

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 van de betreffende toets 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 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. AVI is een onderdeel van het Cito Volgsysteem primair en speciaal onderwijs en is bedoeld voor leerlingen in groep 3 tot en met halverwege groep 8. De AVI-toets is een papieren toets, bedoeld om het vaardigheidsniveau en de ontwikkeling van leerlingen op het gebied van technisch lezen in kaart te brengen. Onderstaande beschrijving en beoordeling is gebaseerd op de Handleiding (Cito, 2017) en op de Wetenschappelijke verantwoording (van Til, Kamphuis, Keuning, Gijsel, & de Wijs, 2018).

Meetpretentie

Binnen het leesonderwijs op de basisschool wordt een onderscheid gemaakt tussen technisch lezen en begrijpend lezen. Bij technisch lezen gaat het erom of leerlingen in staat zijn om geschreven woorden correct en vlot te verklanken. Bij begrijpend lezen staat het begrip van geschreven teksten centraal. Technisch lezen is geen doel op zich, maar wordt gezien als een voorwaardelijke activiteit voor begrijpend lezen. Het neemt in het basisonderwijs, in ieder geval in de onder- en middenbouw, een volwaardige plaats in op het lesrooster. De AVI-toets is bedoeld om vast te stellen wat het vaardigheidsniveau van leerlingen is op het gebied van technisch lezen en hoe deze vaardigheid zich ontwikkelt in de loop der jaren.

Doelgroep

De AVI-toets is bedoeld voor leerlingen van groep 3 tot en met halverwege groep 8 van het primair en speciaal onderwijs. Voor leerlingen in het speciale onderwijs zijn geen aparte normen vastgesteld. Wel mogen leerlingen met een visuele beperking gebruikmaken van vergrote leeskaarten (op A3-formaat) en mogen zij langer over het lezen doen dan andere leerlingen. Voor leerlingen die de Nederlandse taal onvoldoende beheersen en voor leerlingen met ernstige spraakmoeilijkheden is de AVI niet geschikt.

Gebruiksdoel en functie

Het hoofddoel van AVI is tweeledig: enerzijds het in kaart brengen van het vaardigheidsniveau en anderzijds het zichtbaar maken van de ontwikkeling van leerlingen op het gebied van technisch lezen. Daarnaast biedt AVI diagnostische informatie, zodat inzicht geboden kan worden in het leesgedrag van de leerling.

Inhoudelijke theoretische inkadering:

Hoewel technisch lezen in de vakliteratuur niet altijd op precies dezelfde manier wordt omschreven, is het mogelijk een min of meer universele, formele definitie van de vaardigheid te geven. Deze houdt in dat technisch lezen gelijk staat aan het correct en

vlot ontsleutelen van geschreven woorden, al dan niet in context, en het herkennen van deze woorden als dragers van betekenis.

Inhoud van het toetspakket

Het toetspakket AVI bestaat uit de volgende documenten:

- Handleiding, deze bevat informatie over:
 - de afname van de toets (hfdst. 2),
 - rapportages (hfdst. 3),
 - interpretatie en analyse 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
- Toetsmateriaal:
 - afname-instructies AVI
 - score-instructies AVI
 - leeskaarten
 - scoreformulieren

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), mevrouw Pauly K. Berding-Oldersma MSc (secretaris) en mevrouw P. Jennifer Roubiës MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen

In schooljaar 2015/2016 zijn data verzameld met een cross-sectioneel design, waarin leerlingen uit groep 3 tot 7 op twee momenten in het jaar getoetst werden (een M en een E meting) en leerlingen uit groep 8 alleen in januari/februari (M-meting). Elke jaargroep werd opgedeeld in 12 subgroepen van minimaal 45 leerlingen. De nieuwe AVI toetsen werden opgedeeld in 3 subtoetsen/leeskaarten. Via een gelinkt design werden de subtoetsen afgenomen bij de subgroepen waarbij elke subtoets door ten minste 6 subgroepen werd gemaakt. Het totaal aantal leerlingen per leeskaart varieerde van 480 tot 801 en het totaal aantal leerlingen varieerde per toetsmoment van 572 tot 789. Het benodigde aantal waarnemingen (450 per leeskaart, 540 per toetsmoment) werd ruimschoots gehaald.

