

Wetenschappelijke verantwoording LVS-toetsen AVI

Alma van Til, Frans Kamphuis, Jos Keuning, Martine Gijzel en
Anja de Wijs

cito.nl



Wetenschappelijke verantwoording LVS-toetsen AVI

Alma van Til, Frans Kamphuis, Jos Keuning, Martine Gijzel, Anja de Wijs

© Cito B.V. Arnhem (2018)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Voorwoord

Bij de totstandkoming van de AVI-toets waren een groot aantal personen en instanties betrokken. Een aantal ervan willen we op deze plaats bedanken. Dat zijn in de eerste plaats de scholen die hebben deelgenomen aan de proeftoets en het normeringsonderzoek. In de tweede plaats gaat onze dank uit naar de kinderboekenauteurs die de teksten voor de AVI-kaarten voor hun rekening hebben genomen. Dit zijn:

- Annemarie Dragt
- Thea Dubelaar
- Merel Leene
- Jan Paul Schutten

Ten slotte willen we onze dank uitspreken naar de Nederlandse en Vlaamse vakexperts die ons in verschillende stadia van het project hebben geadviseerd. Dit zijn de volgende personen:

- Wim van den Broeck (CODE-Thomas Moore)
- Maud van Druenen (Expertisecentrum Nederlands)
- Thoni Houtveen (Hogeschool Utrecht)
- Chruf van Kempen (Samenwerkingsverband Waterland)
- Luuk Koning (PRAVOO)
- Karin van de Mortel (CPS)
- Marjolein Noé (Thomas More)
- Ron Oostdam (Universiteit van Amsterdam/Hogeschool van Amsterdam)
- Anneke Smits (Hogeschool Windesheim)
- Chris Struiksma (Stichting Dyslexie Nederland/Nederlands Kwaliteitsinstituut Dyslexie)
- Hanneke Wentink (Radboud in'to Languages)

Inhoud

Voorwoord 3

1 Inleiding 7

2 Uitgangspunten van de toetsconstructie 9

- 2.1 Meetpretentie 9
- 2.2 Doelgroep 9
- 2.3 Gebruiksdoel 9
- 2.4 Theoretische inkadering 11

3 Beschrijving van de toets 17

- 3.1 Opbouw en structuur van de toets 17
- 3.2 Inhoudsverantwoording 18
 - 3.2.1 Uitwerking domeinbeschrijving 19
 - 3.2.2 Totstandkoming leeskaarten 19
 - 3.2.3 Gerealiseerde leeskaarten 21
 - 3.2.4 Screening en try-out 23
 - 3.2.5 Proeftoetsing 23

4 Kalibratie en normering 25

- 4.1 Dataverzameling 25
- 4.2 Representativiteit van de normeringssteekproef 29
- 4.3 Beschrijving van het meetmodel 37
 - 4.3.1 Modelleren leesfouten 37
 - 4.3.2 Modelleren leestijden 39
- 4.4 Kalibratie 41
- 4.5 Normering 44

5 Betrouwbaarheid en meetnauwkeurigheid 47

- 5.1 Betrouwbaarheid 47
- 5.2 Nauwkeurigheid 50

6 Validiteit 57

- 6.1 Inhoudsvaliditeit 57
- 6.2 Begripsvaliditeit 57
- 6.3 Convergente en divergente validiteit 59
- 6.4 Verschillen tussen relevante subgroepen 61

7 Samenvatting 63

Literatuurlijst 65

Bijlagen 69

- 1 Classificatietabellen van leerlingen bij de oude en nieuwe AVI-toets 70
- 2 Classificatie-/misclassificatietabellen voor de leestijden op de AVI-toetskaarten 76
- 3 Classificatie-/misclassificatietabellen voor de leesfouten op de AVI-toetskaarten 78
- 4 Classificatie-/misclassificatietabellen voor de combinatie van leesfouten en leestijden op de AVI-leeskaarten 80

1 Inleiding

Deze wetenschappelijke verantwoording heeft betrekking op de AVI-toets. De afkorting AVI staat voor Analyse Van Individualiseringsvormen. In het onderwijs wordt altijd alleen de afkorting gebruikt, wij sluiten in deze verantwoording (en in het toetsmateriaal) bij dit gebruik aan. De AVI-toets is bestemd voor groep 3 tot en met halverwege groep 8 van het primair en speciaal onderwijs. 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.

Deze wetenschappelijke verantwoording biedt samen met de inhoud van het toetspakket AVI alle informatie die nodig is voor een snelle en efficiënte beoordeling van de kwaliteit van het meetinstrument.

Het genoemde materiaal maakt een beoordeling van de AVI-toets mogelijk op de volgende aspecten:

- uitgangspunten van de toetsconstructie;
- kwaliteit van het toetsmateriaal;
- kwaliteit van de handleiding;
- normering;
- betrouwbaarheid;
- validiteit.

Het laatste aspect betreft alleen begripsvaliditeit en géén criteriumvaliditeit. Omdat de toetsen van het Cito Volgstelsel primair en speciaal onderwijs niet bedoeld zijn voor 'voorspellend gebruik' is criteriumvaliditeit niet van toepassing.

Deze verantwoording heeft betrekking op een nieuwe uitgave van de AVI-toets. De vorige uitgave (Cito, 2009), is voorgelegd aan de Cotan in 2010 en kreeg het oordeel 'goed' toegekend voor alle van toepassing zijnde beoordelingscriteria. Met een nieuwe uitgave van de AVI-toets zijn de toetsinhoud en de normering weer actueel. Daarnaast wordt met de nieuwe uitgave voorkomen dat de inhoud van de leeskaarten overbekend raakt.

Het voorliggende document heeft met name betrekking op de uitgangspunten van de constructie (de hoofdstukken 2 en 3), de normen (hoofdstuk 4), de betrouwbaarheid (hoofdstuk 5) en de validiteit (hoofdstuk 6) van de AVI-toets. De kwaliteit van het toetsmateriaal en de handleiding is te bepalen door kennis te nemen van de inhoud van het toetspakket dat bestaat uit toetsmappen, het Computerprogramma LOVS en aanvullende bestanden die zijn geplaatst op Cito Portal.

2 Uitgangspunten van de toetsconstructie

2.1 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 hoe leerlingen technisch kunnen lezen en hoe deze vaardigheid zich ontwikkelt in de loop der jaren.

2.2 Doelgroep

De AVI-toets is bedoeld voor leerlingen in groep 3 tot en met halverwege groep 8 van het primair en speciaal onderwijs. Voor leerlingen in het speciaal 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 (de leestijdnormen zijn voor leerlingen met een visuele beperking verruimd).

Voor anderstalige leerlingen die de Nederlandse taal nog onvoldoende beheersen, is de AVI-toets ongeschikt. Leerlingen moeten het Nederlands voldoende beheersen om de kaarten te kunnen lezen. Ook voor kinderen met ernstige spraakmoeilijkheden is de toets niet geschikt, omdat leerlingen de teksten hardop moeten lezen. Een verkeerde uitspraak leidt dan onterecht tot fouten en geeft een vertekend beeld van de technische leesvaardigheid van de leerlingen.

Er zijn twee uitgaven van de AVI-toets beschikbaar: een uitgave voor het onderwijs en een uitgave voor zorgaanbieders. De uitgave voor het onderwijs kan gebruikt worden door (groeps)leerkrachten en intern begeleiders. Daarnaast kunnen schooldirecteuren en schoolbesturen gebruikmaken van de resultaten van deze toets om het onderwijs te evalueren. De uitgave voor zorgaanbieders kan worden gebruikt door behandelaars in de gezondheidszorg, zoals orthopedagogen, psychologen of logopedisten. De twee uitgaven worden gekenmerkt door een verschillende toetsinhoud, de normering en rapportage is daarentegen vergelijkbaar.

2.3 Gebruiksdoel

Het hoofddoel van de AVI-toets 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 de AVI-toets diagnostische informatie, zodat inzicht geboden kan worden in het leesgedrag van de leerling. De verschillende doelen worden hieronder toegelicht.

Niveaubepaling

Het AVI-niveau, zoals bepaald met de AVI-toets, kan gebruikt worden om een indicatie te geven van het technisch leesniveau van een leerling. Daarnaast kan de leestechische moeilijkheid van een tekst ook worden aangeduid met een AVI-niveau. Door beide AVI-niveaus te combineren kunnen leesboeken geselecteerd worden met een moeilijkheid die afgestemd is op de leestechiek waarover een leerling beschikt.

Bij de vorige uitgave van de AVI-toets zijn 12 AVI-niveaus gedefinieerd: AVI-start, AVI-M3, AVI-E3, AVI-M4, AVI-E4, AVI-M5, AVI-E5, AVI-M6, AVI-E6, AVI-M7, AVI-E7 en AVI-Plus. Voor al deze AVI-niveaus geldt dat

er tekstkenmerken zijn vastgelegd (zie hoofdstuk 3) die specificeren waar een tekst aan moet voldoen om in aanmerking te komen om te worden aangeduid met een bepaald AVI-niveau. Bij de constructie van de AVI-toets zijn de tekstkenmerken als uitgangspunt genomen. Vervolgens zijn voor elke AVI-niveau, met uitzondering van AVI-start, prestatie-eisen bepaald. Die prestatie-eisen bestaan uit grenswaarden voor tijd en fouten. Aan de hand van de grenswaarden kan worden bepaald of een leerling een leeskaart op beheersings-, instructie-, of frustratieniveau heeft gelezen.

Het beheersingsniveau houdt in dat de tekst vloeiend en met relatief weinig fouten wordt gelezen. Als leerlingen zelfstandig willen lezen, sluiten teksten op dit niveau goed aan. Door het lezen van teksten op beheersingsniveau kunnen leerlingen hun technische leesvaardigheid onderhouden en ook een beetje uitbreiden, omdat de leeshandeling verder wordt geautomatiseerd, wat de leessnelheid ten goede komt. Als een leeskaart op instructieniveau wordt gelezen, dan wordt de tekst langzamer en/of met meer fouten gelezen dan op beheersingsniveau. Als een leerkracht of een zorgaanbieder het technisch lezen wil oefenen met een leerling, dan sluiten teksten op dit niveau goed aan. Het frustratieniveau houdt ten slotte in dat de tekst eigenlijk te moeilijk is voor een leerling. Dit betekent overigens niet dat het leerlingen verboden is om leesboeken uit te kiezen die boven het AVI-niveau van een leerling liggen. Als een leerling graag een boek wil lezen omdat hij of zij dat leuk vindt, is het niet erg als dit iets te moeilijk is.

De prestatie-eis voor het beheersingsniveau is bij de vorige uitgave van de AVI-toets zo gekozen dat deze aansloot bij het technisch leesniveau van de groep gemiddelde leerlingen op een bepaald afnamemoment. Zo verwijst AVI-niveau M3 naar de gemiddelde technische leesvaardigheid in het midden van groep 3. Als Tim de leeskaart AVI-E5 op beheersingsniveau heeft gelezen, komt zijn technische leesvaardigheid overeen met die van de groep gemiddelde leerlingen einde groep 5. Het maakt daarbij niet uit in welke groep Tim zelf zit. Als Tim in het midden van groep 4 zit, dan leest hij bovengemiddeld goed. Zou hij in het midden van groep 6 zitten, dan scoort hij onder het gemiddelde. De prestatie-eis voor het instructieniveau was afgestemd op het niveau van de 10 à 25 procent zwakst presterende leerlingen.

Van de 12 AVI-niveaus zijn er 2 niveaus met een afwijkende aanduiding: AVI-start betreft het niveau van leerlingen die nog maar net begonnen zijn met het leren verklanken van woorden en zinnen. Omdat deze leerlingen aan het begin van het proces van het leren lezen staan, heeft het geen zin om dit niveau met behulp van de AVI-toets vast te stellen. Als een leerling de leeskaart AVI-M3 op frustratieniveau maakt, wil dit zeggen dat de leerling het niveau AVI-start heeft. AVI-Plus is het hoogste niveau en geeft aan dat leerlingen in principe voldoende vaardig zijn om teksten te verklanken die worden aangeboden op de middelbare school.

De normering van de huidige AVI-toets is gebaseerd op de vorige uitgave van de toets. De prestatie-eisen die ten grondslag liggen aan de grenswaarden voor frustratie, instructie en beheersing, komen dus overeen met die van de vorige AVI-toets. Ter controle zijn de individuele scores van alle leerlingen in een landelijke referentiegroep (zie paragraaf 4.2 voor de beschrijving van de referentiegroep) langs beide dimensies (tijd en fouten) gerangschikt, om na te gaan of de prestatie-eis voor de beheersingsniveaus nog correspondeerden met het niveau van de groep gemiddelde leerlingen. Dit bleek inderdaad het geval te zijn. Ook de prestatie-eis voor de instructieniveaus blijkt nog steeds overeen te komen met het niveau van de 10 à 25 procent zwakst presterende leerlingen.

Progressiebepaling

Het volgen van leerlingen in hun groei, ook wel aangeduid als progressiebepaling, is een van de belangrijkste functies van het Cito Volgstelsel primair en speciaal onderwijs (LVS). De toetsen van het LVS geven de leerkracht (en ouders en leerlingen zelf) informatie over de ontwikkeling van de vaardigheden van de leerlingen, individueel en als groep, gedurende (vrijwel) de gehele basisschoolperiode. De toetsen geven antwoord op vragen als: is er sprake van vooruitgang, achteruitgang of van stabilisering? Is de vooruitgang – gelet op de gemiddelde vooruitgang in de populatie – volgens verwachting?

De oplopende AVI-niveaus geven oplopende vaardigheidsniveaus weer. Daardoor is het heel eenvoudig om uitspraken te doen over de ontwikkeling van de technische leesvaardigheid van een leerling. Als Mia op afnamemoment E3 een AVI-niveau M4 behaalt, en op afnamemoment M4 een AVI-niveau E4, dan kunnen we er zeker van zijn dat ze vooruitgegaan is. Als Mia op beide afnamemomenten hetzelfde AVI-niveau behaalt, is ze niet (meetbaar) vooruitgegaan.

Diagnostiek

Door afname van de AVI-toets wordt inzicht verkregen in de leessnelheid en de accuratesse van leerlingen. Doorgaans is de leessnelheid voor de technische leesvaardigheid van zowel goede als zwakke lezers de bepalende factor. Zwakke technisch-lezers onderscheiden zich met name door een geringe leessnelheid. Uit de individuele analyse van de leesprestaties van laag scorende, dus doorgaans langzaam lezende leerlingen kan echter blijken dat deze niet alleen moeite hebben met het tempo maar ook met de accuratesse waarmee gelezen wordt. De leerkracht kan dan nagaan welke leesfouten de leerling maakt. Door op de scoreformulieren bij de afgenomen kaarten het type fouten te noteren, krijgt de leerkracht meer zicht op het leesgedrag van de leerling. Maakt hij fouten in het verklanken van woorden of leest hij nog vaak spellend? Slaat hij woorden over of lukt het hem niet om een woord te verklanken? Individuele leerlingen die steeds dezelfde soorten leesfouten maken, zullen wellicht baat hebben bij extra instructie en gerichte oefeningen. Let wel: de mogelijkheid om patronen te signaleren in gemaakte fouten staat geheel los van de niveau- en progressiebepaling en is niet meegenomen in de kalibratie- en normeringsonderzoeken. Het overzicht op het scoreformulier heeft dan ook geen wetenschappelijke status of pretentie en is bedoeld als extra handreiking aan de gebruiker.

2.4 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.

Hoe leren leerlingen in technisch lezen?

