

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 Volgsysteem primair 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 voor groep 4 zijn een onderdeel van het Cito Volgsysteem primair onderwijs en zijn bedoeld voor leerlingen in groep 4 van het primair onderwijs. Onderstaande beschrijving is gebaseerd op de Handleiding.

Meetpretentie

De toetsen in de toets pakketten Spelling 3.0 van het Cito Volgsysteem 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 4 zijn bedoeld voor leerlingen in groep 4 van het primair en speciaal onderwijs, maar kunnen ook gebruikt worden voor leerlingen uit andere jaargroepen die werken op het niveau van groep 4 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 4 specifiek het schrijven van eenvoudige klankzuivere en niet-klankzuivere woorden van één of twee 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 en lezen) 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 4 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 veelgestelde vragen (hfdst 8) en
 - enkele bijlagen
- Vier toetsen:
 - Toets M4 (Medio groep 4)
 - Toets E4 (Eind groep 4)
 - Toets E3M4 (makkelijke variant van de toets M4)
 - Toets M4E4 (makkelijke variant van de toets E4)
- Afnamekaarten met aanwijzingen voor de papieren of de digitale afname van de toetsen
- Nakijkkaarten
- Antwoordbladen
- Tabellen voor de vier toetsen voor het bepalen van de vaardigheidsscore en – niveau.

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 Paulyn 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 leerlingvolgsystemen toegepast. Een belangrijk aspect van die werkwijze is om de normeringssteekproef (voor 50 procent) te baseren op gegevens uit het zogenaamde embedded field onderzoek en (voor 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 en sekse. Op pagina 47 van de 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 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. Ook bleek de gemiddelde score op de Cito Eindtoets Basisonderwijs, en specifiek de score voor Spelling 3.0 voor Groep 4, 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 afweken 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 3.0 voor Groep 4 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 verantwoording worden de aantallen leerlingen gegeven die meegenomen zijn bij de normering. Voor M4 betreft dat 6717 leerlingen (3115 leerlingen uit de steekproef en 3602 leerlingen via dataretour) afkomstig van 264 scholen en voor E4 6094 leerlingen (2854 leerlingen uit de steekproef en 3240 leerlingen via dataretour) afkomstig van 220 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 3 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor de toetsen Spelling groep 4 heeft kalibratieonderzoek plaatsgevonden tijdens het proefonderzoek en het normeringsonderzoek. Tijdens het proefonderzoek zijn in 2012 halverwege en aan het einde van leerjaar 4 opgaven voorgelegd aan leerlingen. Daarbij zijn op het afnamemoment medio groep 4 225 nieuwe opgaven geproeftoetst en op het afnamemoment eind groep 4 wederom 225 opgaven. Elke deelnemende school maakte één taak met 25 nieuwe opgaven en de LVS-toets Spelling van de tweede generatie. 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 een 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 niet in de wetenschappelijke verantwoording vermeld. 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 (p-waarden en rir-waarden uit de klassieke testtheorie) criteria werden 40 opgaven voor de toets M4, 40 opgaven voor de toets E3M4, 40 opgaven voor de toets M4E4, en 40 opgaven voor de toets E4 geselecteerd. In vergelijking met de Spelling 3.0 voor groep 3 was het voor groep 4 gemakkelijker om geschikte woorden voor de toets te vinden. Toch is ervoor gekozen om ook in groep 4 toetsen aan te bieden met maar 20 opgaven per taak (in totaal dus 40 opgaven), zodat leerlingen hun concentratie niet zo lang hoeven vast te houden hetgeen voor leerlingen in de lagere groepen prettig is.

Hoofdstuk 4 van de verantwoording betreft de kalibratie en normering en presenteert een zogenaamd macro-design 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. Ook de kalibratie van leerjaar 4 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 M4 en E4. M4 bestaat uit zes toetsversies/boekjes en E4 eveneens uit zes toetsversies/boekjes. Boekje 1 van M4 bestaat uit items M4 LVS 2e generatie, E3M4 deel 1 en M4E4 deel 1, boekje 2 bestaat uit items M4 LVS 2e generatie, E3M4 deel 2 en M4 deel 1, boekje 3 bestaat uit items van M4LVS 2e generatie, M4 deel 1 en M4 deel 2, boekje 4 bestaat uit items van M4 LVS 2e generatie, M4 deel 2 en M4E4 deel 1, boekje 5 bestaat uit items van M4LVS 2e generatie, E3M4 deel 2 en M4E4 deel 2 en boekje 6 bestaat uit M4 LVS 2e generatie, E3M4 deel 1 en M4E4 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 214 verschillende items uit het normeringsonderzoek M4 zijn door 1634 leerlingen gemaakt. De 25 items uit de bestaande taak Spelling werden door alle 1634 leerlingen gemaakt en de overige items werden door gemiddeld 904 leerlingen gemaakt. De 210 verschillende items uit het normeringsonderzoek E4 zijn in 2013 door 1539 leerlingen gemaakt. De 25 items uit bestaande taak Spelling van de tweede generatie werden alleen door de 1539 leerlingen uit het 2013-onderzoek gemaakt en de overige items door gemiddeld 934 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 4 het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

In tabel 4.9 van de verantwoording wordt voor M4 en E4 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. 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 verantwoording wordt voor M4 en E4 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.