Informatie is verzameld over regio, urbanisatiegraad, schooltype en sekse. Vervolgens zijn normsteekproeven genomen voor M3, E3, ... , M8. De steekproeven zijn vergeleken met populatiegegevens die zijn verkregen via DUO en CBS. De steekproeven voor de verschillende groepen lijken veel op elkaar, omdat ze op dezelfde scholen genomen zijn.

- De zuidelijke provincies en plattelandsgebieden zijn oververtegenwoordigd.
- Wat het schooltype betreft zijn scholen met weinig achterstandsl leerlingen oververtegenwoordigd.
- Wat sekse betreft waren er geen afwijkingen.

Met regressieanalyse is uitgezocht of er een systematisch effect was tussen de samenstelling van de normsteekproef en de vaardigheid. Alhoewel er voor specifieke groepen een effect lijkt te bestaan, is dit effect niet systematisch. Er lijkt eerder sprake te zijn van toeval.

Conclusie

Op aspect S1.1 wordt de toets AVI als '**voldoende**' beoordeeld.

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

Bevindingen

Er was sprake van een onvolledig maar gelinkt design, waardoor de scores op de nieuwe AVI toetsen wat technische leesvaardigheid en tijd vergeleken konden worden met die van de oude AVI toetsen. Daarvoor moeten de verdelingen van de leerlingen wat betreft frustratie-, instructie- en beheersingsscores voor ten minste 90 procent overeenkomen. De classificatie ten aanzien van frustratie- en beheersingsniveau voldoen aan deze eis. De

classificatie van het instructieniveau scoort daar met 89,1 procent net onder. In totaal wordt 96,5 procent identiek geclassificeerd.

Conclusie

Op aspect S1.2 wordt de toets AVI als '**voldoende**' beoordeeld.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen

Het totaal aantal leerlingen per leeskaart varieerde van 480 tot 801 en het totaal aantal leerlingen varieerde per toetsmoment van 572 tot 789 voor de normeringssteekproeven. Het benodigde aantal waarnemingen (450 per leeskaart, 540 per toetsmoment) werd ruimschoots gehaald.

Conclusie

Op aspect N1.2.1 wordt de toets AVI als '**voldoende**' beoordeeld.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen

De representativiteit van de steekproeven is besproken bij punt S1.1. Hier werd reeds geconstateerd dat deze niet representatief zijn, maar dat dit geen impact had.

Conclusie

Op aspect N1.2.2 wordt de toets AVI als '**voldoende**' beoordeeld.

Betrouwbaarheid

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

Bevindingen

Voor de leessnelheid en de leesfouten is er de beschikking over een zuivere schatter, waardoor het vrij makkelijk is om een schatting van de betrouwbaarheid te maken. In de wetenschappelijke verantwoording wordt in detail beschreven wat de verschillende stappen zijn om tot een schatting van de betrouwbaarheid te komen.

Conclusie

Op aspect B1.1. wordt de toets AVI als '**voldoende**' beoordeeld.

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

Bevindingen

De betrouwbaarheid van de schatting van de leessnelheid en van de leesfouten moet worden geschat. Voor de verschillende leeskaarten varieert de test-hertest

betrouwbaarheid van de leessnelheid van 0,905 tot 0,973. Dit is zeker hoog te noemen. Voor de leesfouten varieert de test-hertest betrouwbaarheid van 0,565 tot 0,791. Vooral voor de hogere groepen wordt de COTAN eis van een minimale betrouwbaarheid van 0,70 voor een voldoende score niet gehaald. De reden is dat er in de hogere groepen nauwelijks leesfouten worden gemaakt.

Daarnaast is onderzoek gedaan naar het percentage misclassificaties. Ten aanzien van de leestijden valt op dat het percentage misclassificaties voor het instructieniveau varieert van 28,2% tot 52,9%. Voor de leesfouten zijn er vooral veel misclassificaties op het instructie- en het beheersingsniveau. Voor deze beide niveaus varieert het percentage van 37% tot 74,3%.