De basis voor de ontwikkeling van de technische leesvaardigheid ligt in de voor- en voerschoolse periode (Verhoeven & Aarnoutse, 2000). In een omgeving waarin jonge kinderen op betekenisvolle wijze met geschreven taal in aanraking komen, kunnen zij de functie daarvan al vroeg leren kennen. Hun lees- en schrijfgedrag is echter nog slechts oriënterend; vaak bootsen zij lees- en schrijfgedrag na. Veel kleuters leren hun eigen naam en enkele andere bekende woorden. Ook kennen zij een aantal letters van het alfabet en kunnen ze die aanwijzen in woorden. Sommige kinderen komen zelfs zo ver dat zij min of meer spontaan woorden synthetiseren.

De ervaring leert echter dat veel jonge kinderen moeite hebben met het doorzien van ons schriftsysteem. Deze kinderen wordt, doorgaans in jaargroep 3, in een gestructureerd curriculum stap-voor-stap het inzicht in het alfabetische principe bijgebracht. Daarin wordt in de regel uitgegaan van twee fasen. In de eerste fase leren kinderen dat woorden zijn opgebouwd uit klanken en dat grafemen die klanken representeren. In een tijdsbestek van gemiddeld vier maanden leren zij hoe de klanken van onze taal door middel van letters kunnen worden weergegeven. Zo ontwikkelen zij op systematische wijze een fonemisch bewustzijn, leren zij de relatie tussen letters en klanken, en leren zij van daaruit woorden te lezen (decoderen) en te spellen (coderen). Deze fase beperkt zich tot de zogeheten klankzuivere woorden met een eenvoudige mkm-structuur (medeklinker-klinker-medeklinker, bijvoorbeeld: *mes* en *pan*). In de daaropvolgende fase leren kinderen de elementaire leeshandeling te versnellen en uit te breiden naar woorden met medeklinkercombinaties en meerlettergrepige woorden (vgl. Wentink, 1997).

Doordat kinderen de eerder verworven vaardigheden in steeds hoger tempo leren toepassen, kunnen zij met toenemend gemak eenvoudige teksten lezen. Op het eind van jaargroep 5 kunnen zij ongeveer 3000 woorden vlot lezen en begrijpen. (Hun mondelinge woordenschat bedraagt dan zo'n 9000 woorden.)

Onder invloed van het gegeven dat kinderen de relaties tussen grafemen, spraakklanken en woorden steeds sneller doorzien, krijgen leesprocessen steeds meer een automatisch verloop. Kinderen bouwen hun leesvaardigheid vanaf dat moment min of meer zelfstandig verder uit. Daarbij geldt dat het lezen van onbekende woorden ook van de geoefende lezer bewuste aandacht vraagt. Het aantal bekende woorden neemt in de loop van het basisonderwijs echter steeds verder toe en de leerlingen kunnen tegen het einde van de basisschool ongeveer evenveel woorden in geschreven vorm lezen en begrijpen als die zij kennen in gesproken vorm.

Het onderwijs in technisch lezen vindt plaats aan de hand van een lesmethode. Educatieve uitgevers brengen elk hun eigen methode op de markt; scholen zijn vrij om uit het aanbod te kiezen.

De AVI-toets is een methodeonafhankelijk instrument voor het meten van de technische leesvaardigheid. Het maakt voor de afname van de AVI-toets niet uit welke lesmethode op school gehanteerd wordt.

Welke processen spelen bij technisch lezen een rol?

Bij het technisch lezen spelen twee processen een essentiële rol: decoderen en woordherkenning. Decoderen, ook wel ontsleutelen of verklanken genoemd, is het omzetten van een visuele code (geschreven woorden) in een klankcode ('spraakklanken'). De aanhalingstekens rond het woord spraakklanken zijn bedoeld om aan te geven dat decoderen niet per definitie neerkomt op het hardop uitspreken van woorden; het verklanken van woorden kan hoorbaar of niet-hoorbaar gebeuren. Bij het decoderen gaat het om het leggen van verbindingen tussen woorden in hun orthografische vorm (i.e. de manier waarop woorden gespeld worden) en de daarmee geassocieerde klankvorm. De sterkte van die verbindingen, en de snelheid en nauwkeurigheid waarmee ze gelegd kunnen worden, bepalen iemands decodeervaardigheid. Het aanvankelijk (technisch-)leesonderwijs is er dan ook op gericht om leerlingen stabiele en vlot beschikbare orthografie-klankassociaties bij te brengen.

Woordherkenning betreft het activeren van inhoudelijke informatie om aan het woord betekenis toe te kennen. Bij de start van het aanvankelijk leesonderwijs in jaargroep 3 van de basisschool hebben leerlingen al een behoorlijke (mondelinge) woordenschat. Het proces van woordherkenning komt in dat licht neer op het relateren van bekende semantische informatie (woordbetekenissen) aan op dat moment nog tamelijk onbekende orthografische informatie (geschreven woorden). Uiteraard verloopt de woordherkenning efficiënter en sneller naarmate de woordenschat van de leerling groter is.

Het decodeerproces, het proces van de woordherkenning en de relatie tussen deze twee processen worden beschreven in diverse theoretische leesmodellen, die we hieronder kort bespreken. In principe is er bij het lezen van een woord rechtstreekse woordherkenning mogelijk op basis van de orthografie van dat woord; zeker bij een gevorderde lezer. Maar er is nog een andere optie: de orthografie kan eerst omgezet – gedecodeerd – worden in klankvormen die vervolgens worden gebruikt bij de woordherkenning.

Deze opties worden *direct* en *indirect* genoemd. De mate waarin de ene of de andere optie wordt toegepast is een klassiek thema in het leesonderzoek, met consequenties voor de wijze waarop het lezen aangeleerd zou moeten worden (Harm & Seidenberg, 2004, maar zie ook Bosman & Van Orden, 2003).

Sommige onderzoekers schrijven aan de fonologie geen enkele rol toe als het om woordherkenning gaat, terwijl andere onderzoekers fonologie daarbij juist noodzakelijk achten. Ook zijn er onderzoekers met een 'verzoenende' visie, die inhoudt dat beide opties belangrijk zijn maar onder verschillende omstandigheden; bijvoorbeeld bij de herkenning van zogeheten regelmatige versus onregelmatige woorden: woorden die conform hun uitspraak gespeld worden, zoals 'jas', versus woorden waarvoor dat niet geldt, zoals 'jus'.

Modellen die het leesproces beschrijven, moeten deze en eventueel nog andere opties kunnen verantwoorden. Het linguïstisch analysemodel, dat tot in de jaren zestig van de vorige eeuw de theorievorming over leesprocessen heeft gedomineerd, kan dat niet. Lezen wordt in dat model beschouwd als het omzetten van individuele grafemen in fonemen op basis van een systeem van regels: de zogeheten

grafeem-foneemcorrespondentieregels. Een regelsysteem kan het lezen van onregelmatige woorden (jas – jus) echter niet verklaren. Voor het correct verklanken van deze woorden is aanvullende informatie nodig. Deze informatie heeft een plaats gekregen in het dubbele-routemodel.

Dubbele-routemodel

Het oorspronkelijke dubbele-routemodel (Coltheart, 1978) beschrijft het leesproces in termen van twee onafhankelijke routes: een fonologische route en een lexicale route. Daarnaast is in het model een belangrijke rol weggelegd voor het mentale lexicon; een intern systeem waarin informatie over woorden is opgeslagen: orthografische, fonologische, syntactische en semantische informatie. Dit model veronderstelt dat bij het (indirecte) lezen via de fonologische route een woord eerst van links naar rechts, letter voor letter, verklankt wordt. Vervolgens worden de losse klanken op basis van fonologische informatie in het mentale lexicon samengevoegd, als woord herkend en van betekenis voorzien. Bij het lezen via de lexicale route zou de orthografische informatie in het mentale lexicon direct en als één geheel geactiveerd worden, waardoor de betekenis van het woord beschikbaar komt zonder dat eerst een fonologische omzetting heeft plaatsgevonden.

De indirecte of fonologische aanpak kost meer tijd en is dus minder efficiënt dan de directe of lexicale aanpak. De lexicale route zou echter alleen mogelijk zijn bij woorden die een lezer relatief vaak heeft gelezen. Voor het lezen van woorden waarvoor weinig of geen informatie in het mentale lexicon beschikbaar is (onbekende woorden, laagfrequente woorden, pseudoworden), zou de fonologische route gebruikt moeten worden. Experimenteel onderzoek heeft enkele resultaten opgeleverd die door het dubbele-routemodel goed verklaard kunnen worden: bestaande woorden worden sneller gelezen dan pseudoworden; woorden met een regelmatige orthografie-klankrelatie worden sneller gelezen dan onregelmatige woorden en hoogfrequente woorden worden sneller gelezen dan laagfrequente woorden (Bosman, 2000). Ook recente studies geven evidentie voor het dubbele-routemodel, door aan te tonen dat lezers gaandeweg meer op basis van orthografie tot woordherkenning komen (Grainger & Ziegler, 2011; Ziegler, Bertrand, Lété, & Grainger, 2013).

Daartegenover staan bevindingen die de houdbaarheid van het model in zijn oorspronkelijke vorm serieus hebben aangetast. Allereerst is duidelijk geworden dat er bij het lezen geen sprake kan zijn van een complete onderlinge onafhankelijkheid tussen de fonologische en lexicale route (Coltheart & Rastle, 1994; Paap & Noel, 1991; Rapcsak et al., 2007). Ten tweede blijkt het omzetten van een visuele code naar een klankcode niet alleen te spelen op het niveau van het grafeem maar ook op het niveau van grotere eenheden (Paap & Noel, 1991). Ten derde toont experimenteel onderzoek aan dat lezen zonder gebruik te maken van de klank van woorden niet mogelijk is of niet plaatsvindt (Van Orden et al. 1988), en dat dit geldt voor lezers op alle vaardigheidsniveaus. Het indirecte lezen lijkt dus geen tijdelijk stadium te zijn. De fonologische route wordt ook gebruikt door ervaren lezers, en niet alleen tijdens het lezen van onbekende woorden (Bosman, 2000; McLeod et al., 2001).

Deze bevindingen hebben geleid tot bijstellingen van het dubbele-routemodel en tot het ontstaan van alternatieve, connectionistische, modellen (zie de paragraaf hieronder). Een belangrijke bijstelling van het dubbele-routemodel is het dubbele-route-cascademodel of DRC-model (Coltheart et al., 1993). Het DRC-model is uitgevoerd als een computerprogramma waarmee taken die in experimenteel leesonderzoek gebruikelijk zijn, gesimuleerd kunnen worden. De term 'cascade' (waterval) verwijst naar de manier waarop de activatie onbelemmerd kan 'doorstromen' van het ene niveau naar het andere.

Sinds de introductie is het DRC-model in veel empirisch onderzoek op de proef gesteld en heeft het deze tests over het algemeen doorstaan. Ziegler et al. (2000) tonen aan dat het DRC-model ook van toepassing is op het Duits. En in een recentere beschrijving van het DRC-model claimen Coltheart et al. (2001) dat het model succesvol is in het simuleren van een veelheid aan verschijnselen op het gebied van (onder meer) het hardop lezen. Zij eindigen hun artikel met een lange lijst van voorbeelden die deze claim moeten staven. Steun voor het model komt ook van Rapcsak et al. (2007) die resultaten presenteren waaruit blijkt dat het DRC-model de leesprestaties kan voorspellen van patiënten met verworven alexie en agrafie.

Hoewel DRC aldus voldoende ondersteund lijkt te worden vanuit de empirie, is het model niet onweersproken. Een model is slechts een middel; het doel is een theorie die leesgedrag en de neurale fundamenteen daarvan verklaart (Seidenberg, 2007). Deze pretentie heeft het DRC-model niet, in ieder geval niet expliciet, maar hebben connectionistische modellen wél.

Connectionistische modellen

Sinds de late jaren tachtig van de vorige eeuw worden in het leesonderzoek connectionistische modellen toegepast. Het gaat daarbij om computermodellen die gebaseerd zijn op theorieën over hoe lezen in z'n werk gaat, hoe kinderen leren lezen, hoe het lezen van goede lezers verloopt en hoe leesproblemen, zoals dyslexie, verklaard kunnen worden. Connectionistische modellen steunen op aannames over de manier waarop kennis in ons brein gerepresenteerd is; zij toetsen hypothesen over menselijk leesgedrag ten dienste van een algemene theorie: een theorie die omvangrijker is dan louter een leestheorie.

Ontwikkelaars van connectionistische modellen hebben het idee van het mentale lexicon verlaten. In plaats daarvan wordt een uitgebreid netwerk van orthografische, fonologische en semantische eenheden verondersteld, waartussen bij het lezen van een woord verbindingen (connecties; vandaar de naamgeving van deze modellen) worden gelegd. Kern van het connectionisme is dat de verklanking van een woord tot stand komt in één geïntegreerd proces, waarin op basis van kennis van statistische regelmatigheden op verschillende niveaus activatie wordt opgebouwd. Letterclusters, syllaben, morfemen en woorden die vaker zijn omgezet in klanken, worden volgens deze theorie sneller en preciezer herkend (Schijff, 2009). In de literatuur zijn twee belangrijke connectionistische modellen beschreven: het parallelle gedistribueerde verwerkingsmodel en het fonologisch-coherentiemodel.

Parallelle gedistribueerde verwerkingsmodel

Het PDP-model, waarin PDP staat voor *parallel distributed processing*, is ontwikkeld door Seidenberg en McClelland (1989). In het dubbele-routemodel worden de lexicale en fonologische route als onafhankelijk opgevat. Het alternatief van het PDP-model is dat beide routes parallel werken. De term *distributed* verwijst naar de manier waarop informatie gerepresenteerd wordt geacht.

Het oorspronkelijke PDP-model bestaat uit verzamelingen van orthografische en fonologische eenheden die met elkaar verbonden zijn. De verbindingen zijn voorzien van gewichten die bepalen hoeveel informatie doorgegeven wordt. Deze gewichten weerspiegelen een leereffect, de som van alle training die iemand in z'n 'leescarrière' heeft ondergaan.

Harm & Seidenberg (1999) onderzoeken de rol van fonologische informatie bij het vroege lezen en in dyslexie. Hun PDP-model is een uitbreiding van de principes van Seidenberg & McClelland (1989); het simuleert de ontwikkeling van de leesvaardigheid en de stoornissen die tijdens die ontwikkeling kunnen optreden. In een latere versie van het PDP-model is ook een semantische component ingebouwd (Harm & Seidenberg, 2004), met name om een bijdrage te leveren aan het langlopende debat over de rol van fonologische informatie bij het stil lezen. Dit model beschrijft hoe betekenis geactiveerd wordt in een systeem waarin zowel een visuele (van orthografie naar betekenis) als een fonologische route (van orthografie via fonologie naar betekenis) beschikbaar is. Een dergelijk systeem blijkt efficiënter te werken wanneer beide routes gebruikt worden, dan wanneer een van de twee geïsoleerd gebruikt wordt. Aanvankelijk, bij de beginnende lezer, wordt de semantische activatie grotendeels aangedreven via de orthografie-fonologie-betekenisroute. Na verloop van tijd begint de orthografie-betekenisroute meer invloed te krijgen. Uiteindelijk is de situatie zo dat er bij het lezen van de meeste woorden input van beide routes komt.

