

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 8 zijn een onderdeel van het Cito Volgstelsel primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 8 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 8 vormen dan samen met de LVS toetsen Spelling 3.0 groepen 3, 4, 5, 6 en groep 7 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 8 zijn bedoeld voor leerlingen in groep 8 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 8 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.

De toetsen zijn niet geschikt voor leerlingen met een (tijdelijk) beperkt gehoor.

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 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 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 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 B8 (niet-werkwoorden)
 - Toets M8 (niet-werkwoorden)
 - Toets B8 (werkwoorden)
 - Toets M8 (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 Jennifer Roubiès MSc (secretaris).

De kwaliteit van de dataverzameling

S1.1. Is de steekproef representatief?

Bevindingen:

Voor kalibratie en normering is gebruik gemaakt van dezelfde steekproeven. De data zijn verzameld in een embedded field onderzoek om de afname condities zoveel mogelijk op de echte condities te laten lijken. De verzamelde data zijn aangevuld met gegevens uit Cito dataretour.

De toetsen voor groep 8 zijn op twee momenten in het leerjaar af te nemen: half oktober/half november (B8-moment) en half januari/half februari (M8-moment). Deze beoordeling gaat enkel in op afnamemoment M8. De representativiteit van de steekproeven voor **B8 en M8** is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. Bij regio is uitgegaan van de vier landsdelen / regio's van de CBS-indeling. Bij urbanisatiegraad is uitgegaan van een tweedeling in stad en platteland, afgeleid van de CBS-indeling naar vijf niveaus van verstedelijking. Bij schooltype is uitgegaan van de formatiegewichten volgens OCW. Hierin worden drie niveaus onderscheiden die gebaseerd zijn op het opleidingsniveau van de ouders. Bij sekse is een tweedeling gemaakt naar jongens en meisjes. **De steekproefverdelingen wijken voor B8 af van de populatieverdeling voor regio en stratum. De effectgroottes phi waren ruim boven de 0.10 en zijn daarmee niet verwaarloosbaar.** Voor M8 is sprake van een representatieve steekproef.

In de Tabel 4.7b is voor de variabele stratum (gedefinieerd als een indicatie van de aard van de schoolpopulatie in termen van het percentage leerlingen met een afwijkend leerlinggewicht per school) bij M8 een effectgrootte van 0.126 gerapporteerd. Dit is net boven de kritische waarde van 0.100, maar door weging van de steekproef in de normering is dit nog op te lossen. De methode is de kalibratie uit te voeren met aparte vaardigheidsverdelingen voor de verschillende strata en daarna een gezamenlijke virtuele vaardigheidsverdeling via weging te creëren voor de normering. **Echter, bij de variabele stratum bij B8 zijn de effectgroottes veel extremer (0.290 en 0.491) en daar is weging onmogelijk omdat er een aantal strata in het geheel niet geobserveerd zijn. Oversampling gecombineerd met weging zou hier te verkiesbaar zijn geweest. Uit de ruwe scores van de individuele leerlingen uit het embedded field normeringsonderzoek en Cito dataretour (50%/50%) werden plausible values gegenereerd op de nieuw ontwikkelde vaardigheidsschaal. De steekproef voor B8 moet dus als onvoldoende worden beoordeeld.** In reactie hierop is in de handleiding opgenomen dat de wetenschappelijke verantwoording alleen ingaat op de normering van M8. In 2020 zal een addendum over B8 verschijnen.

De normering werd vervolgens gebaseerd op de plausible values van de leerlingen in de normeringssteekproef. De plausible values voor de afnamemomenten **B8 en M8** bleken een normale verdeling te vormen. De schoolverdeling werd bepaald met het intercept-only multilevel model. Dit model werd geschat via een bootstrap procedure. De intraklassecorrelatie (ICC) lag boven de 0.04, wat inhoudt dat een multilevelanalyse zinvol is. Ondanks het feit dat de percentielen van de normgegevens op schoolniveau dichter bij elkaar kwamen te liggen dan in de leerlingverdeling, waren de afstanden groot genoeg om scholen zinvol te classificeren in de verschillende niveaus.

Conclusie:

De procedure voor het samenstellen van de steekproeven is onderbouwd en de omstandigheden waaronder data is verzameld, is redelijk vergelijkbaar met de omstandigheden waaronder de toets wordt afgenomen.

Voor B8 was de steekproef niet representatief. Deze beoordeling behelst echter alleen M8.

Voor M8 was de steekproef wel representatief. Aan aspect S1.1. wordt het oordeel '**voldoende**' toegekend voor de werkwoorden en de niet-werkwoorden toets 3.0 groep 8 afname-moment M8.