De marginal classification accuracy (MCA) voor zowel de leessnelheid als de leesfouten varieert van 84,2 tot 93,6 wat wel weer hoog is. In de technische handleiding wordt een opmerking toegevoegd die expliciet aangeeft dat de toetsen in M6, E3, E6 en M7 de leerlingen op instructieniveau minder nauwkeurig classificeren, zodat daar kritisch met de toetsresultaten moet worden omgegaan.

Conclusie

Op aspect B1.2. wordt de toets AVI 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

Inhoudsvaliditeit

De AVI toetsen zijn onafhankelijk van de verschillende onderwijsmethoden opgesteld. De teksten zijn speciaal geschreven door professionele kinderboekenauteurs die zich voor elk AVI-niveau gebaseerd hebben op een set tekstkenmerken. Na feedback van vakexperts en leerkrachten en na analyse van de try-out en de proeftoetsen zijn de kaarten aangepast. Na herziening voldeden de kaarten aan de vooraf vastgestelde tekstkenmerken. Met deze aanpak is de voorliggende set teksten in vormgeving, inhoudelijk, en qua gevraagde actie van leerlingen een verbetering en een meer herkenbare context voor leerlingen dan de vorige generatie AVI toetsmateriaal.

Begripsvaliditeit

Technische leesvaardigheid wordt gedefinieerd in termen van leessnelheid en leesnauwkeurigheid. De toetsen worden daarom onder lichte tijdsdruk afgenomen en de teksten zijn van een oplopende moeilijkheid. Om de begripsvaliditeit verder te onderbouwen, is er ook een MTMM matrix opgesteld met als soortgenootinstrumenten de DMT en de EMT. Als niet-soortgenootinstrumenten zijn de toetsen Begrijpend Lezen en Woordenschat gebruikt. De correlaties van de AVI toetsen met de DMT en de EMT waren hoog en die met de toetsen Begrijpend Lezen en Woordenschat waren matig. Omdat bij de twee laatste toetsen deels dezelfde vaardigheden gebruikt worden als bij de AVI toetsen, kunnen deze correlaties gezien worden als bewijs voor begripsvaliditeit.

Tenslotte is er nog gekeken of er effecten zijn van sekse, leerlinggewicht en dyslexie. Alleen bij dyslexie werd een klein effect gevonden. Dit komt overeen met de verwachtingen op basis van de literatuur.

Conclusie

Op aspect V1. wordt de toets AVI 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

Voor het modelleren van de leesfouten is gebruik gemaakt van het Poisson model en voor het modelleren van de leestijden van een RPCM. Alle AVI kaarten zijn met deze modellen gekalibreerd op zowel een vaardigheidsschaal voor de leestijden als een vaardigheidsschaal voor de leesfouten. Een grafische analyse laat zien dat de beide modellen op de data passen. Model fit is verder niet analytisch onderzocht.

Conclusie

Op aspect VA1.1 wordt de toets AVI 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 en 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.

Conclusie

Op aspect VA1.2 wordt de toets AVI als '**voldoende**' beoordeeld.

Inzicht in leervorderingen

I1.2 Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/voogden/verzorgers begrijpelijk is?

Bevindingen

De resultaten van de AVI toetsen kunnen zowel handmatig als digitaal verwerkt worden in leerlingrapporten en groepsrapporten. Voor ouders is met name het leerlingrapport of alternatief leerlingrapport informatief, omdat deze rapporten van hun kind individueel de vaardigheid en de groei 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.

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 veel gestelde vragen behandeld die weliswaar voor de leraren bestemd zijn, maar waar de antwoorden voor een deel ook informatief zijn tijdens bijvoorbeeld de tienminutengesprekken.

Conclusie

Op aspect I1.2 wordt de toets AVI 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	I1.2	Voldoende

4. Literatuurlijst

- Alma van Til, Frans Kamphuis, Jos Keuning, Martine Gijssel, Anja de Wijs (2018). *Wetenschappelijke verantwoording LVS-toetsen AVI*. Arnhem: Cito.
- Cito (2017). *AVI*. Arnhem: Cito.