Fonologisch-coherentiemodel

Een ander relevant connectionistisch model is het fonologisch-coherentiemodel (Van Orden et al., 1990), dat gebaseerd is op de principes van de dynamische systeemtheorie. Het model heeft de vorm van een netwerk waarin drie zogeheten knooppamilies worden verondersteld: letterknopen, foneemknopen en semantische kenmerknopen (Bosman, 2000). De activatie tussen de knooppamilies verloopt in beide

richtingen, heen en terug (Stone et al., 1997); een belangrijk onderscheid ten opzichte van het PDP-model, dat een zogeheten *feedforward*-model is waarin de activatie zich slechts in één richting verspreidt (Bosman & Van Orden, 2003). Een essentieel kenmerk van het systeem is dat de verbindingen tussen verschillende families niet even sterk zijn. De sterkste verbindingen zijn die tussen de letterknopen en de foneemknopen. Dat ligt voor de hand: er bestaan zeer consistente relaties tussen letters en fonemen. En dit zou ook verklaren waarom de fonologie een fundamentele rol speelt bij het lezen. Daarentegen zijn de verbindingen tussen spraakklanken en betekenselementen en tussen letters en betekenselementen veel zwakker. Het verband tussen de foneemknopen en semantische-kenmerknopen is overigens sterker dan tussen de letterknopen en semantische-kenmerknopen; dit omdat we leren spreken voordat we leren lezen (Bosman, 2000; Bosman & Van Orden 2003).

Als dit connectionistisch netwerk een geschreven woord krijgt voorgelegd, worden eerst de letterknopen geactiveerd. Zij sturen hun activatie door naar de andere knopen, en ontvangen daarna weer activatie terug. Er ontstaat een dynamisch proces van voor- en terugwaartse activaties waarbij alle drie de knoopfamilies zijn betrokken. In dit proces komt woordherkenning tot stand door onderlinge beïnvloeding van knopen van verschillende families, waaruit relatief stabiele terugkoppelingslusen ontstaan. Relaties tussen knopen uit verschillende knoopfamilies worden tijdelijk als functionele eenheden vastgelegd in het geheugen. Tijdens dat proces wordt de sterkte van de relaties voortdurend aangepast op basis van de leeservaring. Woordherkenningsprocessen gaan hierdoor steeds efficiënter en sneller verlopen.

Amalgammodellen

Nog steeds worden in verschillende studies de verschillende leesmodellen getoetst (zie bijvoorbeeld Coltheart, 2006; Harm & Seidenberg, 2004; Seidenberg, 2007; Peterson, Pennington, & Olson, 2013; Pritchard, Coltheart, Palethorpe, & Castles, 2012), waarbij aspecten van het dubbele-routemodel steeds vaker gecombineerd worden met connectionistische opvattingen. Een goed voorbeeld van zo'n combinatie is het CDP-model (*Connectionist Dual Process Model of Reading Aloud*), in 1998 voor de Engelse taal ontwikkeld door Zorzi, Houghton & Butterworth. Dit computermodel berekent en verklaart de 'werkverdeling' tussen lexicale en niet-lexicale processen in neurale netwerken bij zowel ervaren als beginnende lezers. Ook biedt het model aanknopingspunten voor het verklaren van leesstoornissen zoals dyslexie. Van het CDP-model zijn in de afgelopen jaren diverse nieuwe versies ontwikkeld: CDP+ (Perry et al., 2007), CDP++ (Perry et al., 2010) en CDP++.parser (Perry et al., 2013). De meest recente versie is voorzien van software die woorden en pseudoworden (letterreeksen) kan ontleden in grafemen, maar ook in beginletters/voorvoegsels, eindletters/achtervoegsels en klinkers en medeklinkers.

Tot besluit

Deze verantwoording is niet de plaats om de geldigheid van de onderscheiden leesmodellen tot in detail te bediscussiëren. Voor ons doel, het theoretisch inkaderen van de AVI-Toets, is het belangrijk dat de modellen een aantal eigenschappen gemeen hebben. Om te beginnen voorspellen alle modellen dat bekendheid met en van het te lezen materiaal de nauwkeurigheid en de vlotheid van het lezen ten goede komt. Hoe vaker een leerling bepaalde letters, letterclusters, woorden en woordgroepen heeft gedecodeerd, des te preciezer en sneller verloopt het leesproces.

Daarnaast gaat het in alle modellen om de sterkte van representaties en verbindingen, om doorvoer van activatie, om toegang tot systemen; eigenschappen en processen die de verschillen tussen de zwakkere en de betere lezers verklaren en die in hun samenhang resulteren in de leesvaardigheid van, in het geval van onze doelgroep, individuele leerlingen op de basisschool. Hun niveau van technisch lezen kenmerkt zich door een bepaalde vlotheid en een bepaalde mate van nauwkeurigheid. Hoe deze aspecten van technische leesvaardigheid in de AVI-toets geoperationaliseerd zijn, wordt beschreven in hoofdstuk 3.

3 Beschrijving van de toets

3.1 Opbouw en structuur van de toets

Het toetspakket voor de AVI-toets bestaat uit een handleiding en papieren toetsmaterialen (leeskaarten, scoreformulieren, een kaart met de afname-instructie en een kaart met de score-instructie). De uitgave voor zorgaanbieders bevat tevens registratieformulieren voor handmatige verwerking van de scores. Voor het onderwijs zijn deze registratieformulieren te vinden op Cito Portal. De AVI-toets is een individuele toets. Het is de taak van de leerling om de tekst op een leeskaart hardop, nauwkeurig en vlot te lezen. In de groepen 3 tot en met 7 is de toets op twee afnamemomenten (M en E) genormeerd. In groep 8 is de toets alleen halverwege het schooljaar (M) genormeerd.

Opbouw en structuur

Er zijn in beide uitgaven (die voor het onderwijs en die voor zorgaanbieders) leeskaarten opgenomen voor 11 AVI-niveaus. Op elke kaart staat een tekst waarvan de leestechische moeilijkheid aansluit bij het betreffende AVI-niveau. Zie voor de kenmerken van de teksten op de toetskaarten paragraaf 3.2.2 van deze verantwoording. Elke tekst vormt een afgerond geheel.

De uitgave voor het onderwijs bevat twee parallelversies van elke leeskaart, een A- en een B-versie. De uitgave voor zorgaanbieders bevat slechts één versie van elke kaart, de C-versie. De parallelversies van de leeskaarten bevatten andere teksten, maar zijn qua opzet volledig vergelijkbaar.

Afname

De AVI-toets wordt individueel afgenomen door een toetsleider. Dat kan de leerkracht zijn, maar ook een intern begeleider of zorgaanbieder. De toetsafname begint met een instructie. Na de instructie draait de toetsleider de kaart om en leest zelf de titel voor – deze kan namelijk woorden bevatten die, hoewel onmisbaar, voor de leerlingen nog te moeilijk zijn. Op het moment dat de leerling begint te lezen, start de toetsleider de tijdwaarneming. Tijdens het lezen noteert de toetsleider op een scoreformulier welke woorden fout gelezen worden en eventueel welk type leesfout de leerling daarbij maakt. Wanneer de leerling de gehele tekst gelezen heeft, stopt de toetsleider de tijdwaarneming en noteert de gebruikte leestijd en het aantal leesfouten op het scoreformulier.

De AVI-toets kent twee afnamemomenten voor groep 3 t/m 7: medio (half januari tot half februari) en einde leerjaar (half mei tot eind juni). In jaargroep 8 is er slechts één normeringsmoment: medio leerjaar. Om de resultaten van de leerlingen te kunnen vergelijken met die van andere leerlingen uit het land (zie de laatste alinea van deze paragraaf) is het belangrijk dat de toetsleider zich houdt aan de vermelde toetsperiodes.

Voor de meeste leerlingen bestaat een AVI-afname uit het lezen van meer dan één toetskaart. Afhankelijk van het resultaat op de eerste kaart van de afnamesessie krijgt de leerling een tweede kaart van een lager of hoger AVI-niveau voorgelegd; afhankelijk van het resultaat op deze kaart volgt eventueel een derde kaart, enzovoort. Dit wordt doortoetsen genoemd. Of en met welke kaart er doorgetoetst wordt, hangt af van de behaalde score op de beginkaart (beheersings-, instructie- of frustratieniveau) en de gekozen afnameoptie. Er zijn drie afnameopties mogelijk. Deze worden hierna beschreven.

Optie 1: De leerlingen worden doorgetoetst tot en met het laagste instructieniveau. Hiermee is duidelijk wat het beheersingsniveau van alle leerlingen van de groep is en wat het laagste instructieniveau is.

Optie 2: De leerlingen worden doorgetoetst tot en met het frustratieniveau. Hiermee is duidelijk wat het beheersingsniveau van alle leerlingen van de groep is en wat het hoogste instructieniveau is.

Optie 3: De leerlingen worden doorgetoetst tot en met het afnamemoment van de jaargroep. Alleen leerlingen die het afnamemoment van de jaargroep niet beheersen, worden doorgetoetst tot en met het frustratieniveau. Hiermee is duidelijk welke leerlingen de leeskaart beheersen die correspondeert met het afnamemoment. Van de leerlingen die minder vaardig zijn is duidelijk wat het hoogste instructieniveau is.

Scoring

Tijdens het lezen noteert de toetsleider op het bijbehorende scoreformulier welke woorden fout gelezen worden. De scorevoorschriften zijn beschreven op een kaart met score-instructies. Woorden waarvan de uitspraak niet overeenkomt met de spelling worden fout gerekend. Klemtoonverschillen vallen daar niet onder. Spellend gelezen woorden worden fout gerekend, tenzij ze daarna alsnog correct verklankt worden.

De toetsleider noteert op het scoreformulier de totale leestijd en het aantal gemaakte fouten. Door deze waarden te vergelijken met de grenswaarden die bovenaan het scoreformulier staan, kan eenvoudig bepaald worden of de leerling de kaart heeft gelezen op beheersings-, instructie- of frustratieniveau. Desgewenst kan de toetsleider ook noteren welke soorten fouten de leerling heeft gemaakt. In het Computerprogramma LOVS kunnen zowel de leestijd als het aantal fouten worden ingevoerd.

Verwerking resultaten en interpretatie

Na de toetsafname en de scoring kunnen de toetsresultaten verwerkt worden op speciaal ontwikkelde registratieformulieren. De rapportages op leerlingniveau zijn zowel beschikbaar bij de uitgave voor het onderwijs als de uitgave voor zorgaanbieders. De rapportages op groeps- en schoolniveau zijn uitsluitend beschikbaar bij de uitgave voor het onderwijs. In de handleiding in de toetsmap is een beschrijving opgenomen van alle rapportages en analyses, met een toelichting op het gebruik en de inhoudelijke interpretatie ervan. In de handleiding bij het Computerprogramma LOVS staat informatie over de wijze waarop de rapportages zijn op te vragen en welke keuzemogelijkheden de school hierbij heeft.

Bij de beoordeling van de behaalde resultaten is het belangrijk om te weten hoe de AVI-niveaus zich verhouden tot landelijke gegevens. Zoals in paragraaf 2.3 al is aangegeven, betekent een beheersings-score op leeskaart M3 dat het technisch leesniveau van de leerling aansluit bij dat van de gemiddelde leerling op het M3-afnamemoment. Omdat dit een vrij algemeen criterium is, is in de handleiding de volgende informatie toegevoegd:

- Tweederde tot driekwart van de leerlingen heeft een AVI-niveau dat correspondeert met het afnamemoment of met een later afnamemoment. Dit betekent bijvoorbeeld dat gemiddeld twee derde tot drie kwart van de leerlingen in het midden van groep 5 een beheersingsscore haalt op AVI-M5 of een beheersingsscore op een leeskaart die hoger is;
- Voor één tiende tot een kwart van de leerlingen is de leeskaart die correspondeert met het afnamemoment te moeilijk. Dit betekent bijvoorbeeld dat gemiddeld een tiende tot een kwart van de leerlingen in het midden van groep 5 een frustratiescore haalt op AVI-M5 of een frustratiescore op een leeskaart die lager is.

3.2 Inhoudsverantwoording

In het ontwikkelproces van de AVI-toets zijn een aantal fasen te onderscheiden:

- Uitwerking van de domeinbeschrijving
- Schrijven van teksten voor de leeskaarten
- Screening en try-out
- Proeftoetsing
- Kalibratie en normering
- Uitgave

De eerste vier fasen, tot en met de proeftoetsing, worden hierna toegelicht. In hoofdstuk 4 komen de kalibratie en normering aan de orde.

3.2.1 Uitwerking domeinbeschrijving

In paragraaf 2.4 is technisch lezen omschreven als het correct en vlot ontsleutelen van geschreven woorden, al dan niet in context, en het herkennen van deze woorden als dragers van betekenis. Omdat technische leesvaardigheid wordt gedefinieerd in termen van nauwkeurigheid én snelheid, wordt de AVI-toets onder een zekere tijdsdruk afgenomen.

Met behulp van de AVI-toets kan worden nagegaan hoe goed leerlingen *teksten van uiteenlopende moeilijkheidsgraad hardop* kunnen verklanken. Waarom hiervoor gekozen is, wordt hieronder uitgelegd.

Waarom teksten en geen geïsoleerde woorden?

Een belangrijk verschil tussen het lezen van losse woorden en het lezen van zinnen of tekst is dat de contextuele informatie die de zin of de tekst bevat een rol kan spelen bij het decoderen: door gebruik te maken van informatie uit de context kunnen lezers hun eventuele decodeerproblemen tot op zekere hoogte compenseren doordat de context als het ware aanvullende aanwijzingen verstrekt (Curtis & Glaser, 1983; Goldsmith-Phillips, 1989; Kim & Goetz, 1994; Leu Jr. et al., 1986; Nicholson, 1991; Nicholson et al., 1988; Stanovich, 1980; Stanovich & Stanovich, 1995; Stanovich et al., 1981).

Goede lezers decoderen bijna altijd zonder daarbij gebruik te maken van informatie uit de context of van welke kennis – anders dan die over letters en klanken – dan ook. Zwakke lezers daarentegen steunen dikwijls op de context. Als zij een woord tegenkomen in een context die het woord in hoge mate voorspelbaar maakt, kunnen zij dat woord makkelijker ontsleutelen. Tekortkomingen in het afleiden van informatie uit het woord zelf kunnen op deze manier gecompenseerd worden. Dit mechanisme treedt ook in werking bij de alledaagse leessituatie, waarin leerlingen voornamelijk samenhangende teksten lezen en geen losse woorden. Om deze reden is ervoor gekozen om teksten op te nemen in de AVI-toets. Het authentieke karakter hiervan maakt het mogelijk om een goede indruk te geven van de technische leesvaardigheid die de leerling in de praktijk beheerst.

Waarom teksten van uiteenlopende moeilijkheidsgraad?

Kwantitatieve en kwalitatieve linguïstische kenmerken zijn van invloed op de leesscores van leerlingen. Dit is al geïllustreerd in paragraaf 2.4 van deze verantwoording waarin onderzoek beschreven is dat uitgevoerd werd om de daar gepresenteerde theoretische leesmodellen empirisch te ondersteunen. Het betekent bijvoorbeeld dat een tekst die bestaat uit korte frequente woorden minder technische leesvaardigheid vraagt dan een tekst die bestaat uit lange minder frequente woorden. Bij het schrijven van de teksten voor de AVI-kaarten zijn de tekstkenmerken gehanteerd die voor elk AVI-niveau zijn vastgelegd. Deze tekstkenmerken stellen voor elk hoger AVI-niveau aanvullende eisen. In paragraaf 3.2.2 wordt voor elk AVI-niveau een omschrijving van de tekstkenmerken gegeven.

Waarom hardop lezen?

Lezen kan zowel stil als hardop, maar het meeste lezen gebeurt stil. Uit onderzoeksliteratuur blijkt dat stil lezen en hardop lezen geen identieke processen zijn, maar ook dat leessnelheidsscores die behaald worden op een hardop-leestoets wel degelijk een indicatie kunnen zijn voor het leestempo tijdens het stil lezen (Allington, 1984; Armbruster & Wilkinson, 1991; Burge, 1983; Leu Jr., 1982; Mosenthal, 1976-1977; Mosenthal, 1978; Schumm & Baldwin, 1989; Søvik et al., 2000). Er is voor gekozen om leerlingen bij de AVI-toets hardop te laten lezen, omdat snelheid en accuratesse dan direct in kaart kunnen worden gebracht en niet hoeven te worden afgeleid.