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

Bevindingen:

Om te komen tot een set van psychometrisch en inhoudelijk geschikte items zijn de opgaven uit het proefonderzoek van en de opgaven uit de daaropvolgende normeringsonderzoeken gekalibreerd. Hiervoor is gebruik gemaakt van het IRT model OPLM. Met dit statistische model zijn de psychometrische kenmerken (moeilijkheidsparameters en discriminatie indices) van de items geschat.

In het kalibratieproces is uitgegaan van een onvolledig maar 'verbonden' design. De opgaven werden gemiddeld door 330 tot 352 leerlingen gemaakt, wat aan het minimum vereiste van 150 voldoet.

Uit het kalibratieonderzoek (S-toetsing, R1c-waarden en de constante 'c') blijkt dat de items passen bij voornoemd IRT model en dat het model ook past voor de toets als geheel. Dit betekent dat er zowel voor de toets Spelling niet-werkwoorden als voor Spelling werkwoorden sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Conclusie:

Het onvolledige maar 'verbonden' design van de proefonderzoeken is adequaat. Aan aspect S1.2 wordt het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

Beide toetsen zijn genormeerd voor de afnamemomenten **B8** en M8.

In januari 2015 zijn 153 niet-werkwoorden items, verdeeld over 7 boekjes, geproeftoetst, waarbij elke nieuwe opgave door minimaal 200 leerlingen gemaakt werd. In januari 2016 zijn 127 werkwoorden items geproeftoetst, verdeeld over 7 boekjes, waarbij elke nieuwe opgave gemaakt werd door minimaal 200 leerlingen. Voor niet-werkwoorden werd een tekort gesignaleerd voor een aantal categorieën. Aanvullend werden in het voorjaar 2016 nog 33 nieuwe opgaven voorgelegd aan 98 leerlingen.

Na samenstelling van de nieuwe toetsen is een normeringsonderzoek uitgevoerd. In deze beoordeling wordt enkel ingegaan op M8. De gegevens van de genormeerde toets Spelling niet-werkwoorden zijn gebaseerd op **1008 leerlingen voor B8** en 1914 leerlingen voor M8. De gegevens van de genormeerde toets Spelling werkwoorden zijn gebaseerd op 1560 leerlingen. Deze leerlingen komen uit het normeringsonderzoek, aangevuld met dataretour.

Voor het afnamemoment M8 werden vaardigheidsverdelingen gepresenteerd op leerlingniveau en op schoolniveau. Dit betreft de gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75, P80 en P90. Van hieruit kunnen de beide niveau indelingen (de symmetrische niveau indeling I t/m V en de asymmetrische niveau indeling A t/m E) worden bepaald.

Conclusie:

Er is sprake van relatieve normen, de steekproeven zijn groot genoeg. Daarmee wordt aan aspect N1.2.1. het oordeel **'voldoende'** toegekend voor afnamemoment M8.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

~~Zoals beschreven wordt op pag. 56 van~~ In de handleiding is aangegeven dat er zijn er in de normeringssteekproef voor B8 afwijkingen geconstateerd zijn voor regio en stratum, zowel voor niet-werkwoorden als werkwoorden. De normen voor de niet-werkwoorden toets 3.0 groep 8 en de werkwoorden toets 3.0 groep 8 voor het afname moment B8 zijn voorlopig vastgesteld en geldig tot en met 2019. (Zie ook de bevindingen onder S1.1.). Onderstaande beoordeling heeft dus alleen betrekking op M8.

Bevindingen:

De normeringssteekproef voor M8 is representatief voor de variabelen regio, urbanisatiegraad, stratum en sekse. De normen voor de werkwoorden toets 3.0 groep 8 en de niet-werkwoorden toets 3.0 groep 8 afnamemoment M8 zijn geldig tot en met 2027.

Conclusie:

Aan aspect N1.2.2. wordt het oordeel '**voldoende**' toegekend voor de werkwoorden en de niet-werkwoorden toets 3.0 groep 8 afname-moment M8.

Betrouwbaarheid

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

Bevindingen:

Om relevante gegevens bij de toetsen te genereren, is gebruik gemaakt van het programma OPLAT. Binnen dit programma wordt de coëfficiënt MAcc ('Accuracy of Measurement') berekend. Deze coëfficiënt vertoont qua interpretatie grote overeenkomst met de betrouwbaarheidscoëfficiënt uit de KTT. Deze coëfficiënt wordt in de psychometrische literatuur beschreven en als correct aangemerkt.