Conclusie:

De normgroepen zijn groot genoeg (zie S1.1). Op aspect N1.2.1 wordt aan de toetsen Spelling 3.0 groep 4 het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

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

Voor Spelling 3.0 Groep 4 geldt dat de normen geldig zijn tot en met 2023.

Conclusie:

Op aspect N1.2.1 wordt aan de toetsen Spelling 3.0 groep 4 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, 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 4 het oordeel '**voldoende**' toegekend.

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

Bevindingen:

Voor E3M4, M4, M4E4 en E4 zijn drie betrouwbaarheidsgegevens berekend: standaardmeetfout, M_{Acc} en een gesimuleerde test-hertest betrouwbaarheidscoëfficiënt. Voor E3M4 zijn de gegevens gelijk aan 2,121, 0,88 en 0,88, voor M4 gelijk aan 2,389, 0,89 en 0,89, voor M4E4 gelijk aan 2,434, 0,88 en 0,88 en voor E4 gelijk aan 2,535, 0,90 en 0,90. 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 4, 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 niveauperdelingen (I t/m V en A t/m E) voor E3M4, M4, M4E4 en E4. De tabellen laten zien dat bij toets E4 met een M_{Acc}-betrouwbaarheid van 0,90, de leerlingen betrouwbaarder geclassificeerd worden dan bij de drie andere toetsen. In het algemeen worden de leerlingen met de hoogste en de laagste niveau-indicatie voldoende betrouwbaar geclassificeerd. 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, dwz dat gebruikers rekening dienen te houden met maximaal 1 niveau verschil.

Conclusie:

De betrouwbaarheid van de toetsen Spelling 3.0 groep 4 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 basis van het voorgaande wordt op aspect B1.2 aan de toetsen Spelling 3.0 groep 4 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:

In de groepen 3 en 4 zijn de toetsen bedoeld om vast te stellen hoe goed een leerling eenvoudige klankzuivere en niet-klankzuivere woorden van één of twee lettergrepen kan schrijven. Dit gebeurt door actieve toetsing (dictee) wat een goede keuze is. Logischerwijs wordt een beperkt aantal spellingcategorieën getoetst. De zinnen in de toetsen zijn ook gezien de doelgroep kort (bijv. Onze hond is heel trouw) en de woorden die de kinderen opschrijven zijn voor de meeste kinderen bekend en passen binnen de spellingcategorieën die worden gehanteerd. In de handleiding wordt via tabellen inzicht gegeven welke spellingcategorieën worden gehanteerd en welke woorden in de toetsen eronder vallen. Ook wordt inzicht gegeven in de moeilijkheid van opgaven op basis van vaardigheidsscores. Leervorderingen worden systematisch weergegeven via allerlei rapportages en er worden handvatten geboden voor analyse.

Er wordt een duidelijke relatie gelegd met het Referentiekader. In de toetsen die nog op de markt komen voor de bovenbouw zal echter duidelijk moeten worden of de school kan nagaan of een leerling referentieniveau 1F of 1S behaalt.

Merk op dat de beoordeling van dit aspect zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat verricht is.

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 en dat de moeilijkheidsgraad van de items een goede spreiding had. De samenhang tussen de afzonderlijke items met de totaalscore bleek hoog te zijn. Onderzoek naar de samenhang tussen de toetsen Spelling met andere LVS toetsen zoals Rekenen-Wiskunde E4, Begrijpend lezen E4, Technisch lezen E4 en Woordenschat E4 lieten correlaties zoals verwacht zien. Onderzoek naar differentieel itemfunctioneren (dif) met betrekking tot jongens en meisjes liet zien dat er geen sprake was van dif naar sekse. Onderzoek naar verschillen tussen relevante subgroepen liet zien dat: 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 lager scoren dan leerlingen met leerlinggewicht 0.

Conclusie:

Op aspect V1.1 wordt aan de toetsen Spelling 3.0 groep 4 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 4 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 4 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 4 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. 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. 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 veelgestelde 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 4 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	<i>I1</i>	Voldoende
----------------------------	-----------	------------------

4. Literatuurlijst

Bij deze beoordeling zijn de volgende, door Cito B.V. verstrekte, materialen gebruikt:

- M. Tomesen, J. Wouda, A. Mols & L. Horsels (2015). *Wetenschappelijke verantwoording van de LVS-toetsen Spelling 3.0 voor groep 4*. Arnhem: Cito B.V.
- *Leerkrachtmap Spelling 3.0 voor groep 4*. Arnhem: Cito B.V.