3.2.2 Totstandkoming leeskaarten

De teksten zijn speciaal voor de AVI-toets geschreven door professionele auteurs van jeugdliteratuur. Aan hen is gevraagd om teksten te schrijven van een ongeveer 100 woorden voor het laagste AVI-niveau tot ongeveer 220 woorden voor het hoogste AVI-niveau. De auteurs werd eveneens verzocht om qua tekstinhoud aan te sluiten bij de belevingswereld van de verschillende doelgroepen (medio jaargroep 3,

einde jaargroep 3, medio jaargroep 4, enzovoort). Ten slotte is aan de auteurs gevraagd om rekening te houden met de tekstkenmerken die gelden voor de verschillende AVI-niveaus (zie tabel 3.1). Deze tekstkenmerken hebben betrekking op woordtypen, zinskenmerken en de CILT-score (Staphorsius & Verhelst, 1997). CILT staat voor Cito Index voor de LeesTechniek. Dit is een index voor de technische leesbaarheid van een tekst. De CILT wordt vastgesteld met behulp van een formule, waarbij het percentage hoogfrequente woorden in de tekst en het gemiddeld aantal letters per woord een rol spelen.

Tabel 3.1 Overzicht tekstkenmerken per AVI-niveau

AVI-niveau	Woordtypen	Zinskenmerken	CILT-waarde
AVI-M3	éénlettergrepige woorden (mk, km, mkm) met alle tekens en klanken derde persoons -t zoals zegt	korte zinnen één zin per regel nevenschikkende samengestelde zinnen betekenisvol verdeeld over twee regels mogen voorkomen geen hoofdletters	CILT ≤ 56,9
AVI-E3	de woordtypen van de vorige leeskaart én: woorden beginnend met –sch woorden eindigend op –ng een- en tweelettergrepige woorden met twee of drie medeklinkers na elkaar zoals staart, botst, sprong, bankje, knappe, winkel enkele eenvoudige twee-lettergrepige woorden met de volgende leesmoelijkheden: voorvoegsel, stomme -e, verkleinwoorden (-je), samengestelde woorden enkele woorden eindigend op -nk, -b, -d(t), -ch(t), -ooi, -aai, -oei, -eeuw, -ieuw, -uw	korte zinnen één zin per regel nevenschikkende samengestelde zinnen betekenisvol verdeeld over twee regels mogen voorkomen hoofdletters mogen voorkomen	56,9 < CILT ≤ 58,9
AVI-M4	de woordtypen van de vorige leeskaarten én: moeilijkere twee- en drielettergrepige woorden met voorvoegsel be- ge-, ver- samengestelde woorden met drie lettergrepen enkele woorden met de volgende leesmoelijkheden: th-, –aaien, /i/ uitgesproken als /ie/, woorden eindigend op -end, -ig, -lijk, -ing, -eren, -etje, -ennetje enkele drielettergrepige woorden als bibberen, stilletjes, andere	samengestelde zinnen (nevenschikkend en onderschikkend) mogen voorkomen zinnen mogen, betekenisvol afgebroken, doorlopen op de volgende regel iedere nieuwe zin begint op een nieuwe regel	58,9 < CILT ≤ 61,9
AVI-E4	de woordtypen van de vorige leeskaarten én: enkele drie- en vierletter-grepige woorden met een open lettergreep zoals bovenbenen, stevige woorden met -be-, -ge- of -ver- in het midden enkele woorden met de volgende leesmoelijkheden: woorden eindigend op -tie, hoogfrequente leenwoorden zoals race	een nieuwe zin hoeft niet meer te beginnen op een nieuwe regel	61,9 < CILT ≤ 63,9

AVI-M5	de woordtypen van de vorige leeskaarten én: enkele drie- en meerletter-grepige woorden eindigend op open lettergreep enkele woorden met de volgende leesmoelijkheden: /y/ uitgesproken als /ie/, /y/ uitgesproken als /j/, /c/ uitgesproken als /k/, /c/ uitgesproken als /s/ woorden met een -x- woorden met 's	er gelden geen beperkingen ten aanzien van de zinskenmerken	63,9 < CILT ≤ 65,9
AV-E5	de woordtypen van de vorige leeskaarten én: enkele woorden met de volgende leesmoelijkheden: beginnend met /ch/ uitgesproken als /sj/, eindigend op –ge, uitgesproken als /zje/, eindigend op –isch, woorden met klinkerreeks, leenwoorden met eau, é of è	er gelden geen beperkingen ten aanzien van de zinskenmerken	65,9 < CILT ≤ 67,9
AVI-M6	de woordtypen van de vorige leeskaarten én: enkele woorden met de volgende leesmoelijkheden: woorden eindigend op -ele, -eal, -ueel, -iaal of -ieel, woorden met een trema, minder frequente leenwoorden	er gelden geen beperkingen ten aanzien van de zinskenmerken	67,9 < CILT ≤ 69,9
AVI-E6	Alle woordtypen kunnen voorkomen. Voor het AVI-niveau is uitsluitend de CILT-score bepalend.	er gelden geen beperkingen ten aanzien van de zinskenmerken	69,9 < CILT ≤ 71,9
AVI-M7	Alle woordtypen kunnen voorkomen. Voor het AVI-niveau is uitsluitend de CILT-score bepalend.	er gelden geen beperkingen ten aanzien van de zinskenmerken	71,9 < CILT ≤ 73,9
AVI-E7	Alle woordtypen kunnen voorkomen. Voor het AVI-niveau is uitsluitend de CILT-score bepalend.	er gelden geen beperkingen ten aanzien van de zinskenmerken	73,9 < CILT ≤ 74,9
AVI-Plus	Alle woordtypen kunnen voorkomen. Voor het AVI-niveau is uitsluitend de CILT-score bepalend.	er gelden geen beperkingen ten aanzien van de zinskenmerken	CILT >74,9

Nadat de auteurs de leeskaarten hadden voltooid, werd nagegaan of deze aan alle tekstkenmerken voldeden. Zo nodig werden de leeskaarten aangepast. Vervolgens werd opnieuw gecontroleerd of aan alle tekstkenmerken was voldaan en volgde een tweede aanpassingsronde. Dit proces werd net zolang herhaald tot alle kaarten in overeenstemming waren met de tekstkenmerken.

3.2.3 Gerealiseerde leeskaarten

In deze paragraaf wordt de samenstelling van de gerealiseerde leeskaarten beschreven. In tabel 3.2 staan de kwantitatief-linguïstische kenmerken van de teksten op de AVI-leeskaarten vermeld. De tabel laat zien dat de CILT-waarde van de gerealiseerde leeskaarten in overeenstemming zijn met de beoogde CILT-waarde zoals vermeld in tabel 3.1. Uit de tabel valt ook af te leiden dat de verdeling van één-, twee-, drie-, vier-, of meerlettergrepige woorden over de kaarten in overeenstemming is met de beoogde verdeling. Zo is bijvoorbeeld te zien dat de kaarten op niveau AVI-M3 uitsluitend uit éénlettergrepige woorden bestaan en dat vanaf niveau AVI-E3 tweelettergrepige woorden worden geïntroduceerd.

Tabel 3.2 Kwantitatief-linguïstische kenmerken van de teksten op de AVI-leeskaarten

Leeskaart	Titel	Woorden	Gemiddelde woordlengte	Percentage frequente woorden	CILT	Range CILT	Aantal 1-lettergrepen	Aantal 2-lettergrepen	Aantal 3-lettergrepen	Aantal 4- of meerlettergrepen
AVI-M3A	De mus	99	3,05	88,89	48,43	CILT ≤ 56,9	99	0	0	0
AVI-M3B	De pup	99	3,17	87,88	50,21	CILT ≤ 56,9	99	0	0	0
AVI-M3C	De reus	98	3,28	90,62	50,79	CILT ≤ 56,9	98	0	0	0
AVI-E3A	Knappe hond	105	3,75	83,81	58,51	56,9 < CILT ≤ 58,9	99	6	0	0
AVI-E3B	Sneeuw	111	3,71	83,64	58,02	56,9 < CILT ≤ 58,9	97	14	0	0
AVI-E3C	Floor zwemt af	115	3,74	88,50	57,09	56,9 < CILT ≤ 58,9	103	12	0	0
AVI-M4A	De nieuwe meester	118	3,96	82,20	61,49	58,9 < CILT ≤ 61,9	97	18	3	0
AVI-M4B	Ridder Rosa	124	3,85	80,65	60,66	58,9 < CILT ≤ 61,9	102	17	5	0
AVI-M4C	Het hertje	122	4,02	86,07	61,23	58,9 < CILT ≤ 61,9	100	18	4	0
AVI-E4A	Blubber	164	4,09	82,32	63,03	61,9 < CILT ≤ 63,9	121	35	7	1
AVI-E4B	Een stoere sport	164	4,07	78,66	63,83	61,9 < CILT ≤ 63,9	125	30	7	2
AVI-E4C	Spinnen	163	3,99	75,31	63,79	61,9 < CILT ≤ 63,9	128	28	4	3
AVI-M5A	Trefbal	172	4,23	80,23	65,43	63,9 < CILT ≤ 65,9	121	35	15	1
AVI-M5B	De popgroep	169	4,30	81,66	65,81	63,9 < CILT ≤ 65,9	110	45	10	4
AVI-M5C	De redding	176	4,20	78,98	65,44	63,9 < CILT ≤ 65,9	126	36	12	2
AVI-E5A	Wij gaan winnen!	164	4,34	76,22	67,90	65,9 < CILT ≤ 67,9	109	38	13	4
AVI-E5B	Onderwaterwereld	166	4,39	80,12	67,42	65,9 < CILT ≤ 67,9	109	41	11	5
AVI-E5C	De baas	163	4,25	75,46	67,00	65,9 < CILT ≤ 67,9	118	29	14	2
AVI-M6A	De waakhond	172	4,48	76,16	69,65	67,9 < CILT ≤ 69,9	112	39	15	6
AVI-M6B	Het monster	163	4,40	77,30	68,38	67,9 < CILT ≤ 69,9	110	41	11	1
AVI-M6C	De dader	165	4,47	80,61	68,21	67,9 < CILT ≤ 69,9	106	42	13	4
AVI-E6A	De dief	171	4,64	77,78	71,18	69,9 < CILT ≤ 71,9	110	35	20	6
AVI-E6B	De telefoon	171	4,71	78,36	71,81	69,9 < CILT ≤ 71,9	113	33	22	3
AVI-E6C	Het optreden	166	4,63	74,70	71,84	69,9 < CILT ≤ 71,9	110	36	14	6
AVI-M7A	Een geweldig resultaat	210	4,76	74,76	73,43	71,9 < CILT ≤ 73,9	131	42	26	11
AVI-M7B	Het Portugees oorlogsschip	215	4,80	76,74	73,46	71,9 < CILT ≤ 73,9	141	36	28	10
AVI-M7C	Een ijzige wedstrijd	211	4,77	74,88	73,59	71,9 < CILT ≤ 73,9	129	48	28	6
AVI-E7A	De computer van de toekomst	222	4,84	73,87	74,73	73,9 < CILT ≤ 74,9	136	42	35	9
AVI-E7B	De speurtocht	205	4,83	75,61	74,14	73,9 < CILT ≤ 74,9	124	55	16	10
AVI-E7C	Is er buitenaards leven?	224	4,86	75,89	74,35	73,9 < CILT ≤ 74,9	131	50	33	10
AVI-Plus A	Termieten	222	4,98	72,07	76,90	CILT >74,9	137	49	21	15
AVI-Plus B	Parfum	223	4,99	73,99	76,53	CILT >74,9	126	57	24	16
AVI-Plus C	Stof tot nadenken	220	4,96	71,36	76,87	CILT >74,9	129	49	25	17

3.2.4 Screening en try-out

Nadat de leeskaarten waren samengesteld zijn ze voorgelegd aan een aantal Nederlandse en Vlaamse vakexperts. Op basis van hun commentaar zijn de kaarten bijgesteld. De bijgestelde leeskaarten zijn besproken met een aantal leerkrachten. Ook is een try-out georganiseerd waarbij klein aantal leerlingen gevraagd is om de kaarten hardop te lezen. Op basis van de deze ervaringen zijn de kaarten herzien. Bij de herziening is ervoor gezorgd dat teksten aan bleven sluiten bij de tekstkenmerken.

3.2.5 Proeftoetsing

De herziene leeskaarten zijn in een proeftoets voorgelegd aan in totaal ongeveer duizend leerlingen in groep 3 t/m 7 van het primair onderwijs. Periode van afname was het E-moment (mei/juni 2015).

De kaarten zijn niet in groep 8 afgenomen, omdat het laatste reguliere afnamemoment in groep 8 niet het E-moment maar het M-moment is.

Om in een later stadium de onderliggende prestatie-eisen voor de instructie- en beheersingsniveaus van de vorige AVI-toets te kunnen overbrengen op de nieuwe leeskaarten, is in de proeftoetsing nagegaan of de nieuwe en de oude leeskaarten van een vergelijkbare moeilijkheidsgraad waren. Daarom zijn zowel nieuwe als oud AVI-kaarten aan leerlingen voorgelegd. De verschillende versies van de nieuwe leeskaarten zijn volgens een 'balanced incomplete block' principe aan subgroepen toegewezen. Dit betekent dat alle versies even vaak, in alle mogelijke combinaties, zijn afgenomen. Tabel 3.3 geeft gedetailleerd weer hoe het proeftoetsdesign eruitzag. De tabel laat zien dat er per jaargroep 4 of 5 subgroepen zijn gemaakt. Voorafgaand aan de proeftoets was van alle leerlingen het AVI-beheersingsniveau opgevraagd. Aan de hand hiervan zijn de leerlingen ingedeeld in de subgroepen, zodat het niveau van de voorgelegde kaarten min of meer op hen kon worden afgestemd. Uit de tabel komt eveneens naar voren dat elke leerling telkens 3 nieuwe leeskaarten en 1 oude leeskaart moest oplezen. Er zijn in totaal 45 leerlingen geworven per subgroep, zodat van elke leeskaart minstens 90 waarnemingen beschikbaar waren. Deze doelstelling is ruimschoots gerealiseerd.

Tabel 3.3 Design proeftoets conceptversie AVI-leeskaarten

afname-moment	subgroep	bevatte de nieuwe leeskaarten			bevatte de oude leeskaart
E3	1	M3A	M3B	M3C	M3
	2	M3B	M3C	E3A	M3
	3	E3A	E3B	E3C	E3
	4	M3A	E3B	M4C	E3
	5	E3C	M4A	M4B	M4
E4	6	M4B	M4C	E4A	M4
	7	E4A	E4B	E4C	E4
	8	M4A	E4B	M5C	E4
	9	E4C	M5A	M5B	M5
E5	10	M5A	M5B	E5A	M5
	11	E5A	E5B	E5C	E5
	12	M5A	E5B	M6C	E5
	13	E5C	M6A	M6B	M6
E6	14	M6B	M6C	E6A	M6
	15	E6A	E6B	E6C	E6
	16	M6A	E6B	M7C	E6
	17	E6C	M7A	M7B	M7
E7	18	M7B	M7C	E7A	M7
	19	E7A	E7B	E7C	E7
	20	M7A	E7B	AVI-PlusC	E7
	21	E7C	AVI-PlusA	AVI-PlusB	Plus
	22	AVI-PlusA	AVI-PlusB	AVI-PlusC	Plus

Na afloop van de proeftoets zijn de scores op de verschillende kaarten geanalyseerd. Daarnaast zijn de logboeken van de toetsleiders bekeken. Op basis van deze informatie is de tekst van een aantal leeskaarten aangepast. Zo is het woord 'plateaus' verwijderd van kaart E5B om deze makkelijker te maken. Aan kaart E7B zijn de woorden 'optimistisch' en 'slechts' toegevoegd om de moeilijkheidsgraad wat te verhogen. Net als bij de try-out is ervoor gezorgd dat de herziene teksten voldeden aan de tekstkenmerken.

