuren

laten zien dat de meetfout kleiner is in de lagere en gemiddelde vaardigheidsregionen dan in de hogere vaardigheidsregionen. De betekenis van de (lokale) meetnauwkeurigheid voor de beslissingen die met een toets genomen worden, is af te leiden uit betrouwbaarheidstabellen 5.2 en 5.3. De tabellen laten zien hoe vaak de werkelijke vaardigheidsscore in dezelfde scoregroep valt als de geschatte vaardigheidsscore. 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 dit soort 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 M8 niet-werkwoorden en M8 werkwoorden is voldoende als aangenomen mag worden dat de toetsen geen zware consequenties voor de leerlingen hebben. Gegeven het voorgaande wordt aan de toetsen voor aspect B2 het oordeel '**voldoende**' toegekend.

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 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 in de vier toetsvarianten, zijn inhoudelijk bijna zonder uitzondering voldoende tot goed gekozen. De keuze voor en gebruiksmogelijkheden van de digitale toetsversies voor de doelgroepen en meetmomenten in groep 8 wordt uitgebreid 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 8. 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 8, 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 van groep 8 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 in de wetenschappelijke verantwoording verwezen naar de wetenschappelijke verantwoording van de papieren toetsen. Bovendien merken de auteurs op dat op basis van de statistische toetsen van het itemresponsmodel OPLM de modelfit voor de digitale opgaven bevredigend is (zie paragraaf 4.2.3) en dat daarmee net als bij de papieren versie voldaan wordt aan eisen van unidimensionaliteit als waarborg voor de constructvaliditeit van de toetsen.

Conclusie:

Aan aspect V2 wordt voor de digitale toetsen Spelling van groep 8 het oordeel '**voldoende**' toegekend.

Het volg-aspect

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

Bevindingen:

Uit de kalibratieonderzoeken blijkt dat de items van de digitale toetsen met behulp van het itemresponsmodel OPLM op een eendimensionale schaal afgebeeld kunnen worden en dat aan de hand van de door de leerling behaalde vaardigheidsscores diens vaardigheid adequaat gemeten en gevolgd kan worden.

Conclusie:

Aan aspect Va1 wordt voor de digitale toetsen Spelling van groep 8 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

In de wetenschappelijke verantwoording van de onderhavige digitale toetsen wordt aan dit aspect geen aandacht besteed. In de beoordeling van de papieren toetsen door de expertgroep wordt verwezen naar de wetenschappelijke verantwoording, waar toegelicht wordt hoe de toetsen ingezet kunnen worden om de ontwikkeling van leerlingen te volgen in de tijd.

Conclusie:

Gebaseerd op het oordeel van de expertgroep op de papieren toetsen wordt aan aspect Va2 voor de digitale toetsen van Spelling groep 8 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

In de beoordeling van de papieren toetsen door de expertgroep wordt verwezen naar onder andere hoofdstuk 7 van de handleiding waar beschreven wordt hoe er met de verschillende gebruikers over de toetsresultaten gecommuniceerd kan worden.

Conclusie:

Gebaseerd op het oordeel van de expertgroep op de papieren toetsen wordt aan aspect Va3 voor de digitale toetsen van Spelling groep 8 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.

Conclusie:

Aan aspect I1 wordt voor de digitale toetsen Spelling van groep 8 het oordeel '**voldoende**' toegekend.

Referentieniveaus

R1 Sluit de inhoud van de toets aan op de kennis en vaardigheden zoals omschreven in de referentieniveaus van het betreffende domein (voor toetsen vanaf groep 6)?

Bevindingen:

De toetsaanbieder geeft gedetailleerde en overzichtelijk informatie over de totstandkoming van de toets als geheel, en 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*:

Beoordeling van LVS toets Spelling 3.0 groep 8 dig. – Cito B.V.,

uitwerking van het referentiekader Nederlandse taal voor het domein begrippenlijst en taalverzorging op de basisschool; inventarisatie van inhoud en uit relevante taalmethodes) is een valide set toetsitems geselecteerd.

Conclusie:

Aan aspect R1 wordt voor de digitale toetsen Spelling van groep 8 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. (2019). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 groep 8*. Arnhem: Cito B.V.
- Tomesen, M., Wouda, J. en Horsels, L. (2020). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 digitaal voor groep 8. Aanvulling bij de wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 voor groep 8*. Arnhem: Cito B.V.
- Cito (2018). *Cito Volgstelsel primair en speciaal onderwijs. Spelling 3.0 groep 8*. Arnhem: Cito B.V.