Daarnaast wordt voor de toetsen B8/M8 niet-werkwoorden en B8/M8 werkwoorden het percentage correcte classificaties per scoregroep afgeleid en worden de plus/minus 1 niveau index en de marginal Classification Accuracy berekend om de meetnauwkeurigheid weer te geven.

Conclusie:

Aan aspect B1.1 wordt het oordeel '**voldoende**' toegekend.

B1.2. 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.

Voor de toets niet-werkwoorden 3.0 groep 8 en werkwoorden 3.0 groep 8 wordt deze coëfficiënt berekend als MAcc voor de afnamemomenten B8 en M8. De afnamecontext van de toets leent zich dankzij een OPLM kalibratie voor een gesimuleerd test-hertest onderzoek onder ideale condities. De test-hertest coëfficiënt is identiek aan de MAcc voor alle afnamemomenten. De resultaten laten zich lezen als standaardmeetfout (niet-werkwoorden B8 2.95; M8 2.84 en werkwoorden B8 2.96; M8 2.91), MAcc (niet-werkwoorden B8 0.90; M8 0.91 en werkwoorden B8 0.86; M8 0.88) en test-hertest simulatie (niet-werkwoorden B8 0.90; M8 0.91 en werkwoorden B8 0.86; M8 0.88) en worden aangemerkt als 'voldoende'.

Aanvullend hierop is de lokale meetnauwkeurigheid weergegeven in betrouwbaarheidstabellen (misclassificaties). Uitgaande van de betrouwbaarheidstabellen worden twee indices voor de nauwkeurigheid van de classificaties gerapporteerd: de

plus/minus 1 niveau-index en de marginal classification index. Uit de hoogte van de indices blijkt dat de laagst en de hoogst scorende leerlingen accuraat te classificeren zijn, maar dat over het algemeen tussen leerlingen in de niveaugroepen B, C en D, respectievelijk II, III en IV, er een licht minder duidelijk onderscheid te maken is. De marginal classification indices lopen uiteen van 62 tot 69 procent en de resultaten stemmen hiermee eveneens tot tevredenheid.

Conclusie:

De betrouwbaarheid van de toetsen Spelling 3.0 groep 8 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft en ingestemd wordt met de beoordelingscriteria voor de betrouwbaarheid van de COTAN.

Op aspect B.1.2. wordt aan de toetsen Spelling 3.0 groep 8 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 inhoudsvaliditeit wordt beoordeeld door de onderwijskundige expert(s). Het theoretisch kader waarop de toetsconstructie technisch en inhoudelijk gebaseerd is, wordt in de wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 voor groep 8 helder beschreven. In paragraaf 3.2 en paragraaf 6.1 worden gemaakte keuzes betreffende relevantie en objectiviteit beschreven en onderbouwd. De inhoudelijke uitgangspunten, in combinatie met de beschreven aan toetsconstructie gerelateerde uitgangspunten geven helder en stapsgewijs de gemaakte keuzes en bevindingen weer. De items in de beoordeelde toetsen Spelling 3.0 voor groep 8 zijn te plaatsen in het beschreven kader in de Referentieniveaus, domein Begrippenlijst en Taalverzorging, en in een gedetailleerd overzicht uitgewerkt in diverse categorieën voor zowel werkwoorden als niet-werkwoorden. De items in elke variant van de toets bevragen de actieve beheersing van een goedgekozen set werkwoorden en niet-werkwoorden die in passende zinnen van context zijn voorzien.

De (psychometrische) begripsvaliditeit wordt uitgewerkt in unidimensionaliteit, itemkwaliteit, itembias, convergente en divergente validiteit en in verschillen tussen relevante subgroepen.

De resultaten van de uitgevoerde kalibratie maken het aannemelijk dat er sprake is van unidimensionaliteit. Dit betekent dat met elke willekeurige subset van items uit de gekalibreerde itembank dezelfde onderliggende vaardigheid in spelling kan worden vastgesteld. Dit wordt tevens bevestigd door de nauwkeurigheid van de itemparameterschattingen (voor twee items uit Spelling niet-werkwoorden en acht items van Spelling werkwoorden, is de constante c iets groter dan 0.20).

De gemiddelde moeilijkheidsgraden van de items (als criterium voor de itemkwaliteit) voldoen. De gemiddelde R_{it} (ook een criterium voor de itemkwaliteit) voldoet eveneens. Voor alle vier de toetsen is de gemiddelde R_{it} waarde 0.38 of hoger en bij geen enkele toets ligt de R_{it} waarde onder de 0.20.

In het onderzoek naar itembias is nauwelijks tot geen sprake van DIF (Differential Item Functioning) naar sekse.