4 Kalibratie en normering

4.1 Dataverzameling

In schooljaar 2015/2016 is een landelijk onderzoek in het reguliere basisonderwijs uitgevoerd ten behoeve van de normering van de AVI-toets. Er is gebruikgemaakt van een cross-sectioneel design waarin leerlingen in de groepen 3 tot en met 7 op twee momenten in het schooljaar getoetst werden, namelijk in januari/februari (mediomoment) en in mei/juni (eindemoment). De leerlingen in groep 8 zijn alleen in januari/februari getoetst.

Er is naar gestreefd om bij de afname op het medio- en eindemoment in een jaargroep dezelfde leerlingen te toetsen, zodat de voortgang van deze leerlingen in beeld kon worden gebracht. Bij de toetsafname kregen leerlingen twee of drie leeskaarten uit de nieuwe AVI-toets voorgelegd. Daarnaast las een deel van de leerlingen een of meer leeskaarten uit de vorige uitgave van de AVI-toets (Cito, 2009). Hierdoor was het mogelijk om de prestatie-eisen van de vorige AVI-toets over te brengen op de huidige. Ten slotte zijn met het oog op de validering van de AVI-toets enkele andere toetsen afgenomen. Het ging om de Eén-Minuu-Test (Brus en Voeten, 1979) en de DMT (Cito, 2017). Om overbelasting bij de leerlingen te voorkomen, werden de toetsafnames, zowel voor het medio- als voor het eindemoment, op verschillende dagen gepland. De tabellen 4.1 tot en met 4.4 geven gedetailleerd weer hoe het onderzoeksdesign eruitzag voor de jaargroepen 3, 4, 7 en 8. Het onderzoeksdesign voor de jaargroepen 5 en 6 is vergelijkbaar met dat voor jaargroep 4.

Tabel 4.1 Onderzoeksdesign jaargroep 3

sub-groep	medio 3		einde 3			
	AVI-M3	AVI-M3	AVI-M3	AVI-E3	AVI-E3	AVI-M4
1	A	B	C	A	B	
2	A	C	B	A	C	
3	B	A	C	B	A	
4	B	C	A	B	C	
5	C	A	B	C	A	
6	C	B	A	C	B	
7	A	B		A	B	C
8	A	C		A	C	B
9	B	A		B	A	C
10	B	C		B	C	A
11	C	A		C	A	B
12	C	B		C	B	A

Tabel 4.2 Onderzoeksdesign jaargroep 4

sub-groep	medio 4					einde 4				
	AVI-E3	AVI-M4	AVI-M4	AVI-E4	AVI-M5	AVI-M4	AVI-E4	AVI-E4	AVI-M5	AVI-E5
1	C	A	B			C	A	B		
2	A	B	C			A	B	C		
3	B	C	A			B	C	A		
4	C	A		B		C	A		B	
5	A	B		C		A	B		C	
6	B	C		A		B	C		A	
7		B	A	C			B	A	C	
8		C	B	A			C	B	A	
9		A	C	B			A	C	B	
10		A		B	C		A		B	C
11		B		C	A		B		C	A
12		C		A	B		C		A	B

Tabel 4.3 Onderzoeksdesign jaargroep 7

sub-groep	medio 7					einde 7			
	AVI-E6	AVI-M7	AVI-M7	AVI-E7	AVI-M8	AVI-M7	AVI-E7	AVI-E7	AVI-Plus
1	C	A	B			C	A	B	
2	A	B	C			A	B	C	
3	B	C	A			B	C	A	
4	C	A		B		C	A		B
5	A	B		C		A	B		C
6	B	C		A		B	C		A
7		B	A	C			B	A	C
8		C	B	A			C	B	A
9		A	C	B			A	C	B
10		A		B	C		C	A	B
11		B		C	A		A	B	C
12		C		A	B		B	C	A

Tabel 4.4 Onderzoeksdesign jaargroep 8

sub-groep	medio 8		
	AVI-E7	AVI-Plus	AVI-Plus
1	A	A	B
2	B	A	C
3	C	B	A
4	B	B	C
5	C	C	A
6	A	C	B
7	C	A	B
8	A	A	C
9	B	B	A
10	A	B	C
11	C	C	A
12	B	C	B

In de tabellen 4.1 tot en met 4.4 is te zien dat elke jaargroep tijdens het normeringsonderzoek is opgedeeld in 12 subgroepen. Naast de leeskaarten van de nieuwe AVI-toets, maakten alle subgroepen de DMT. Subgroepen 1 en 7 van alle jaargroepen kregen ook de Eén-Minuut-Test voorgelegd en ten slotte werd bij de leerlingen die nog niet niveau AVI-Plus hadden bereikt, 1 of meer leeskaarten afgenomen van de vorige uitgave van de AVI-toets. De verschillende parallelversies van de nieuwe AVI-leeskaarten zijn volgens een 'balanced incomplete block' principe aan subgroepen toegewezen. Dit betekent dat alle parallelversies (A, B of C) van de leeskaarten even vaak, in alle mogelijke combinaties, zijn afgenomen, en dat leerlingen tijdens het eindemoment nooit dezelfde parallelversie maakten als tijdens het mediomoment. Om tijdens het normeringsonderzoek gelijke afnamecondities te hebben voor alle leerlingen zijn toetsleiders ingezet. Deze toetsleiders zijn geworven onder pabo-studenten en studenten logopedie en onderwijskunde. De selectie van toetsleiders vond plaats op basis van bekwaamheid en op basis van hun reistijd tot de geworven scholen. De toetsleiders zijn voorafgaand aan het onderzoek getraind. Tijdens de training is aandacht besteed aan het doel en de opzet van het onderzoek, de inhoud van de materialen, de toetsprocedure en het invullen van de scoreformulieren. Zowel tijdens het medio- als tijdens het eindemoment waren ongeveer 45 toetsleiders actief.

Bij het normeringsonderzoek is ernaar gestreefd om elke subgroep te laten bestaan uit 45 leerlingen. Per afnamemoment zouden dan 540 leerlingen moeten deelnemen ($45 \text{ leerlingen} \times 12 \text{ subgroepen}$) en in totaal zou elke leeskaart van de AVI-toets minimaal 450 keer moeten worden afgenomen (minimaal 6 subgroepen op het afnamemoment dat overeenkomt met het AVI-niveau $\times 45$ leerlingen en daarnaast nog minimaal 4 subgroepen op de naastliggende afnamemomenten $\times 45$ leerlingen). Deze doelstelling is ruimschoots gerealiseerd. Tabel 4.5 laat zien hoeveel leerlingen bij elke afname hebben deelgenomen en hoe vaak elke leeskaart is opgelezen. Bij alle afnamemomenten is het gerealiseerde leerlingaantal groter dan het streefaantal van 540 en ook het aantal waarnemingen per leeskaart is steeds groter dan het streefaantal van 450.

Tabel 4.5 Aantal leerlingen in de steekproef uitgesplitst naar afnamemoment en AVI- leeskaart

Afname	M3	E3	M4	E4	M5	E5	M6	E6	M7	E7	M8	Totaal aantal leer- lingen
AVI-M3A	386	104	0	0	0	0	0	0	0	0	0	490
AVI-M3B	375	105	0	0	0	0	0	0	0	0	0	480
AVI-M3C	383	104	0	0	0	0	0	0	0	0	0	487
AVI-E3A	0	411	96	0	0	0	0	0	0	0	0	507
AVI-E3B	0	404	92	0	0	0	0	0	0	0	0	496
AVI-E3C	0	405	97	0	0	0	0	0	0	0	0	502
AVI-M4A	0	95	291	136	0	0	0	0	0	0	0	522
AVI-M4B	0	101	294	136	0	0	0	0	0	0	0	531
AVI-M4C	0	101	287	133	0	0	0	0	0	0	0	521
AVI-E4A	0	0	144	396	94	0	0	0	0	0	0	634
AVI-E4B	0	0	144	393	98	0	0	0	0	0	0	635
AVI-E4C	0	0	148	401	96	0	0	0	0	0	0	645
AVI-M5A	0	0	49	198	290	121	0	0	0	0	0	658
AVI-M5B	0	0	50	196	284	120	0	0	0	0	0	650
AVI-M5C	0	0	48	191	286	120	0	0	0	0	0	645
AVI-E5A	0	0	0	59	142	361	99	0	0	0	0	661
AVI-E5B	0	0	0	64	143	363	98	0	0	0	0	668
AVI-E5C	0	0	0	64	143	360	98	0	0	0	0	665
AVI-M6A	0	0	0	0	47	180	305	125	0	0	0	657
AVI-M6B	0	0	0	0	48	178	302	126	0	0	0	654
AVI-M6C	0	0	0	0	48	180	306	122	0	0	0	656
AVI-E6A	0	0	0	0	0	59	151	372	97	0	0	679
AVI-E6B	0	0	0	0	0	60	155	373	98	0	0	686
AVI-E6C	0	0	0	0	0	61	151	377	98	0	0	687
AVI-M7A	0	0	0	0	0	0	53	185	305	109	0	652
AVI-M7B	0	0	0	0	0	0	53	184	302	111	0	650
AVI-M7C	0	0	0	0	0	0	53	187	302	109	0	651
AVI-E7A	0	0	0	0	0	0	0	62	151	374	195	782
AVI-E7B	0	0	0	0	0	0	0	61	156	375	209	801
AVI-E7C	0	0	0	0	0	0	0	58	153	379	205	795
AVI-Plus	0	0	0	0	0	0	0	0	54	163	409	626
AVI-Plus	0	0	0	0	0	0	0	0	51	160	406	617
AVI-Plus	0	0	0	0	0	0	0	0	54	158	403	615
Totaal aantal waar- nemingen per afname- moment	1144	1830	1740	2367	1719	2163	1824	2232	1821	1938	1827	
Totaal aantal leerlingen per afname- moment	572	610	580	789	573	721	608	744	607	646	609	20605

4.2 Representativiteit van de normeringssteekproef

In deze paragraaf beschrijven we de samenstelling van de AVI-steekproef in relatie tot de landelijke populatie. Doorgaans streven we ernaar om een normeringsteekproef een representatieve afspiegeling te laten zijn van de populatie, maar voor de normering van de AVI-toets is representativiteit weliswaar fraai, maar geen vereiste. Zoals al aangegeven in paragraaf 2.3 is de normering van de nieuwe uitgave van de AVI-toets gebaseerd op de prestatie-eisen die golden bij de vorige uitgave. Dit wil zeggen dat we er bij het opstellen van de normtabellen voor gezorgd hebben dat leerlingen op basis van de nieuwe AVI-toets (nagenoeg) hetzelfde geclassificeerd worden als op basis van de oude AVI-toets (zie paragraaf 4.5). Ook voor de constructie van de meetschaal (kalibratie, zie paragraaf 4.4) is representativiteit niet noodzakelijk. Hoewel representativiteit dus niet per se nodig is, bespreken we de AVI-steekproeven hieronder voor de volledigheid wel in dergelijke termen. In de analyses, waar we kijken naar prestatieverschillen, zijn we uitgegaan van de leestijden van de leerlingen, omdat deze variabele het meest onderscheidend is als het gaat om het verschil tussen goede en zwakke leerlingen (zie ook Verhoeven & Leeuwe, 2003).

Tijdens de dataverzameling is informatie verzameld over regio, urbanisatiegraad, schooltype en sekse. De variabelen zijn als volgt gedefinieerd:

- Regio. Bij de definitie van de variabele Regio is uitgegaan van de CBS-indeling naar landsdeel. Dit betekent dat er vier regio's te onderscheiden zijn. Regio Noord omvatte de provincies Groningen, Friesland en Drenthe; regio Oost de provincies Overijssel, Gelderland en Flevoland; regio West de provincies Utrecht, Noord-Holland, Zuid-Holland en Zeeland en regio Zuid de provincies Noord-Brabant en Limburg.
- Urbanisatiegraad. De gehanteerde verdeling naar urbanisatiegraad of mate van verstedelijking is terug te voeren op de bij het CBS gebruikelijke indeling naar vijf niveaus, namelijk zeer sterk, sterk, matig, weinig en niet verstedelijkt. Voor het onderzoek is de gebruikelijke vijfdeling echter gereduceerd tot een tweedeling in enerzijds niet tot matig verstedelijkt (platteland) en anderzijds sterk tot zeer sterk verstedelijkt (stad). Een dergelijke tweedeling blijkt in de praktijk goed te volstaan (cf. Van Boxtel & Hemker, 2009).
- Schooltype. Bij de definitie van deze variabele is gebruikgemaakt van de formatiegewichten van de leerlingen binnen een school. Ten tijde van het onderzoek onderscheidde OCW drie niveaus die gebaseerd waren op het opleidingsniveau van de ouders:
 - 0.0 een 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 een van de ouders heeft een opleiding gehad uit categorie 1 en de ander een opleiding uit categorie 1 óf 2In deze indeling wordt verwezen naar de volgende categorieën in het opleidingsniveau van de ouders: 1 = maximaal basisonderwijs of (V)SO-ZMLK, 2 = maximaal LBO/VBO, praktijkonderwijs of VMBO basis- of kaderberoepsgerichte leerweg, en 3 = overig VO en hoger. Leerlingen met een formatiegewicht van 0.3 of 1.2 zijn te definiëren als achterstandsleerlingen. Scholen zijn ingedeeld naar het percentage achterstandsleerlingen volgens een indeling in twee typen: (1) percentage achterstandsleerlingen [0, .15) en (2) percentage achterstandsleerlingen [.15, 1]. Dezelfde indeling wordt door de Inspectie van het Onderwijs gehanteerd bij de opbrengstbeoordeling in het basisonderwijs.
- Sekse. In de analyses omtrent sekse is vanzelfsprekend een tweedeling naar jongens en meisjes gehanteerd.

Voor het bepalen van de populatiegegevens is gebruikgemaakt van gegevens van DUO (data.duo.nl) en CBS (statline.cbs.nl).

De representativiteitsanalyse bestond uit 2 onderdelen. Ten eerste is per achtergrondvariabele een vergelijking gemaakt tussen de verdeling van leerlingen in een steekproef en de verdeling van leerlingen in

de populatie. Via een χ^2 -toetsing is nagegaan of eventuele afwijkingen significant waren. Bij grotere steekproeven zegt significantie echter niet zoveel. Daarom is bij de interpretatie vooral gekeken naar de effectgrootte ϕ van de χ^2 -toetsing. Een ϕ van .10 kan aangemerkt worden als een klein effect, een ϕ van .30 als een gemiddeld effect, en een ϕ van .50 als een groot effect (Cohen, 1988). Ten tweede is nagegaan of de prestaties van groepen leerlingen betekenisvol van elkaar verschilden. In het regressiemodel waren de leestijden de afhankelijke variabele en regio, urbanisatiegraad, schooltype, en sekse de verklarende variabelen. De regressieparameters zijn geïnterpreteerd in termen van significantie (z) en relevantie (d). De z -waarden zijn bepaald door de regressieparameters te delen door de standaardfout en de effectgroottes d door de regressieparameters te delen door de residuele standaarddeviatie. We spreken van een klein effect als $d = .20$, van een gemiddeld effect als $d = .50$, en van een groot effect als $d = .80$ (Cohen, 1988).

