

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 Begrijpend Luisteren groep 7 zijn een onderdeel van het Cito Volgsysteem primair onderwijs (tweede generatie toetsen) en zijn bedoeld voor leerlingen in groep 7 van het primair onderwijs. De toets vormt samen met de LVS toetsen Begrijpend Luisteren voor groep 3 tot en met 8 een systeem om vast te stellen hoe leerlingen met begrip kunnen luisteren en hoe hun luistervaardigheid zich in de basisschoolperiode ontwikkelt. Onderstaande beschrijving is gebaseerd op de Handleiding en de Wetenschappelijke verantwoording.

Meetpretentie

De toetsen in de toets pakketten Begrijpend Luisteren groep 7 van het Cito Volgsysteem primair en speciaal onderwijs zijn bedoeld om vast te stellen hoe goed een leerling een betekenis kan toekennen aan gesproken taal en of het adequaat kan reageren op gesproken taal en hoe de vaardigheid in begrijpend luisteren van de leerling zich in de loop van de jaren ontwikkelt.

Doelgroep

De toetsen Begrijpend Luisteren 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.

Voor leerlingen die nog maar pas in Nederland verblijven of gehoorproblemen hebben, zijn de toetsen ongeschikt. Een leerling moet voldoende taalvaardig in het Nederlands zijn (vergelijkbaar met Nederlandstalige leerlingen in groep 7) voordat de toets Begrijpend Luisteren kan worden afgenomen.

Gebruiksdoel en functie

Doel van de toetsen Begrijpend Luisteren groep 7 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van begrijpend luisteren. Hiervoor wordt de behaalde vaardigheidsscore normgericht geïnterpreteerd op basis van de vaardigheidsverdeling in een adequate, landelijke, referentiegroep. De vaardigheidsscore wordt uitgedrukt in de symmetrische niveau indeling I t/m V en in de asymmetrische niveau indeling A t/m E. De toetsen maken het mogelijk om:

- De vaardigheid begrijpend luisteren 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 begrijpend luisteren 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 toets Begrijpend Luisteren is gebaseerd op de door het OCW geformuleerd kerndoelen en de door het project TULE ontwikkelde tussendoelen en leerlijnen. De kerndoelen voor het onderwijs in begrijpend luisteren zijn grotendeels ondergebracht bij 'Mondeling taalonderwijs', kerndoel 1 en bij 'Taalbeschouwing', kerndoel 12. In het project TULE zijn de kerndoelen uitgewerkt in tussendoelen en leerlijnen per basisschoolgroep. Het construct luisteren wordt hierbij gedefinieerd als de interactie tussen begrijpen, interpreteren en reflecteren wat plaatsvindt binnen een context die steeds vaker wordt gevoed met zowel auditieve als visuele stimuli ('de beeldcultuur').

Inhoud van het toetspakket

Het toetspakket Begrijpend Luisteren 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),
 - 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
- Één toets:
 - Toets M7/E7
- Afnamekaarten met aanwijzingen voor het afnemen van de toetsen
- Nakijkaarten
- Antwoordbladen
- Tabellen om de vaardigheidsscore en -niveau te bepalen
- 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 Paulyn Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

In het kalibratieonderzoek van mei 2013 zijn 188 items voorgelegd aan 2126 leerlingen van groep 7. Er is gebruik gemaakt van een gelinkt design. Het gemiddeld aantal leerlingantwoorden per item was 432. Op grond van de gegevens uit het kalibratieonderzoek is een selectie van items gemaakt voor de normeringsonderzoeken.

Geconstateerd kan worden dat de steekproefgrootte voldoet aan het doel. Verder worden er tabellen m.b.t. de representativiteit gepresenteerd. De representativiteit van de steekproeven is onderzocht met betrekking tot leerlinggewicht, schoolgrootte, regio, verstedelijking en geslacht. Er wordt echter niet beschreven hoe de steekproef tot stand is gekomen. Gezien de tot nu toe m.b.t. Begrijpend Luisteren gemaakte beoordelingen, is het niet consequent om op basis van dit tekort een beoordeling onvoldoende te geven. Voor toekomstige beoordelingen door de Expertgroep kan afwezigheid van informatie over de methode van steekproeftrekking wel tot een onvoldoende leiden.

Conclusie:

Er worden adequate gegevens verstrekt over de representativiteit van de kalibratiesteekproef, maar niet over de methode van steekproeftrekking. Op aspect S1.1 wordt de toets Begrijpend Luisteren voor groep 7 als 'voldoende' beoordeeld, met de aantekening dat in de toekomst de methode van steekproeftrekking wel beschreven dient te worden.

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

Bevindingen:

De 188 items waren verdeeld over 9 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit ongeveer 38 items. Elk item kwam in twee boekjes voor. Het gemiddeld aantal leerlingeantwoorden per item was 432. De items zijn gekalibreerd met OPLM en de modelfit is getoetst. Zowel op individueel item als op het niveau van de gehele verzameling items past het model.

