

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 toets Begrijpend Luisteren groep 8 is een onderdeel van het Cito Volgsysteem primair onderwijs (tweede generatie toetsen) en is bedoeld voor leerlingen in groep 8 van het primair onderwijs. De toets vormt samen met de LVS toetsen Begrijpend Luisteren voor groep 3 tot en met 7 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 toets in de toets pakketten Begrijpend Luisteren groep 8 van het Cito Volgsysteem primair en speciaal onderwijs is 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 toets Begrijpend Luisteren groep 8 is bedoeld voor leerlingen in groep 8 van het primair en speciaal onderwijs, maar kan 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.

Voor leerlingen die nog maar pas in Nederland verblijven of gehoorproblemen hebben, is de toets ongeschikt. Een leerling moet voldoende taalvaardig in het Nederlands zijn (vergelijkbaar met Nederlandstalige leerlingen in groep 8) voordat de toets Begrijpend Luisteren kan worden afgenomen.

Gebruiksdoel en functie

Doel van de toets Begrijpend Luisteren groep 8 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 geformuleerde 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 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 groepsniveau (hfdst 4),
 - algemene aandachtspunten voor het schoolplan (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
- Één toets:
 - Toets M8
- Afnamekaarten met aanwijzingen voor het afnemen van de toetsen
- Nakijkkaarten
- Antwoordbladen
- Tabellen om de vaardigheidsscore en -niveau te bepalen
- DVD met teksten
- 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 Paulyne Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

In januari 2015 zijn in het kalibratieonderzoek M8 190 items en 20 ankeritems voorgelegd aan 1.848 leerlingen van groep 8. Er is gebruik gemaakt van een gelinkt design. Het gemiddeld aantal leerling antwoorden per item was 395. 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.

Voor de steekproef t.b.v. het kalibratieonderzoek werden scholen geworven op basis van eenzelfde steekproefkader als voor het normeringsonderzoek. Hiermee werd spreiding nagestreefd naar regio, schoolgrootte en stratum, waarmee bij de inrichting van het proefonderzoek reeds rekening werd gehouden met de representativiteit die vereist is bij het daaropvolgende normeringsonderzoek. Een evaluatie van de representativiteit van de steekproef voor het proefonderzoek wordt niet vermeld.

Conclusie:

Er worden adequate gegevens verstrekt over de representativiteit van de kalibratiesteekproef en over de methode van steekproeftrekking. Op aspect S1.1 wordt de toets Begrijpend Luisteren voor groep 8 als '**voldoende**' beoordeeld.

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

Bevindingen:

De 190 items en 20 ankeritems van groep 8 waren verdeeld over 9 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit ongeveer 45 items. Elk item kwam in twee boekjes voor. Het gemiddeld aantal leerlingantwoorden per item was 395. De items zijn gekalibreerd met OPLM en de modelfit is getoetst.

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

Conclusie:

Op aspect S1.2 wordt de toets Begrijpend Luisteren groep 8 als '**voldoende**' beoordeeld.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

De toets is in de periode januari 2016 en begin februari 2016 genormeerd voor afnamemoment M8 (halverwege het schooljaar). Voor dit afnamemoment M8 hebben 940

leerlingen van 41 scholen meegedaan aan het normeringsonderzoek. Voor afnamemoment M8 worden vaardigheidsverdelingen gepresenteerd. 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 asymmetrische niveau indeling A t/m E en de symmetrische niveau indeling I t/m V) worden bepaald.

De geldigheid van de normen is vastgesteld tot en met 2025. Cito monitort periodiek de normering. Mochten er systematische verschuivingen plaatsvinden, dan wordt de normering indien nodig aangepast.

Conclusie:

Op aspect N1.2.1 wordt de toets Begrijpend Luisteren groep 8 als '**voldoende**' beoordeeld.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

Voor het normeringsonderzoek M8 werden scholen geworven na het trekken van een representatieve steekproef, waarbij rekening gehouden werd met verdeling over leerlinggewicht (strata), schoolgrootte, regio, verstedelijking en sekse. De normeringsgroep bestond deels uit herhalingscholen die ook op het afnamemoment E7 hadden meegedaan aan het normeringsonderzoek. Van de aangeschreven herhalingscholen bleek 98% bereid deel te nemen aan normeringsonderzoek M8. Van de scholen uit de aanvullende steekproef bleek slechts 7% bereid deel te nemen. Voor het afnamemoment M8 deden uiteindelijk 940 leerlingen van 41 scholen mee.