Representativiteit normgroep 1: medio groep 3

Voor afnamemoment medio 3 bestond de steekproef uit 572 leerlingen. Tabel 4.6 laat zien in hoeverre de steekproef representatief was voor regio, urbanisatiegraad, schooltype en sekse. We zien dat de steekproef niet representatief is te noemen als het gaat om de achtergrondvariabelen regio, urbanisatiegraad en schooltype. Hoewel de afwijkingen niet extreem groot zijn, zien we een oververtegenwoordiging van de zuidelijke provincies, de plattelandsgebieden en de scholen met relatief weinig achterstandsl leerlingen. Van deze variabelen is regio echter de enige die er enigszins toe doet. De zuidelijke provincies presteren zwakker dan de overige provincies ($z > 1.96$; $d > 0.20$). Bij de variabelen urbanisatiegraad en schooltype verschillen de groepen leerlingen niet betekenisvol van elkaar. De verhouding tussen het aantal jongens en meisjes komt in hoge mate overeen met de verhouding die we ook in de landelijke populatie vinden ($\phi = .07$). Bovendien scoren jongens niet significant hoger of lager dan meisjes op de AVI-toets ($z = -0.23$, $d = -0.02$).

Tabel 4.6 Resultaten representativiteits- en regressieanalyse medio groep 3

Variabele	Categorie	Procentuele verdeling in		Percentuele verdeling in de steek-						
		de populatie	N steekproef	proef	χ^2	df	p	ϕ	z	d
Regio	noord	9.91	58	10.14	11.39	3	0.01	0.14	---	---
	oost	22.23	102	17.83					-1.52	-0.26
	west	48.22	273	47.73					-1.01	-0.16
	zuid	19.63	139	24.30					-2.02	-0.34
Urbanisatie	platteland	54.29	367	64.16	22.46	1	0.00	0.20	---	---
	stad	45.71	205	35.84					-0.96	-0.09
Schooltype	[0, .15)	81.23	512	89.51	25.72	1	0.00	0.21	---	---
	[.15, 1]	18.77	60	10.49					-1.70	-0.23
Sekse	jongen	50.46	256	47.15	2.39	1	0.12	0.07	---	---
	meisje	49.54	287	52.85					-0.23	-0.02

Representativiteit normgroep 2: eind groep 3

Voor afnamemoment einde 3 bestond de steekproef uit 610 leerlingen. In tabel 4.7 is te zien dat de sterk en zeer sterk verstedelijkte gebieden ondervertegenwoordigd waren in de steekproef. Datzelfde geldt voor noordelijke en westelijke provincies, en de scholen met relatief veel achterstandsl leerlingen. De steekproef is wel een adequate afspiegeling van de populatie als het gaat om de variabele sekse. Als het om prestatieverschillen gaat, blijken niet alle variabelen er ook daadwerkelijk toe te doen. Alleen de prestaties van leerlingen op de scholen met relatief veel en relatief weinig achterstandsl leerlingen verschillen opvallend van elkaar ($z = -3.64$, $d = -0.64$).

Tabel 4.7 Resultaten representativiteits- en regressieanalyse eind groep 3

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef						
				χ^2	df	p	φ	z	d	
Regio	noord	9.91	41	6.72	10.95	3	0.01	0.13	---	---
	oost	22.23	158	25.90					0.00	0.00
	west	48.22	283	46.39					0.66	0.11
	zuid	19.63	128	20.98					-0.80	-0.16
Urbanisatie	platteland	54.29	480	78.69	146.33	1	0.00	0.49	---	---
	stad	45.71	130	21.31					-1.51	-0.18
Schooltype	[0, .15)	81.23	558	91.48	42.00	1	0.00	0.26	---	---
	[.15, 1]	18.77	52	8.52					-3.64	-0.64
Sekse	jongen	50.46	251	46.83	2.83	1	0.09	0.07	---	---
	meisje	49.54	285	53.17					-1.35	-0.12

Representativiteit normgroep 3: medio groep 4

Voor afnamemoment medio 4 bestond de steekproef uit 580 leerlingen. In tabel 4.8 is te zien dat de verdeling van leerlingen over de onderscheiden groepen grote gelijkens vertoont met de verdelingen die we eerder vonden bij de normeringssteekproeven voor medio en einde groep 3. Opnieuw zijn de zuidelijke provincies en de plattelandsgebieden oververtegenwoordigd en is de verhouding tussen het aantal jongens en meisjes in overeenstemming met de landelijke populatie. Ook de scholen met relatief veel achterstandsl leerlingen zijn representatief ondervertegenwoordigd in de normeringssteekproef. Prestatieverschillen tussen leerlingen kunnen we verklaren vanuit variabelen urbanisatiegraad en sekse ($z > 1.96$; $d > 0.20$).

Tabel 4.8 Resultaten representativiteits- en regressieanalyse medio groep 4

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef						
				χ^2	df	p	φ	z	d	
Regio	noord	9.91	35	6.03	20.51	3	0.00	0.19	---	---
	oost	22.23	113	19.48					0.57	0.11
	west	48.22	285	49.14					1.40	0.26
	zuid	19.63	147	25.34					1.06	0.21
Urbanisatie	platteland	54.29	346	59.66	6.73	1	0.01	0.11	---	---
	stad	45.71	234	40.34					-2.90	-0.28
Schooltype	[0, .15)	81.23	472	81.38	0.01	1	0.93	0.00	---	---
	[.15, 1]	18.77	108	18.62					1.80	0.20
Sekse	jongen	50.46	288	51.99	0.52	1	0.47	0.03	---	---
	meisje	49.54	266	48.01					2.66	0.23

Representativiteit normgroep 4: einde groep 4

Voor afnamemoment einde 4 bestond de steekproef uit 789 leerlingen. In tabel 4.9 is te zien dat plattelandsgebieden, de zuidelijke provincies en de scholen met relatief weinig achterstandsl leerlingen oververtegenwoordigd waren in de steekproef. Met name bij de variabelen regio en urbanisatiegraad zijn de afwijkingen tussen de verdelingen in de steekproef en de populatie behoorlijk groot ($\varphi > .30$).

De voorspellende waarde van de variabelen blijkt echter gering. Jongen en meisjes, leerlingen in verschillende regio's en leerlingen op verschillende typen scholen blijken vergelijkbaar te presteren.

Leerlingen in verstedelijkte gebieden scoren wel beduidend lager dan hun jaargroepgenoten op het platteland ($d > 0.30$).

Tabel 4.9 Resultaten representativiteits- en regressieanalyse eind groep 4

Variabele	Categorie	Procentuele		Percentuele		χ^2	df	p	ϕ	z	d
		verdeling in	N steek-	verdeling in	de steek-						
		de populatie	proef	de steek-	proef						
Regio	noord	9.91	53	6.72	83.33	3	0.00	0.33	---	---	
	oost	22.23	206	26.11					-0.90	-0.14	
	west	48.22	289	36.63					0.93	0.14	
	zuid	19.63	241	30.54					0.49	0.09	
Urbanisatie	platteland	54.29	561	71.10	89.87	1	0.00	0.34	---	---	
	stad	45.71	228	28.90					-2.78	-0.34	
Schooltype	[0, .15)	81.23	718	91.00	49.41	1	0.00	0.25	---	---	
	.[15, 1]	18.77	71	9.00					0.11	0.02	
Sekse	jongen	50.46	345	49.15	0.49	1	0.49	0.03	---	---	
	meisje	49.54	357	50.85					0.96	0.07	

Representativiteit normgroep 5: medio groep 5

Voor afnamemoment medio 5 bestond de steekproef uit 573 leerlingen. In tabel 4.10 is te zien dat meisjes relatief sterk vertegenwoordigd waren in de steekproef. Waar we verwachten dat ongeveer 50 procent van de leerlingen in de steekproef een jongen is en de andere 50 procent een meisje, is de verhouding in onze steekproef 45 procent jongens tegenover 55 procent meisjes. Verder zijn de verstedelijkte gebieden en de oostelijke provincies iets ondervertegenwoordigd. Het aantal scholen met relatief veel en relatief weinig achterstandsleerlingen is in balans met de aantallen die we vinden in de populatie. Prestatieverschillen tussen leerlingen kunnen we uit geen van de achtergrondvariabelen verklaren; de z-waarden zijn allemaal kleiner dan 1.96 en de effectgroottes zijn klein te noemen ($d < .20$).

Tabel 4.10 Resultaten representativiteits- en regressieanalyse medio groep 5

Variabele	Categorie	Procentuele		Percentuele		χ^2	df	p	ϕ	z	d
		verdeling in	N steek-	verdeling in	de steek-						
		de populatie	proef	de steek-	proef						
Regio	noord	9.91	55	9.60	33.68	3	0.00	0.24	---	---	
	oost	22.23	91	15.88					1.12	0.19	
	west	48.22	264	46.07					0.90	0.14	
	zuid	19.63	163	28.45					0.00	0.00	
Urbanisatie	platteland	54.29	359	62.65	16.15	1	0.00	0.17	---	---	
	stad	45.71	214	37.35					-0.89	-0.08	
Schooltype	[0, .15)	81.23	466	81.33	0.00	1	0.96	0.00	---	---	
	.[15, 1]	18.77	107	18.67					-1.09	-0.12	
Sekse	jongen	50.46	247	45.49	5.37	1	0.02	0.10	---	---	
	meisje	49.54	296	54.51					0.99	0.09	

Representativiteit normgroep 6: einde groep 5

Voor afnamemoment einde 5 bestond de steekproef uit 721 leerlingen. In tabel 4.11 is te zien dat de steekproef voor dit afnamemoment sterk lijkt op de steekproef voor afnamemoment medio 5. Opnieuw heeft geen van de achtergrondvariabelen een betekenisvol effect op de prestaties van leerlingen.

De verstedelijkte gebieden zijn wederom ondervertegenwoordigd en ook de verdeling van leerlingen over de verschillende regio's wijkt opvallend af van de verdeling in de landelijke populatie. Het aantal jongens is nu wel in evenwicht met het aantal meisjes, maar in plaats daarvan is de verhouding tussen de scholen met relatief veel en relatief weinig achterstandsl leerlingen nu iets uit balans ten opzichte van de landelijke populatie.

Tabel 4.11 Resultaten representativiteits- en regressieanalyse eind groep 5

Variabele	Categorie	Procentuele verdeling in		Percentuele verdeling in de steek-						
		de populatie	N steek-proef	proef	χ^2	df	p	ϕ	z	d
Regio	noord	9.91	55	7.63	125.59	3	0.00	0.42	---	---
	oost	22.23	150	20.80					-0.87	-0.14
	west	48.22	257	35.64					0.77	0.12
	zuid	19.63	259	35.92					-0.03	-0.01
Urbanisatie	platteland	54.29	469	65.05	33.63	1	0.00	0.22	---	---
	stad	45.71	252	34.95					-0.39	-0.04
Schooltype	[0, .15)	81.23	628	87.10	16.30	1	0.00	0.15	---	---
	[.15, 1]	18.77	93	12.90					-1.46	-0.20
Sekse	jongen	50.46	339	48.92	0.66	1	0.42	0.03	---	---
	meisje	49.54	354	51.08					2.25	0.17

Representativiteit normgroep 7: medio groep 6

Voor afnamemoment medio 6 bestond de steekproef uit 608 leerlingen. In tabel 4.12 is te zien dat de verschillende regio's representatief vertegenwoordigd waren in de steekproef. Ook de verhouding tussen het aantal jongens en meisjes sloot aan bij de verhouding die we mogen verwachten op basis van landelijke gegevens. Dat geldt niet voor urbanisatiegraad. De plattelandsgebieden waren iets oververtegenwoordigd in de steekproef. De variabele regio is een voorspeller van de toetsscore (leestijd). De overige variabelen hebben geen effect op de prestaties van leerlingen.

Tabel 4.12 Resultaten representativiteits- en regressieanalyse medio groep 6

Variabele	Categorie	Procentuele		Percentuele		χ^2	df	p	φ	z	d
		verdeling in de populatie	N steekproef	verdeling in de steekproef	de steekproef						
Regio	noord	9.91	45	7.40	8.34	3	0.04	0.12	---	---	
	oost	22.23	119	19.57					1.60	0.28	
	west	48.22	319	52.47					2.17	0.36	
	zuid	19.63	125	20.56					1.82	0.33	
Urbanisatie	platteland	54.29	359	59.05	5.54	1	0.02	0.10	---	---	
	stad	45.71	249	40.95					-1.36	-0.13	
Schooltype	[0, .15)	81.23	508	83.55	2.15	1	0.14	0.06	---	---	
	[.15, 1]	18.77	100	16.45					1.04	0.12	
Sekse	jongen	50.46	290	49.83	0.09	1	0.76	0.01	---	---	
	meisje	49.54	292	50.17					1.18	0.10	

Representativiteit normgroep 8: einde groep 6

Voor afnamemoment einde 6 bestond de steekproef uit 744 leerlingen. In tabel 4.13 is te zien dat de verdeling van leerlingen over de verschillende categorieën van de achtergrondvariabelen regio, urbanisatiegraad en schooltype afwijkt van de verdeling die we op grond van landelijke gegevens verwachten. Achterstandsleerlingen en leerlingen in verstedelijkte gebieden zijn iets ondervertegenwoordigd in de steekproef. Datzelfde geldt voor leerlingen in de westelijke provincies. De verhouding tussen het aantal jongens en meisjes is nagenoeg gelijk aan de verhouding in de populatie. Bij de AVI-toets kunnen de prestaties van leerling enigszins voorspeld worden vanuit sekse en regio; meisjes presteren beter dan jongens en de noordelijke provincies presteren zwakker dan de overige provincies.

Tabel 4.13 Resultaten representativiteits- en regressieanalyse eind groep 6

Variabele	Categorie	Procentuele		Percentuele		χ^2	df	p	φ	z	d
		verdeling in de populatie	N steekproef	verdeling in de steekproef	de steekproef						
Regio	noord	9.91	52	6.99	94.47	3	0.00	0.36	---	---	
	oost	22.23	158	21.24					0.34	0.05	
	west	48.22	285	38.31					2.06	0.32	
	zuid	19.63	249	33.47					0.98	0.17	
Urbanisatie	platteland	54.29	448	60.22	10.53	1	0.00	0.12	---	---	
	stad	45.71	296	39.78					-1.56	-0.16	
Schooltype	[0, .15)	81.23	646	86.83	15.29	1	0.00	0.14	---	---	
	[.15, 1]	18.77	98	13.17					-1.20	-0.16	
Sekse	jongen	50.46	360	51.58	0.35	1	0.56	0.02	---	---	
	meisje	49.54	338	48.42					2.02	0.15	

Representativiteit normgroep 9: medio groep 7

Voor afnamemoment medio 7 bestond de steekproef uit 607 leerlingen. De uitkomsten van de representativiteitsanalyse zijn te vinden in tabel 4.14. We zien dat de steekproef op onderdelen afwijkt van de landelijke populatie. De verdeling van leerlingen over de verschillende categorieën van de variabelen

regio, urbanisatiegraad en schooltype is onvoldoende representatief als we ons baseren op de eerder benoemde richtlijn voor het interpreteren van coëfficiënt ϕ . Het aantal jongens en meisjes is wel in voldoende mate in overeenstemming met de landelijke populatie. De regressieanalyse laat zien dat er sprake is van een betekenisvol effect van regio op de prestaties van leerlingen. Bij de overige variabelen bestaan er geen significante verschillen tussen de onderscheiden groepen leerlingen.