Conclusie:

Op aspect S1.2 wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

Voor het afnamemoment M7 hebben 1766 leerlingen meegedaan aan het normeringsonderzoek en voor het afnamemoment E7 1267 leerlingen. Deze aantallen zijn groot genoeg om op percentiële niveau schattingen te maken voor de populatie. Uit de gegevens blijkt dat de gemiddelde vaardigheid in begrijpend luisteren in de periode tussen de afnamemomenten toeneemt, terwijl de spreiding in de scores nagenoeg gelijk blijft.

Conclusie:

Op aspect N1.2.1 wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

Voor de normeringsonderzoeken M7 en E7 werden scholen geworven na het trekken van een representatieve steekproef, waar bij rekening gehouden werd met verdeling over strata, schoolgrootte, regio en verstedelijking. Achteraf is de verdeling over sekse gecontroleerd. Voor het afnamemoment M7 deden 1766 leerlingen van 74 scholen mee. Voor het afnamemoment E7 deden 1267 leerlingen van 51 mee.

De representativiteit van de steekproef is onderzocht met betrekking tot strata, schoolgrootte, regio, urbanisatiegraad en sekse. Bij de definitie van strata is gebruikgemaakt van de formatiegewichten van de leerlingen binnen een school. In navolging van OCW worden op basis van het opleidingsniveau van de ouders drie opleidingsniveaus onderscheiden: 0.0 = één van de ouders of beide ouders heeft of hebben een opleiding gehad uit categorie 3; 0.3 = beide ouders of de ouder die belast is met de dagelijkse verzorging heeft of hebben een opleiding uit categorie 2 gehad; 1.2 = één van de ouders heeft een opleiding gehad uit categorie 1 en de ander een opleiding uit categorie 1 of 2. In deze indeling wordt verwezen naar de volgende categorieën in het opleidingsniveau van de ouders: categorie 1 is maximaal basisonderwijs, categorie 2 is maximaal lbo/vbo en categorie 3 is voortgezet onderwijs en hoger. Voor elke school wordt het percentage leerlingen met een gewicht afwijkend van 0,0 bepaald. Gebaseerd op deze gewichten zijn vier groepen scholen gevormd. Bij schoolgrootte is onderscheid gemaakt tussen klein en groot. Een kleine school telt minder dan 200 leerlingen. Bij regio is uitgegaan van de CBS-indeling naar landsdeel waarbij de regio's noord, oost, west en zuid onderscheiden worden. Bij urbanisatiegraad is gebruik gemaakt van de vijf niveaus die door het CBS worden onderscheiden: zeer sterk, sterk, matig, weinig en niet verstedelijkt. Tenslotte werd bij de variabele sekse een tweedeling naar jongens en meisjes gehanteerd.

Uit de resultaten van de representativiteitsanalyse bleek dat de steekproef op geen van de genoemde variabelen significant afweek van de populatie voor beide afname momenten.

De geldigheid van de normen is vastgesteld tot en met 2024. Cito monitort periodiek de normering. Mochten er systematische verschuivingen plaatsvinden, dan wordt de normering indien nodig aangepast.

Conclusie:

Op aspect N1.2.2 wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

Betrouwbaarheid

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

Bevindingen:

In tabel 5.1 op pagina 46 van de wetenschappelijke verantwoording worden van de twee afnamemomenten van de toets Begrijpend lezen Standaard meetfout, de MAcc en de gesimuleerde Test-hertest betrouwbaarheid vermeld. Er wordt omschreven hoe de simulatie is uitgevoerd. Gegeven de software die er voor deze berekeningen is gebruik, mag er van uitgegaan worden dat voornoemde gegevens correct berekend zijn.

Conclusie:

Op aspect B1.1. wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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

Bevindingen:

De betrouwbaarheidsscores waren hoger dan 0,70 en mogen daarmee als voldoende beoordeeld worden volgens de richtlijnen van de COTAN.

Wat betreft de lokale betrouwbaarheid, wordt in Tabel 5.3a voor afnamemoment M7 en in Tabel 5.3b voor afnamemoment E7 weergegeven hoe vaak de werkelijke vaardigheidsscore en de geschatte vaardigheidsscore in dezelfde categorie vallen. Voor de extreme categorieën is dit percentage voldoende hoog. Bij de middencategorieën is dit percentage vrij laag te noemen. De marginal classification accuracy loopt uiteen van 53 tot 55 procent. Dit is in lijn overigens met de gerapporteerde overall betrouwbaarheidsscores die net boven de 0,75 liggen.