De constructvaliditeit is uitgewerkt in convergente validiteit door de samenhang te onderzoeken met de voorgaande versie Spelling 2.0 en met Spelling 3.0 groep 7. De correlaties met Spelling 2.0 waren hoog (0.91 - 0.92). De correlaties met Spelling 3.0 groep 7 waren ook hoog (0.88 - 0.90). De convergente validiteit is verder onderzocht door te kijken naar het onderdeel Passieve taalverzorging uit de toets Taalverzorging. De correlatie voor niet-werkwoorden was hoog (0.82). De correlatie voor werkwoorden was eveneens hoog (0.84). Tenslotte is gekeken naar de longitudinale vaardigheidstoename. Voor beide onderdelen is te zien dat de gemiddelde latente vaardigheid van de leerlingen voor elke afname toeneemt.

De divergente validiteit is onderzocht door de samenhang met de te onderzoeken met LVS toetsen Technisch lezen – DMT groep 8, Begrijpen lezen groep 8, Woordenschat groep 8 en Rekenen – Wiskunde groep 8. De hogere correlaties met de leestoetsen (0.52 - 0.58) en de lagere correlaties met de Rekenen toets (0.45 - 0.54) zijn conform de verwachtingen. Een uitzondering is de lage correlatie van de werkwoorden toets met de Technisch Lezen – DMT toets ($r = 0.36$). Waarschijnlijk heeft dit te maken met het feit dat er bij werkwoordspelling ook andere aspecten een rol spelen dan alleen het technische aspect.

Conclusie:

Op aspect V1.1 wordt aan de toetsen Spelling 3.0 groep 8 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:

Het algemene (inhoudelijke) uitgangspunt voor de toetsen Spelling 3.0 is dat de (latente) vaardigheden Spelling niet-werkwoorden kunnen worden opgevat als een unidimensioneel continuüm en dat elke leerling kan worden voorgesteld als een punt op deze continuüm. Voor Spelling werkwoorden geldt hetzelfde, alleen zijn deze toetsen alleen ontwikkeld voor de groepen 7 en 8.

Uit het kalibratieonderzoek blijkt dat de items passen bij het gehanteerde IRT model en dat het model ook past als geheel. Dit betekent dat er sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden voor zowel niet-werkwoorden als werkwoorden.

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

Afhankelijk van het aantal items dat een leerling goed maakt, wordt er een vaardigheidsscore toegekend.

Conclusie:

Aan aspect VA1.1. wordt 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 hoofdstuk 7 van de handleiding ('Communiceren over toetsresultaten met leerling en ouders') wordt beschreven hoe er met de verschillende gebruikers over de toetsresultaten kan worden gecommuniceerd. Hierin wordt onderscheid gemaakt tussen 'niveau' en 'groei', wat wordt onderbouwd met diverse rapportage mogelijkheden.

In de wetenschappelijke verantwoording wordt toegelicht hoe de toetsen ingezet kunnen worden om de ontwikkeling van leerlingen te volgen in de tijd, namelijk door het toetsresultaat van een leerling te vergelijken met andere leerlingen en door het toetsresultaat van een leerling te vergelijken met diens andere toetsresultaten. Voor alle vergelijkingen geldt dat uitspraken over de voortgang van leerlingen gerelativeerd moeten worden vanwege de (on)betrouwbaarheid van de toetsen. Betrokkenen bij de toetsen Spelling moeten beseffen dat vaardigheidsgroei zich langzaam in de tijd voltrekt.

Conclusie:

Aan aspect VA1.2. wordt 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). Voor ouders is met name het leerlingrapport of alternatief leerlingrapport informatief, omdat deze rapporten individueel de vaardigheid en de groei van hun kind weergeven.

In de Handleiding wordt in hoofdstuk 7 aandacht besteed aan de wijze waarop 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 numeriek en grafisch gepresenteerd worden.

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

Daarnaast wordt de leraar 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. De informatie biedt goede handvatten voor de gesprekken met ouders. In hoofdstuk 8 worden veelgestelde vragen behandeld die weliswaar voor de leraren bestemd zijn, maar waar de antwoorden voor een deel ook informatief zijn tijdens bijvoorbeeld de tienminutengesprekken. Voor de interpretatie van toetsresultaten is ook een folder ouderinformatie beschikbaar die men via de website van het Cito kan downloaden.

Conclusie:

Op aspect I1.1 wordt aan de toetsen Spelling 3.0 groep 8 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	V1.1	Voldoende
Volg-aspect	VA1.1	Voldoende
	VA1.2	Voldoende
Inzicht in leervorderingen	I1.1	Voldoende