Tabel 4.14 Resultaten representativiteits- en regressieanalyse medio groep 7

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef						
				χ^2	df	p	ϕ	z	d	
Regio	noord	9.91	43	7.08	39.35	3	0.00	0.26	---	---
	oost	22.23	86	14.17					2.50	0.47
	west	48.22	317	52.22					0.67	0.11
	zuid	19.63	161	26.52					2.56	0.45
Urbanisatie	platteland	54.29	362	59.64	7.00	1	0.01	0.11	---	---
	stad	45.71	245	40.36					-1.84	-0.17
Schooltype	[0, .15)	81.23	519	85.50	7.27	1	0.01	0.11	---	---
	[.15, 1]	18.77	88	14.50					0.91	0.11
Sekse	jongen	50.46	275	47.83	1.60	1	0.21	0.05	---	---
	meisje	49.54	300	52.17					0.92	0.08

Representativiteit normgroep 10: einde groep 7

Voor afnamemoment einde 7 bestond de steekproef uit 646 leerlingen. Tabel 4.15 laat zien dat de steekproef grote gelijkenis vertoont met de landelijke populatie als het gaat om sekse en schooltype. De verstedelijkte gebieden blijken echter iets ondervertegenwoordigd en met name bij regio is er sprake van behoorlijk grote afwijkingen. De noordelijke provincies maakten slechts beperkt deel uit van de steekproef, terwijl de zuidelijke provincies zeer ruim vertegenwoordigd waren. Coëfficiënt ϕ wijst op gemiddelde effecten. Geen van de variabelen blijkt van invloed op de prestaties van leerlingen ($z < 1.96$; $d < .20$).

Tabel 4.15 Resultaten representativiteits- en regressieanalyse eind groep 7

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef						
				χ^2	df	p	ϕ	z	d	
Regio	noord	9.91	32	4.95	96.74	3	0.00	0.39	---	---
	oost	22.23	109	16.87					-0.34	-0.07
	west	48.22	284	43.96					0.12	0.02
	zuid	19.63	221	34.20					1.12	0.23
Urbanisatie	platteland	54.29	405	62.69	18.38	1	0.00	0.17	---	---
	stad	45.71	241	37.31					-0.86	-0.09
Schooltype	[0, .15)	81.23	541	83.75	2.68	1	0.10	0.06	---	---
	[.15, 1]	18.77	105	16.25					-0.25	-0.03
Sekse	jongen	50.46	293	49.16	0.40	1	0.53	0.03	---	---
	meisje	49.54	303	50.84					-0.39	-0.03

Representativiteit normgroep 11: medio groep 8

Voor afnamemoment medio 8 bestond de steekproef uit 609 leerlingen. Tabel 4.16 laat de uitkomsten van de representativiteitsanalyse zien. Er is sprake van een grote gelijkens met de steekproef voor einde 7. Waar er sprake is van een disbalans bij de variabelen regio en urbanisatiegraad, is de steekproef wel voldoende in overeenstemming met de landelijke populatie als we kijken naar de variabelen schooltype en sekse. Opnieuw bestaat er geen verband tussen de achtergrondvariabelen en de prestaties van leerlingen.

Tabel 4.16 Resultaten representativiteits- en regressieanalyse medio groep 8

Variabele	Categorie	Procentuele verdeling in		Percentuele verdeling in		χ^2	df	p	ϕ	z	d
		de populatie	N steekproef	de steekproef	de steekproef						
Regio	noord	9.91	43	7.06	25.68	3	0.00	0.21	---	---	
	oost	22.23	105	17.24					1.41	0.26	
	west	48.22	301	49.43					0.49	0.08	
	zuid	19.63	160	26.27					1.67	0.30	
Urbanisatie	platteland	54.29	375	61.58	13.03	1	0.00	0.15	---	---	
	stad	45.71	234	38.42					-1.76	-0.16	
Schooltype	[0, .15)	81.23	505	82.92	1.15	1	0.29	0.04	---	---	
	[.15, 1]	18.77	104	17.08					-0.56	-0.06	
Sekse	jongen	50.46	275	48.08	1.30	1	0.25	0.05	---	---	
	meisje	49.54	297	51.92					0.72	0.06	

Samenvattend leveren de representativiteitsanalyses het volgende beeld op:

- De normeringssteekproeven laten een grote onderlinge gelijkens zien. Dit is niet vreemd als we bedenken dat het grotendeels om kinderen van dezelfde scholen gaat.
- De normeringssteekproeven hebben met elkaar gemeen dat de verdelingen naar regio en urbanisatiegraad afwijken van de populatieverdeling. De zuidelijke provincies en de plattelandsgebieden zijn oververtegenwoordigd.
- In de meeste normeringssteekproeven is er sprake van afwijkingen in de verdeling naar schooltype. Met uitzondering van afnamemomenten medio 4, medio 5, medio 6, einde 7 en medio 8 zijn de scholen met weinig achterstandsleerlingen oververtegenwoordigd ten opzicht van de scholen met veel achterstandsleerlingen.
- In vrijwel alle normeringssteekproeven is de verdeling naar sekse in overeenstemming met de populatieverdeling. Op dit punt zijn de normsteekproeven dus representatief te noemen. Afnamemoment medio 5 vormt een uitzondering; daar zijn de meisjes licht oververtegenwoordigd ten opzichte van de jongens.
- De regressieanalyses laten geen consistent beeld zien over de afnamemomenten. Soms lijkt er een relatie te bestaan tussen een bepaalde variabele en de prestaties van leerlingen, maar een duidelijk patroon over de afnamemomenten ontbreekt. Er lijkt eerder sprake te zijn van toeval dan van een systematisch effect.

Hoewel uit de representativiteitsanalyses blijkt dat de AVI-steekproeven qua samenstelling soms afwijken van de landelijke populatie is dat, zoals al eerder aangegeven, voor de kalibratie en normering niet erg. De kalibratie en normering zijn dan ook op de hiervoor beschreven steekproeven gebaseerd.

4.3 Beschrijving van het meetmodel

De AVI-toets is geanalyseerd binnen het raamwerk van het Poisson model van Rasch (RPCM; Rasch, 1960). Het RPCM is een uni-dimensionaal item respons model en is geschikt voor het analyseren van situaties waarin de toetsscore gelijk is aan het aantal successen (of mislukkingen) binnen de gestelde tijd/omvang, of waarin de toetsscore gelijk is aan de tijd die nodig is om een bepaald aantal woorden te lezen. Bij de AVI-toets moeten leerlingen de leeskaarten hardop voorlezen. Tijdens de toetsafname wordt de leestijd en het aantal fouten bijgehouden. Er is voor gekozen om de leestijden en het aantal leesfouten los van elkaar te modelleren.

4.3.1 Modelleren leesfouten

Een veelgebruikt model om variabiliteit in frequenties (aantallen) te beschrijven is het Poisson-model. In zijn meest elementaire vorm beschrijft dit model de kans op een score s als

$$(1) \quad P(s) = \frac{\lambda^s}{s!} e^{-\lambda}$$

waarin $\lambda (> 0)$ de parameter van de verdeling is en de uitdrukking $s!$ (lees: s -faculteit of s -factorieel) het product aanduidt $s \times (s - 1) \times (s - 2) \times \dots \times 2 \times 1$. Als $s = 0$, wordt $0!$ gedefinieerd als gelijk aan 1. Rasch (1960) heeft dit model uitgebreid door aan te nemen dat de waarde van de parameter λ afhankelijk is van de persoon en van de taak. Toegepast op de AVI-toets leidt dit tot het volgende model:

$$(2) \quad \lambda_{vj} = \tau_j \sigma_j \theta_v$$

waarin λ_{vj} de waarde van de Poisson-parameter voorstelt wanneer leerling v taak j maakt. Deze parameter is *multiplicatief* samengesteld en bestaat uit drie componenten. De grootte τ_j is geen parameter maar een bekende constante die de omvang van de taak uitdrukt. Bij de AVI-toets is dit het totale aantal woorden op een leeskaart, variërend van 98 (kaart M3C) tot 224 (kaart E7C). De parameter σ_j drukt de moeilijkheid van de taak uit: hoe kleiner deze parameter hoe makkelijker het is om de woorden binnen de taak correct te lezen, en de grootte θ_v is de technische leesvaardigheid van de leerling v . De drie componenten van de Poisson-parameter zijn continue variabelen die alleen positieve waarden of de waarde nul kunnen aannemen. De doelstelling van het gebruik van dit meetmodel is een inferentie te maken omtrent de vaardigheid van de leerling (θ_v) die gebaseerd is op de observatie die we van deze leerling hebben wanneer hij kaart j leest: de score s_{vj} .

De kansverdeling van de scores is afhankelijk van de Poisson-parameter λ , die op zijn beurt weer samengesteld is uit 3 componenten. Als de 3 componenten gegeven zijn, ligt hun product vast, maar als het product gegeven is liggen de 3 componenten niet uniek vast; we kunnen namelijk altijd een component delen door een positieve constante c en 1 van de 2 andere vermenigvuldigen met c , en een dergelijke operatie laat het product onveranderd. Om zinvol over de waarde van de componenten te kunnen spreken moeten we 2 *normalisaties* invoeren. Voor de AVI-toets hebben we dit als volgt gedaan:

- 1 We kiezen een lengtemaat. Voor de AVI-toets is als lengtemaat het aantal eenheden van 50 woorden gekozen.
- 2 De AVI-toets bestaat uit 33 verschillende leeskaarten en voor elke leeskaart moet een σ -parameter worden geschat. We hebben de restrictie opgelegd dat het product van deze parameters gelijk moet zijn aan 1.

Een echte of denkbeeldige kaart waarvan de σ -parameter gelijk is aan 1, noemen we een *standaardkaart*. De interpretatie van de vaardigheid θ_v in dit model is zeer elegant. Als een variabele Poisson-verdeeld is (formule (1)), dan is het gemiddelde of de verwachte waarde van de verdeling gelijk aan de Poisson-

parameter λ . Als leerling v kaart i leest, dan is zijn score S_{vi} een toevalsvariabele, en door (1) en (2) te combineren krijgen we dus

$$(3) \quad E(S_{vj}) = \lambda_{vj} = \tau_j \sigma_j \theta_v$$

Als we nu alle grootheden gaan benoemen, zien we het volgende:

- S_{vj} is een aantal foutief gelezen woorden; zijn verwachte waarde is dus ook een aantal foutief gelezen woorden.
- τ_j is de gekozen lengtemaat van 50 woorden.
- σ_j vatten we op als een onbenoemd getal, een soort correctiefactor die controleert voor de verschillende moeilijkheid van de woorden op de kaarten. De standaardkaart heeft een σ -waarde gelijk aan 1.
- Om het product van de rechterzijde in (3) in dezelfde eenheid uit te drukken als de linkerzijde, moeten we θ_v benoemen als het aantal foutief gelezen woorden per eenheid van 50 woorden op een standaardkaart.

Lokale stochastische onafhankelijkheid

Wat tot hiertoe beschreven is, is de modellering van de uitkomsten wanneer een leerling één enkele kaart leest. Maar er moet ook iets gezegd worden over de verdeling van de uitkomsten wanneer een leerling twee of meer kaarten leest. Daarom moet er een extra veronderstelling aan het model worden toegevoegd en deze veronderstelling wordt meestal aangeduid als lokale stochastische of conditionele onafhankelijkheid. De veronderstelling bestaat eigenlijk uit twee delen:

- 1 De vaardigheid van de leerling θ_v blijft onveranderd bij het lezen van meerdere kaarten op één en hetzelfde afnamemoment.
- 2 Bij elke kaart is de (Poisson-)verdeling van de score alleen afhankelijk van de eigenschappen van de kaart en de leestijd (de σ - en de τ -parameter) en van de vaardigheid (θ_v). En niet van de score die de leerling op een van de andere kaarten heeft behaald.

Als we deze veronderstelling aannemen, kunnen we gebruikmaken van een andere eigenschap van de Poisson-verdeling: als een (eindig) aantal toevalsvariabelen S_1, S_2, \dots, S_k onafhankelijk Poisson-verdeeld is met parameters $\lambda_1, \lambda_2, \dots, \lambda_k$, dan is hun som $S = S_1 + S_2 + \dots + S_k$ Poisson-verdeeld met parameter $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_k$. Hier is een voorbeeld: veronderstel dat leerling v drie van de negen kaarten leest, waarna voor elke kaart zijn score wordt bepaald. De som van deze drie scores $S_v = S_{v1} + S_{v2} + S_{v3}$ is Poisson-verdeeld met parameter

$$\tau_1 \sigma_1 \theta_v + \tau_2 \sigma_2 \theta_v + \tau_3 \sigma_3 \theta_v = \theta_v (\tau_1 \sigma_1 + \tau_2 \sigma_2 + \tau_3 \sigma_3)$$

Dit voorbeeld veralgemeniseren we nu als volgt. Stel dat er in het proefonderzoek in totaal k kaarten zijn gebruikt, en dat elke leerling een aantal van deze kaarten heeft gelezen (volgens een vooraf vastgesteld design), dan definiëren we voor elke leerling de grootheid

$$(4) \quad \delta_v = \sum_{j=1}^k d_{vj} \tau_j \sigma_j$$

waarin d_{vj} een designvariabele is die de waarde 1 aanneemt als leerling v kaart j heeft gelezen en 0 als dit niet het geval is. In het voorbeeld hierboven is $d_{vj} = 1$ voor $j = 1, 2, 3$ en 0 voor alle andere kaarten. Met deze notatie kunnen we het model specificeren wanneer een leerling meerdere kaarten leest: de somscore is Poisson-verdeeld met parameter

$$(5) \quad \lambda_v = \theta_v \delta_v$$

Hoewel formule (5) er heel eenvoudig uitziet, dient men te bedenken dat de parameter δ_v een behoorlijk complexe structuur heeft: hij is afhankelijk van het dataverzamelingsdesign (de specifieke kaarten die de leerling heeft gelezen), van het aantal woorden op de leeskaarten en van de onbekende moeilijkheidsparameters σ (zie formule (4)).

Schatting van de individuele vaardigheid θ_v

De eerste stap is het schatten van de taakparameters σ_i . De procedure waarmee dit gebeurt, staat gedetailleerd beschreven in Verhelst & Kamphuis (2009) en wordt hier verder niet uiteengezet. Belangrijk is dat deze moeilijkheidsparameters op een consistente manier kunnen worden geschat zonder dat men een aanname hoeft te maken over de verdeling van de technische leesvaardigheid in de populatie. Omdat de normeringssteekproef zeer groot is in vergelijking met het aantal te schatten parameters, is de schattingsfout van deze parameters vrij klein en kan ze voor praktische doeleinden worden verwaarloosd. Dit wil zeggen dat we de schattingen van deze parameters verder kunnen behandelen als de echte waarden en dus ook dat we voor elke leerling de parameter δ_v met behulp van formule (4) kunnen uitrekenen. De schatting van de vaardigheid van leerling v wordt gegeven door

$$(6) \quad \hat{\theta}_v = \frac{s_v}{\delta_v}$$

en de standaardfout (SE) door

$$(7) \quad SE(\hat{\theta}_v) = \frac{\sqrt{s_v}}{\delta_v}$$

In beide formules betekent s_v de geobserveerde somscore die door leerling v is behaald. Merk op dat in dit model de standaardfout toeneemt met de behaalde score.