Voor wat betreft 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 van voortgezet onderwijs en hoger; 0.3 = beide ouders of de ouder die belast is met de dagelijkse verzorging heeft of hebben een opleiding van maximaal lbo/vbo gehad; 1.2 = één van de ouders heeft een opleiding gehad van maximaal basisonderwijs en de ander een opleiding van maximaal basisonderwijs of maximaal lbo/vbo. Voor elke school wordt het percentage leerlingen met een gewicht afwijkend van 0,0 bepaald. Gebaseerd op deze gewichten zijn vier groepen scholen gevormd. In de steekproef van het normeringsmoment M8 is stratum 2 licht ondervertegenwoordigd en zijn stratum 3 en 4 licht oververtegenwoordigd. De verschillen waren niet significant.

Bij schoolgrootte is onderscheid gemaakt tussen klein en groot. Een kleine school telt minder dan 200 leerlingen. De verschillen tussen de 2 groepen in de steekproef van het normeringsmoment M8 waren niet significant.

Bij regio is uitgegaan van de CBS-indeling naar landsdeel waarbij de regio's noord, oost, west en zuid onderscheiden worden. De scholen in regio oost waren wat ondervertegenwoordigd en de scholen in zuid en noord wat oververtegenwoordigd. Ook hier waren de verschillen niet significant.

Bij verstedelijking (urbanisatiegraad) is gebruik gemaakt van de vijf niveaus die door het CBS worden onderscheiden: zeer sterk, sterk, matig, weinig en niet verstedelijkt. De scholen in matig en weinig verstedelijkt gebied waren wat oververtegenwoordigd en de

scholen in zeer sterk en sterk verstedelijkt gebied waren wat ondervertegenwoordigd. De verschillen waren, opnieuw, niet significant.

Tenslotte werd bij de variabele sekse een tweedeling naar jongens en meisjes gehanteerd. Jongens waren iets ondervertegenwoordigd, maar ook dit verschil was niet significant.

Conclusie:

Op aspect N1.2.2 wordt de toets Begrijpend Luisteren groep 8 als '**voldoende**' beoordeeld.

Betrouwbaarheid

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

Bevindingen:

Om relevante gegevens bij de toets 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.

Conclusie:

Op aspect B1.1. wordt de toets Begrijpend Luisteren groep 8 als '**voldoende**' beoordeeld.

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

Bevindingen:

Voor afnamemoment M8 worden drie betrouwbaarheidsgegevens berekend en vermeld in tabel 5.1 op pagina 45 van de wetenschappelijke verantwoording: standaardmeetfout (2,36), MAcc (0,77) en de gesimuleerde test-hertest betrouwbaarheidscoëfficiënt (0,77). Er wordt verwezen naar het beoordelingssysteem van COTAN waar voor tests die geen zware consequentie voor leerlingen hebben, zoals de toets Begrijpend luisteren volgens de auteurs, een betrouwbaarheidscoëfficiënt tussen 0,70 en 0,80 als 'voldoende' aangemerkt wordt.

Wat betreft de lokale betrouwbaarheid, wordt in tabel 5.2 voor afnamemoment M8 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 ligt boven de 56 procent. Dit is in lijn overigens met de gerapporteerde overall betrouwbaarheidsscores van 0,77.

Conclusie:

Op aspect B1.2. wordt Begrijpend Luisteren groep 8 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:

Opgaven worden als volgt verdeeld naar vaardigheid en inhoudsaspecten:

Begrijpen van gesproken taal:

- Expliciete betekenistoekenning: betekenis woord en woordgroep
- Specifieke inhoudselementen: elementen die expliciet in de tekst aan de orde gesteld worden (o.a. meningen, hoofdgedachte, tijdsperioden, handelingen)
- Eenvoudige expliciete verbanden: verbanden gericht op o.a. vergelijkingen, tegenstellingen)
- Complexe expliciete verbanden (o.a. oorzaak-vervolg, deel-geheel, hoofd en bijzaken, opeenvolgende stappen)

Interpreteren van gesproken taal:

- Impliciete betekenistoekenning: afleiden van de betekenis van het woord
- Inzet van voorkennis: afleiden van informatie uit de tekst (o.a. ontbrekende info, mening, functionele betekenis)
- Globale inhoud: globale inhoud van de tekst (o.a. verbinden, en vergelijken van informatie, dus: hoofdgedachte, hoofdpersoon, doel en publiek herkennen, samenvatten)
- Manier van spreken (klemtoon, tempo, toon)