Conclusie:

Op aspect B1.2. wordt Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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 wetenschappelijke verantwoording zijn tabellen opgenomen van de verdeling van de opgaven naar vaardigheid (begrijpen en interpreteren) en inhoudsaspecten. Begrijpen en interpreteren liggen op dezelfde vaardigheidsschaal.

Opgaven worden als volgt verdeeld naar vaardigheid en inhoudsaspecten:

Begrijpen van gesproken taal:

- Expliciete betekenistoekenning: betekenis woord en woordgroep
- Specifieke inhoudselementen die expliciet in de tekst aan de orde zijn (o.a. meningen, voorwerpen, aantallen, tijdsperioden, handelingen)
- Eenvoudige expliciete verbanden (o.a. vergelijkingen, tegenstellingen, vraag en antwoord)
- Complexe expliciete verbanden (o.a. oorzaak-gevolg, middel-doel, hoofd en bijzaken, volgorde van opeenvolgende stappen)

Opgaven bij interpreteren:

- Impliciete betekenistoekenning: afleiden van betekenis van een woord
- Inzet van voorkennis: afleiden van informatie uit de tekst (o.a. ontbrekende info, bedoeling of mening van spreker, functionele betekenis van tekstdelen)

- Globale inhoud van de tekst: o.a. verbinden, vergelijken en doorzien van informatie en verbanden daarin: hoofdgedachte, hoofdpersoon, doel en publiek herkennen; samenvatten)
- Manier van spreken: vragen over klemtoon, tempo, toon, accent, register en conventies, maar ook vragen naar verbanden tussen tekstuele informatie en kennis van het taalsysteem.

De toets als geheel bestaat uit 33 opgaven: 13 voor 'Begrijpen' en 20 voor 'Interpreteren'.

In tabel 3.2 in de Wetenschappelijke verantwoording wordt het aantal opgaven en hun nummer(s) weergegeven per inhoudsaspect voor begrijpen en interpreteren. In de Handleiding ontbreekt helaas een vergelijkbaar overzicht van wat getoetst wordt bij welke tekst en bijbehorende vragen. Dit is jammer voor de leerkracht, zeker als de toetsen ook handvatten moeten geven waar kinderen op uitvallen en waar extra aandacht aan zou moeten worden besteed. Toevoeging van dit overzicht aan de Handleiding heeft meerwaarde.

Er wordt in de toets voor groep 7 gebruik gemaakt van twee teksttypen (fictie en non-fictie) en vijf tekstgenres: verhaal/film/drama; nieuwsbericht; documentaire/verslag; betoog; interview/gesprek.

De teksten zijn als geheel van goede kwaliteit en lengte en hebben een duidelijke opbouw en structuur. Enige kanttekening daarbij is de tekst 'De brief voor de koning'. Deze tekst bestaat uit een compilatie van fragmenten uit de gelijknamige speelfilm. (On-)bekendheid met het boek en/of de film kan van invloed zijn op de prestatie van leerlingen: anders dan in alle andere gebruikte teksten zitten er namelijk grote 'gaten' in het verhaalverloop. Het is begrijpelijk dat niet de hele film gebruikt kan worden, maar wellicht had er één langer, aangesloten fragment gekozen kunnen worden om meerdere vragen over te stellen.

De teksten sluiten aan bij de leefwereld van de kinderen (Jeugdjournaal: geld en crisis; interview: stotteren; Klokhuis aflevering etc.). De herhaalde fragmenten voor elke vraag of set vragen zijn doorgaans opnieuw erg lang. Aangezien de leerlingen de vraag bijna altijd pas na afloop van het fragment te horen krijgen, is het beroep dat op het auditief geheugen gedaan wordt, nogal eens behoorlijk. De noodzaak tot het herhalen van dergelijke grote deelfragmenten voor soms een 'kleine' vraag is niet altijd evident: toetsen we daarmee begrip of geheugen?

Uitgangspunten bij de toetsconstructie waren met name de kerndoelen primair onderwijs en de tussendoelen mondelinge communicatie. Zowel bij de kerndoelen als de tussendoelen is begrijpend luisteren één van de aspecten van mondeling taalonderwijs. Er wordt hierdoor maar één onderdeel getoetst, namelijk dat leerlingen leren informatie te verwerven uit gesproken taal. Dit is één van de drie kerndoelen rond mondelinge taal. Maar het leren argumenteren of het leren uit te drukken (de andere twee kerndoelen) vallen hier niet onder. Dat wordt teruggezien t.a.v. welke tussendoelen herkenbaar zijn in de toetsen. Dit zijn vooral de doelen rond het begrijpen van het verhaal, informatie uit het verhaal halen, het afleiden van de hoofdgedachte, gebruik maken van de structuur van het verhaal en niet het samenvatten van het verhaal in eigen woorden, bewust zijn van de impact van media etc. Opvallend is dat er in de handleiding bij de theoretische inkadering wel kort aandacht wordt besteed aan de uitwerking van de expertgroep Doorlopende Leerlijnen, maar dat de tussendoelen en leerstoflijnen uitgangspunt zijn geweest bij de opzet en uitwerking van de toetsen. Omdat er toegewerkt wordt naar

tenminste 1F niveau zou het een meerwaarde hebben als er ook aandacht wordt besteed aan de referentieniveaus.