In groep 8 zijn er per deeltoets 2 kijk-luisterteksten. In totaal zijn er drie deeltoetsen en 33 opgaven. Tweedere van de opgaven zijn gericht op interpreteren en een derde van de opgaven op begrijpen wat ook logisch is gezien wat verwacht mag worden van het vaardigheidsniveau van leerlingen uit groep 8. Opgaven rond begrijpen en interpreteren lopen regelmatig in elkaar over wat logisch is. In de handreiking wordt ook aangegeven dat begrijpen en interpreteren niet zo heel duidelijk van elkaar te onderscheiden zijn. In schema's wordt het aantal vragen weergegeven per inhoudsaspect voor begrijpen en interpreteren. In de wetenschappelijke verantwoording wordt aangegeven hoe interpreteren meer accent krijgt dan begrijpen en wordt inzicht geboden in welke inhoudsaspecten bevraagd worden en in welke frequentie. Daar wordt duidelijk (en blijkt ook tijdens het beluisteren van de fragmenten) dat het accent bij begrijpen en interpreteren met name ligt op eenvoudige expliciete verbanden bij begrijpen en bij interpreteren op afleiden van informatie en hoe informatie met elkaar verbonden moet worden door vragen naar de globale inhoud. Dit accent is voor groep 8 goed gekozen maar zou ook expliciet verwoord/geïllustreerd kunnen worden in de handleiding. Hierdoor kan de leraar ook bij het analyseren van de toets en het handelen nadien deze informatie inzetten.

Er wordt gebruik gemaakt van twee teksttypen: fictie en non-fictie; tekstgenres (soms in combinatie ingezet): Documentaire/verslag; Betoog/reclame; Interview/gesprek, instructie/uitleg en een lied.

Het fragment rond het Tropenmuseum wordt als reclame aangegeven. Het start meer als een documentaire door aan te geven hoe het museum zich door de jaren heen heeft ontwikkeld. Daarna wordt aangegeven wat je allemaal digitaal of in het museum kunt zien.

Met wat kinderen gewend zijn aan reclames op internet of TV, waar ze overtuigd worden om iets te kopen of ergens heen te gaan, komt het als reclame oubollig over. Bij het kijk-luisterfragment rond conditie gaat het vooral om uitleg en niet zozeer om instructie.

De genres film/drama/verhaal en nieuwsbericht ontbreken. Maar ook instructie en reclame komen niet duidelijk naar voren. Hier wordt in de verantwoording al naar verwezen. In de verantwoording wordt aangegeven dat, vanwege de lengte van de toets, niet alle genres aan bod kunnen komen.

In de kijk-luisterfragmenten zijn vooral t.a.v. tekstgenres de combinaties documentaire/verslag, betoog en interview uitgewerkt. Hoewel er maar 6 luisterteksten zijn, zou het van meerwaarde zijn, ook gezien waar leerlingen mee te maken krijgen in hun leefomgeving, om ook een instructietekst en een nieuwsbericht op te nemen. In referentieniveau 2F wordt nog gesproken over discussieprogramma's over vertrouwde onderwerpen. Dat zou in principe ook passen bij deze toets, hoewel dergelijke programma's voor kinderen, zonder een te emotionele lading, niet makkelijk te vinden zijn.

De kijk-luisterfragmenten zijn per fragment max. 7 minuten met als reden dat kinderen hun aandacht kunnen bewaren. De fragmenten hebben een duidelijke opbouw en structuur. Het woordgebruik bestaat uit frequente woorden en woorden die goed kunnen worden afgeleid uit het kijk- en luisterfragment. De lengte van de teksten sluiten aan op niveau 1F (max. 10 minuten). De informatie is herkenbaar geordend en heeft een lage informatiedichtheid. Verbanden worden duidelijk aangegeven. In de luisterteksten zie je kenmerken van 1F maar ook van 2F terug. De teksten sluiten aan bij de leefwereld van de kinderen of staan iets verder weg, maar zijn wel herkenbaar (bijv. Conditie, Tropenmuseum, blindengeleidehond, Katja's Stichting, ecologische viaducten).

Uitgangspunten bij de toetsconstructie waren met name de kerndoelen primair onderwijs en de tussendoelen mondelinge communicatie. Er is ook een koppeling gemaakt met het referentiekader en wel met de vaardigheid 'Luistervaardigheid'.

Zowel bij de kerndoelen als de tussendoelen is begrijpend luisteren één van de aspecten van mondeling taalonderwijs. Je toetst hierdoor één onderdeel namelijk dat leerlingen leren informatie te verwerven uit gesproken taal. Begrijpend luisteren is één van de drie kerndoelen rond mondelinge taal. Het leren argumenteren of je leren uit te drukken (de andere twee kerndoelen) worden niet getoetst. Van de tussendoelen zijn de doelen rond het begrijpen van een verhaal, informatie uit een verhaal halen, afleiden hoofdgedachte, gebruik maken van de structuur van een verhaal en samenvatten herkenbaar in de toetsen. Doelen rond het bewust zijn van de impact van media etc. worden niet getoetst.

Het domein Mondelinge taalvaardigheid bij de referentieniveaus heeft een niveaubeschrijving uitgewerkt rond luistervaardigheid waar bij niveau 1F wordt aangegeven dat een kind kan luisteren naar eenvoudige teksten over alledaagse onderwerpen of onderwerpen die aansluiten bij de leefwereld. Twee van de vier kenmerken van de taakuitvoering richten zich net als de hier beoordeelde Cito toets op het begrijpen en interpreteren. Van de andere twee kenmerken komt het evalueren van de tekst wel aan de orde maar niet bijvoorbeeld gericht op het aantekeningen maken, maar door het kiezen van de beste samenvatting bij de antwoordalternatieven. Dit kan ook niet anders als er alleen met meerkeuzenvragen wordt gewerkt. Het evalueren van de tekst komt niet aan de orde, wat logisch is als voor de uitwerking wordt gekozen om met meerkeuzevragen te werken. De taken bij de niveaubeschrijving in het referentiekader zijn gericht op het

luisteren naar instructies, luisteren als lid van een live publiek en luisteren naar radio en tv en naar gesproken tekst op internet. Bij het luisteren naar tv en radio wordt nadrukkelijk aangegeven dat ook via een vooraf gestelde vraag de info uit het verhaal gehaald kan worden. In de toetsen begrijpend luisteren gebeurt dat frequent. In de Cito-toets begrijpend luisteren wordt alleen naar audiovisuele fragmenten geluisterd (kinderen leven in een beeldcultuur volgens de toetsontwikkelaars) maar niet naar radiofragmenten. Het luisteren naar een live-publiek is vanzelfsprekend geen onderdeel van de toets.

In de wetenschappelijke verantwoording wordt niet duidelijk gemaakt hoe 1F en 2F zich verhouden tot elkaar in de toets. Wel staat in de wetenschappelijke verantwoording (p. 49) dat de uitgangspunten in grote lijnen overeenkomen met wat er in het referentiekader staat. De visie was al geformuleerd voordat het referentiekader Taal verscheen.

Merk op dat de beoordeling van dit aspect zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat verricht is inzake validiteit en dan specifiek begripsvaliditeit.

De toets Begrijpend luisteren 3.0 groep 8 is niet bedoeld voor voorspellend gebruik. Daarmee is de criteriumvaliditeit niet van toepassing. De (psychometrische) begripsvaliditeit wordt uitgewerkt in unidimensionaliteit, itemkwaliteit, convergente en divergente validiteit, in itembias 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 luistervaardigheid kan worden vastgesteld. Dit wordt tevens bevestigd door de nauwkeurigheid van de itemparameterschattingen (voor de toets M8 geldt dat de R_{1c} minder dan anderhalf maal het aantal vrijheidsgraden bedraagt en voor de toets M8 is voor geen enkele opgave de waarde van de constante 'c' groter dan 0,30).

De gemiddelde moeilijkheidsgraden van de items (als criterium voor de itemkwaliteit) voldoen voor M8 (een range van 0,65 – 0,75 en een gemiddelde van 0,73). De gemiddelde R_{it} waarde (ook een criterium voor de itemkwaliteit) voldoet eveneens. Met een gemiddelde van 0,34 is deze te kenschetsen als goed. Geen enkele opgave heeft een lagere R_{it} -waarde dan 0,20.

De constructvaliditeit is uitgewerkt in convergente validiteit door de samenhang te onderzoeken met andere taaltoetsen, te weten Woordenschat en Begrijpend lezen. De correlaties, gecorrigeerd voor attenuatie, waren 0,76, respectievelijk 0,75. De correlaties met niet-semantiche taalonderdelen waren laag ($r=0,41$ voor Spelling niet-werkwoorden en $r=0,10$ voor Technisch lezen – leestempo). De divergente (soortgenoot)validiteit is onderzocht door de samenhang met de toets LOVS-Luisteren 3 (Cito, 1996) te onderzoeken. Dit resulteerde in een hoge correlatie van $r=0,97$.