Concluderend is het oordeel dat wat er getoetst wordt aan begrijpend luisteren en de genoemde inhoudsaspecten middels de toetsen begrijpend luisteren, valide is. Omdat er gekozen is voor meerkeuzenvragen is een beperking waarneembaar van wat er getoetst zou kunnen worden bij de vaardigheid begrijpend luisteren. Sommige fragmenten zijn behoorlijk lang, terwijl maar over een beperkt deel ervan vragen gesteld worden. Daarnaast kan ter overweging genomen worden niet alleen te kiezen voor audiovisuele fragmenten, maar ook een fragment zonder beeld in te zetten: luisteren pur sang.

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

Begripsvaliditeit

In het kader van de begripsvaliditeit is onderzoek gedaan naar:

1. *Unidimensionaliteit.* Op basis van de grafische voorstellingen van de S-toetsen, de R1c toets en de 'c' uit het COTAN systeem kan geconcludeerd worden dat er bij de toets Begrijpend luisteren een beroep gedaan wordt op één onderliggende vaardigheid.
2. *De itemkwaliteit.* De gemiddelde moeilijkheidsgraad van de toetsen ligt op het gewenste niveau en de gemiddelde samenhang tussen de items uitgedrukt in Rit en Rir is goed.
3. *Convergente en divergente validiteit.* Correlatieonderzoek met verschillende andere LVS-onderdelen en de Entreetoets liet zien dat de verwachtingen uit de literatuur werden bevestigd.
4. *Item bias.* Er is onderzoek gedaan naar DIF met betrekking tot afnamemoment en sekse. Bij geen enkel item werd een significant effect gevonden.
5. *Prestaties van relevante subgroepen.* Meisjes scoren enigszins hoger dan jongens. In termen van effectgrootte is er sprake van een verwaarloosbaar effect.

Conclusie:

Op aspect V1. wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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:

In paragraaf 2.3 van de wetenschappelijke verantwoording wordt beschreven hoe proefvoetsingen in de leerjaren groep 4 tot en met groep 7 laat zien dat er sprake is van een bescheiden gemiddelde vaardigheidsgroei. De gemiddelde groei over een vol jaar lijkt steeds kleiner te worden en te stabiliseren.

In de verschillende leerjaren wordt met verschillende toetsen gemeten die met behulp van IRT op één en dezelfde vaardigheidsschaal zijn gebracht.

Conclusie:

Op aspect VA1.1 wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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 in hoofdstuk 4 ingegaan op hoe de resultaten van de toets geïnterpreteerd dienen te worden. Daarin wordt beschreven hoe gekeken kan worden of een leerling zich ontwikkelt in lijn met de gehele populatie. De gerapporteerde score is een vaardigheidsscore en door deze te vergelijken met voorafgaande periodes kan de groei van een leerling in kaart gebracht worden.

In de handleiding worden de leerkrachten gewezen op de betrouwbaarheid van de gerapporteerde scores. Die betrouwbaarheid wordt ook vermeld op het leerlingrapport.

Conclusie:

Op aspect VA1.2 wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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:

In de handleiding zijn registratieformulieren opgenomen voor een leerlingrapport, groepsrapport en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken), in een versie I-V en A-E. Deze kunnen door de leraar handmatig of digitaal worden ingevuld.

Voor ouders is met name het leerlingrapport of het alternatief leerlingrapport informatief. Er is in de Handleiding een apart hoofdstuk opgenomen over het communiceren over de toetsresultaten door leraren met de leerlingen en met de ouders. Daarin worden voorbeelden gegeven hoe met ouders gecommuniceerd kan worden over de resultaten met het onderscheid: kijken naar niveau en kijken naar groei. De wijze waarop de registratieformulieren zijn vormgegeven met de uitleg erbij geven inzichtelijk het niveau en de groei weer. De gemiddelde ouder heeft de toelichting van de leerkracht, in een gesprek, wel nodig om een dergelijk rapport te begrijpen.

Conclusie:

Op aspect I1. wordt de toets Begrijpend Luisteren groep 7 als '**voldoende**' beoordeeld.

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

4. Literatuurlijst

- Berkel, S. van, Engelen, R., Hilde, M., Kamphuis, F., & Zanden, M. van der (2016). *Wetenschappelijke verantwoording Begrijpend luisteren groep 7*. Arnhem: Cito
- Cito (2015). *Begrijpend luisteren groep 7*. Cito Volgstelsel primair onderwijs (LVS). Arnhem: Cito.