Het onderzoek naar DIF (Differential Item Functioning) over afnamemoment en sekse per item liet geen enkel item significante verschillen zien op het 1%-niveau.

Wat betreft de verschillen tussen relevante subgroepen scores meisjes enigszins hoger dan jongens. In termen van effectgrootte is er voor M8 sprake van een klein effect ($d = 0,19$).

Conclusie:

Overall wordt geconstateerd dat wat er getoetst wordt aan begrijpend luisteren en de

genoemde inhoudsaspecten op de toets begrijpend luisteren groep 8 valide is. Omdat er gekozen is voor meerkeuzenvragen is een beperking waarneembaar van wat er zou kunnen worden getoetst binnen het domein begrijpend luisteren. Op aspect V1. wordt Begrijpend Luisteren groep 8 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:

Het algemene (inhoudelijke) uitgangspunt voor de toets Begrijpend luisteren groep 8 is dat de (latente) vaardigheid begrijpend luisteren kan worden opgevat als een unidimensioneel continuüm en dat elke leerling kan worden voorgesteld als een punt op dit continuüm. Uit het kalibratieonderzoek blijkt dat de items passen bij het gehanteerde IRT model en dat het model ook past voor de toets M8 als geheel. Dit betekent dat er sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden. Afhankelijk van het aantal items dat een leerling goed maakt, wordt er een vaardigheidsscore toegekend. Meisjes scoren iets beter dan jongens.

In paragraaf 2.3 van de wetenschappelijke verantwoording wordt aan de hand van de proefvoetsingen beschreven hoe de vaardigheid in Begrijpend Luisteren zich in de leerjaren groep 4 tot en met groep 8 ontwikkelt. Het blijkt dat de gemiddelde toename in vaardigheid in Begrijpend luisteren over de jaren gezien steeds kleiner wordt (de gemiddelde toename is steeds aanmerkelijk kleiner dan de spreiding in vaardigheid binnen de groep op enig afnamemoment). De enige uitzondering hierop is groep 8. De eenmalige flinke stijging in vaardigheidsgroei in groep 8 zou te maken kunnen hebben met de 'mentale sprong voorwaarts' die veel leerlingen tijdens de beginnende puberteit lijken te maken.

Conclusie:

Op aspect VA1.1 wordt de toets Begrijpend luisteren groep 8 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 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 rapportagemogelijkheden.

Conclusie:

Op aspect VA1.2 wordt de toets Begrijpend luisteren groep 8 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 in een versie I-V en een versie A-E voor een leerlingrapport, groepsrapport, groepsoverzicht en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken en de groei en vaardigheid in de tijd wordt weergegeven). Deze kunnen door de leraar handmatig of digitaal worden ingevuld. De wijze waarop de registratieformulieren zijn vormgegeven met de uitleg erbij geven inzichtelijk de vaardigheidsgroei en vaardigheidsniveau weer van een individuele leerling en van een groep leerlingen.

In hoofdstuk 7 van de handleiding (Communiceren over toetsresultaten) wordt beschreven hoe er met de ouders over de toetsresultaten kan worden gecommuniceerd. 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. Daarnaast wordt de leraar gewezen op misverstanden die zich bij ouders kunnen voordoen bij de interpretatie van de niveau-indelingen. Alleen uitgaan van een niveau-aanduiding geeft onvoldoende inzicht in de ontwikkeling van leerlingen. Ook moeten de leraren aan ouders het verschil tussen methode-onafhankelijke en methodegebonden toetsen (beheersingstoetsen gericht op de leerstof die is aangeboden) duidelijk maken en de ouders erop wijzen dat de verschillende toetsen de leerling anders (kunnen) beoordelen. Voor ouders is met name het leerlingrapport of alternatief leerlingrapport informatief. De gemiddelde ouder heeft de toelichting van de leerkracht, in een gesprek wel nodig om een dergelijk rapport te begrijpen. Zijn de ouders er vanaf het begin van de schoolloopbaan van hun kind bekend mee dan zal dit minder het geval zijn.

Conclusie:

Op aspect I1. wordt de toets Begrijpend luisteren groep 8 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, Hilde, M., Kamphuis, F., & Zanden, M. van der (2017). *Wetenschappelijke verantwoording Begrijpend luisteren groep 8*. Arnhem: Cito
- Cito (2016). *Begrijpend luisteren groep 8*. Cito Volgstelsysteem primair onderwijs (LVS). Arnhem: Cito.