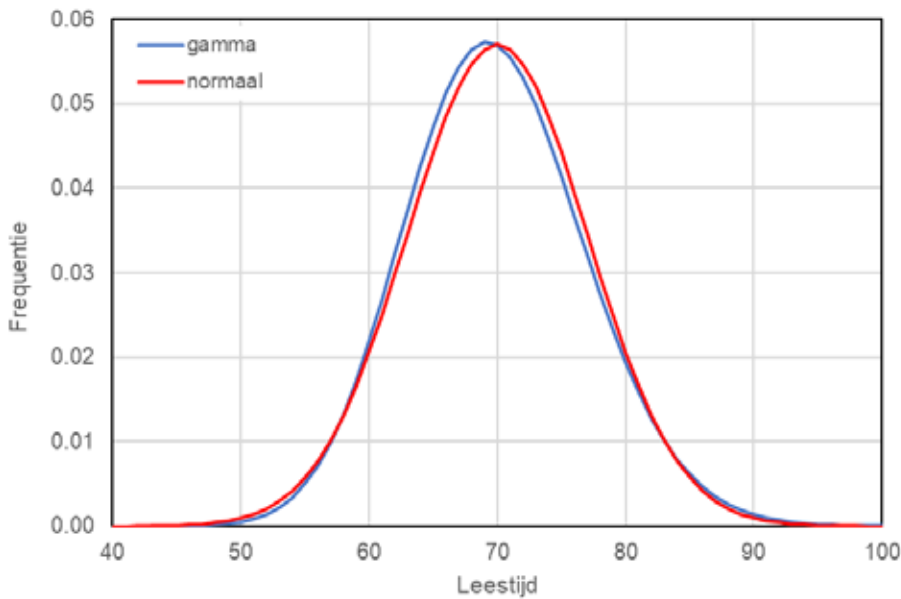
4.3.2 Modelleringsleestijden

Op ingenieuze wijze heeft Rasch (1960) aangetoond dat indien het aantal gelezen woorden binnen een vastgestelde tijd Poisson-verdeeld is, de benodigde tijd om een aantal woorden te lezen de gamma-verdeling moet volgen, waarbij een van de parameters van die verdeling gelijk is aan het aantal woorden. Op deze manier zijn twee soorten snelheidstoetsen – met tijdslijm en zonder tijdslijm – op een elegante manier met elkaar in verband gebracht. Omdat de leestijd een continue variabele is, spreken we niet van de kansverdeling maar van de kansdichtheidsverdeling. De formule voor de kansdichtheidsfunctie is

$$(8) \quad f(t) = \frac{\lambda^m}{(m-1)!} t^{m-1} e^{-\lambda t}$$

Waarin m het aantal woorden in de tekst is. In figuur 4.1 wordt deze functie grafisch weergegeven voor $m = 100$ en $\lambda = 1/0.7 = 1.43$. Het gemiddelde van de verdeling is 70 en de standaardafwijking is gelijk aan 7. In dezelfde figuur is ook de normale verdeling met hetzelfde gemiddelde en dezelfde standaardafwijking afgebeeld (de stippellijn).

Figuur 4.1 Een gamma-verdeling en een normale verdeling (gemiddelde = 70; SD = 7)



In figuur 4.1 is te zien dat beide verdelingen erg op elkaar lijken en dat we voor praktische doeleinden eigenlijk evengoed de normale verdeling zouden kunnen gebruiken. De grote overeenkomst tussen de gamma- en de normale verdeling is echter niet algemeen. Voor andere waarden van de parameters hebben beide verdelingen een heel verschillende vorm, zoals we verderop nog zullen zien.

Net zoals bij het eerder beschreven model voor leesfouten veronderstelt Rasch dat de λ -parameter *multiplicatief* is samengesteld. Als leerling v kaart j leest, is de leestijd gamma-verdeeld met parameter

$$(9) \quad \lambda_{vj} = \theta_v \sigma_j$$

Als een leerling meer dan één kaart leest, wordt conditionele onafhankelijkheid verondersteld, zoals te doen gebruikelijk in de meeste psychometrische modellen. De θ - en de σ -parameters kunnen worden geschat met de grootste aannemelijkheidsmethode, waarbij echter niet gegarandeerd is dat de σ -parameters consistent worden geschat. In het gewone Rasch-model is dit niet het geval en dat is duidelijk aangetoond met een mathematisch argument. Of hier consistentie gegarandeerd is, is niet bekend en we moeten de nodige voorzichtigheid betrachten: immers als de tekstparameters niet consistent worden geschat, geldt waarschijnlijk hetzelfde voor de schattingen van de persoonsparameters θ , en op laatstgenoemde schattingen worden beslissingen voor individuele leerlingen gebaseerd. De schattingsmethode waarbij tegelijkertijd tekst- en persoonsparameters worden geschat, wordt in de psychometrische literatuur aangeduid als *joint maximum likelihood* (JML).

Schatting van de individuele vaardigheid θ_v

De schatting voor de persoonsparameter wordt gegeven door

$$(10) \quad \hat{\theta}_v = \frac{M_v}{\sum_j d_{vj} \sigma_j t_{vj}}$$

waarin d_{vj} de designvariabele is die aangeeft of leerling v kaart j wel of niet heeft gelezen (wel = 1; niet = 0) en waarin t_{vj} de benodigde leestijd is. De waarde van de σ -parameters is niet gekend en daarvoor vult men de schattingen in.

De grootheid M_v is het totaal aantal woorden dat leerling v heeft gelezen:

$$(11) \quad M_v = \sum_j d_{vj} m_j$$

waarin m_j het aantal woorden is van tekst j . Een gedetailleerde afleiding van de schattingsprocedure is te vinden in Verhelst & Kamphuis (in druk).

De schattingsvergelijking (12) levert echter een onzuivere schatter op. Men kan voor die onzuiverheid corrigeren; de zuivere schatter is gegeven door

$$(12) \quad \tilde{\theta}_v = \hat{\theta}_v \times \frac{M_v - 1}{M_v} = \frac{M_v - 1}{\sum_j d_{vj} \sigma_j t_{vj}}$$

In het geval van de AVI-toetskaarten is deze onzuiverheid niet erg belangrijk, omdat de correctiefactor $(M_v - 1)/M_v$ weinig verschilt van 1: het aantal woorden per kaart varieert van 98 tot 224 en in het normerings- en kalibratieonderzoek hebben de leerlingen meerdere kaarten gelezen. De standaardfout van beide schatters is gegeven door

$$(13) \quad SE(\hat{\theta}_v) = \frac{\hat{\theta}_v}{\sqrt{M_v}} \quad \text{en} \quad SE(\tilde{\theta}_v) = \frac{\tilde{\theta}_v}{\sqrt{M_v - 1}}.$$

Interpretatie van de parameters

Als we de benodigde leestijd noteren bij een toetsafname, moeten we natuurlijk een bepaalde eenheid van tijd aanhouden. Deze eenheid is in principe willekeurig, maar bij de rapportage over de AVI-toetskaarten is de seconde als eenheid van tijd gebruikt. De meeste leerlingen in ons onderzoek hadden een leestijd tussen de 60 en de 180 seconden nodig om de hun voorgelegde kaarten te lezen.

De σ -parameters kunnen het best opgevat worden als een onbenoemd getal, een soort correctiefactor: makkelijke teksten krijgen een grote waarde, moeilijkere teksten een kleinere waarde. Aan formule (8) kunnen we zien dat de λ -parameters niet veranderen als we alle σ -parameters met een positieve constante c vermenigvuldigen en alle θ -parameters door diezelfde constante delen. Daarom moeten we de parameters *normaliseren*. In geval van de AVI-toets hebben we geëist dat het product van alle σ -parameters gelijk is aan 1. Een (reële of denkbeeldige) tekst met een σ -parameter gelijk aan 1 noemen we een *standaardtekst*.

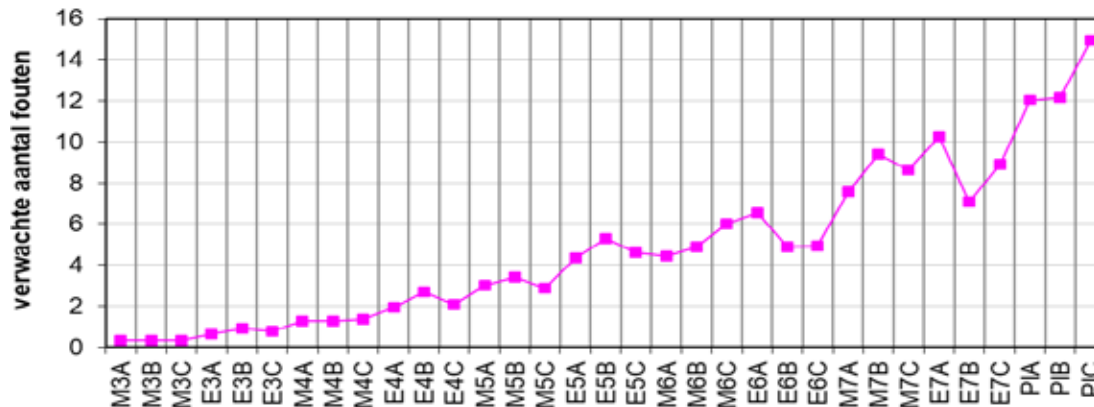
De θ -parameter moet gelezen worden als *aantal woorden per tijdseenheid op een standaardtekst*, in ons geval dus als aantal woorden per seconde. Als men zich daar niet comfortabel bij voelt, is er geen bezwaar tegen een ander tijdseenheid te kiezen: neemt men de minuut als tijdseenheid, dan kan men gewoon de schatting van θ vermenigvuldigen met 60.

4.4 Kalibratie

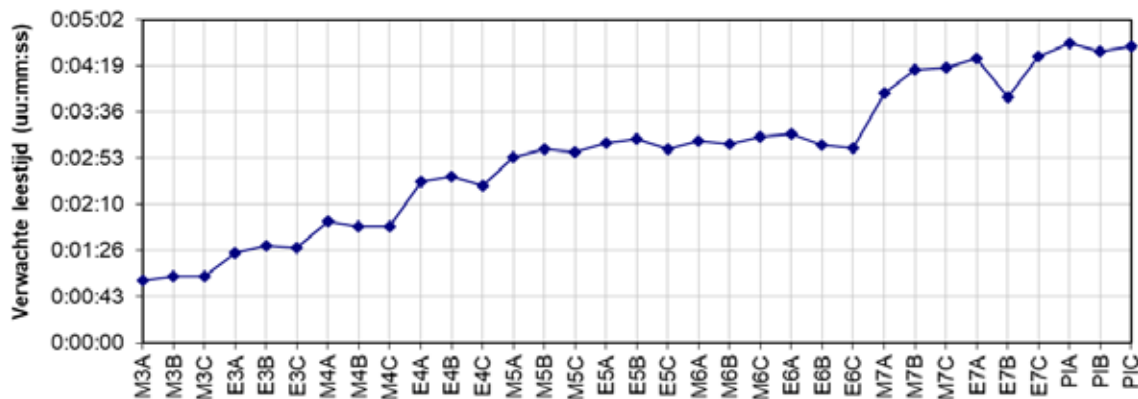
De schatting van de meetmodellen heeft plaatsgevonden met behulp van een speciaal ontwikkeld computerprogramma. De passing van de meetmodellen is vanuit verschillende invalshoeken geëvalueerd. Eerst is gekeken naar de relatie tussen de kaartparameters en de inhoudelijke moeilijkheidsgraad van de kaarten. Vervolgens is onderzocht of de persoonsparameters bij toepassing van de gekozen meetmodellen invariant zijn. Zoals al eerder aangegeven, zijn bij de AVI per kaart twee parameters geschat: één parameter die aangeeft hoe moeilijk het is om de woorden binnen de kaart **correct** te lezen en één parameter die aangeeft hoe gemakkelijk het is om de woorden binnen de kaart **snel** te lezen. In de figuren

4.2 en 4.3 zijn de geschatte parameters (omgezet naar verwachte waarden) voor elke kaart visueel weergegeven.

Figuur 4.2 Verwachte aantal leesfouten in een populatie met leerlingen uit groep 3 tot en met 8



Figuur 4.3 Verwachte leestijd in een populatie met leerlingen uit groep 3 tot en met 8

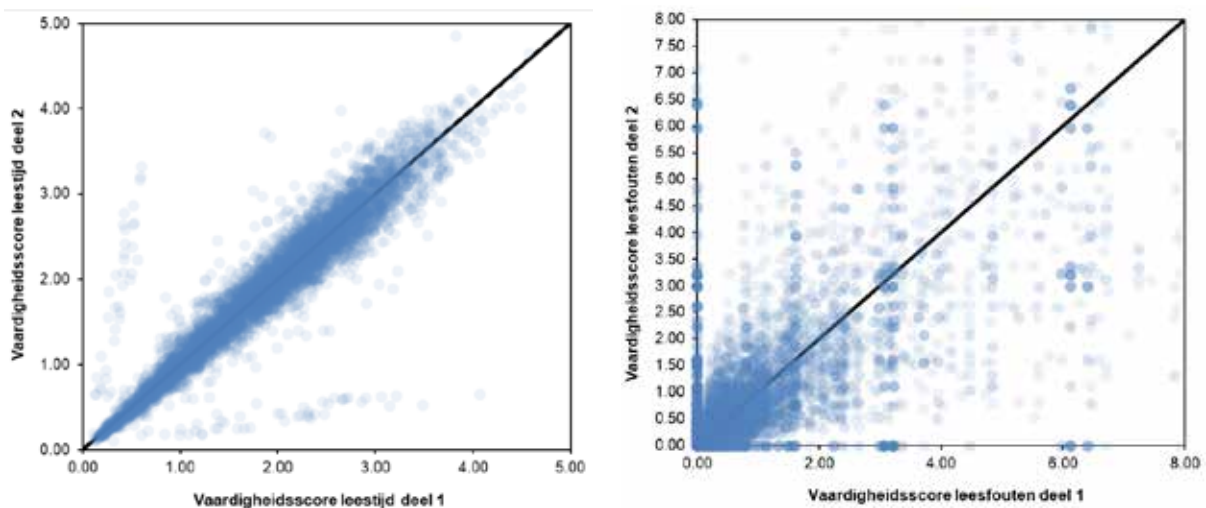


Figuur 4.2 laat zien dat leerlingen bij de kaarten die qua samenstelling passen bij groep 3 en 4 het minst aantal leesfouten maken. Bij de kaarten voor groep 7 en 8 is het verwachte aantal leesfouten aanzienlijk groter. Daarbij is het wel belangrijk om op te merken dat de verwachte waarden gebaseerd zijn op één populatie met leerlingen uit groep 3 tot en met groep 8. Als we een groep 8 kaart voorleggen aan leerlingen uit groep 8 is het verwachte aantal leesfouten lager. Evenzo geldt dat het verwachte aantal fouten op kaart 3 groter is als we uitgaan van een populatie met alleen leerlingen uit groep 3. Het beeld in figuur 4.3 is vergelijkbaar met het beeld in figuur 4.7. Het lezen van de woorden bij de kaarten voor groep 3 neemt gemiddeld gezien het minste tijd in beslag en het lezen van de woorden bij de kaarten voor de hogere groepen neemt meer tijd in beslag. De patronen in figuren 4.2 en 4.3 komt dus goed overeen met het patroon dat we mogen verwachten op basis van de inhoudelijke keuzes die gemaakt zijn bij het samenstellen van de AVI-kaarten.

Tijdens het proefonderzoek hebben alle leerlingen in de steekproef meerdere AVI-kaarten gemaakt. Bij het schatten van de kaartparameters en de vaardigheid (zowel voor leestijd als leesfouten) van leerlingen zijn alle beschikbare gegevens meegenomen. Als het meetmodel geldt moet de vaardigheid van leerlingen echter hetzelfde geschat worden (behoudens meetfout) als we slechts een deel van de afgenomen kaarten meenemen in de schatting. We hebben deze modeleigenschap gecontroleerd door de vaardigheid van leerlingen eerst te schatten op basis van de AVI-kaart die leerlingen als eerste gemaakt hebben en deze

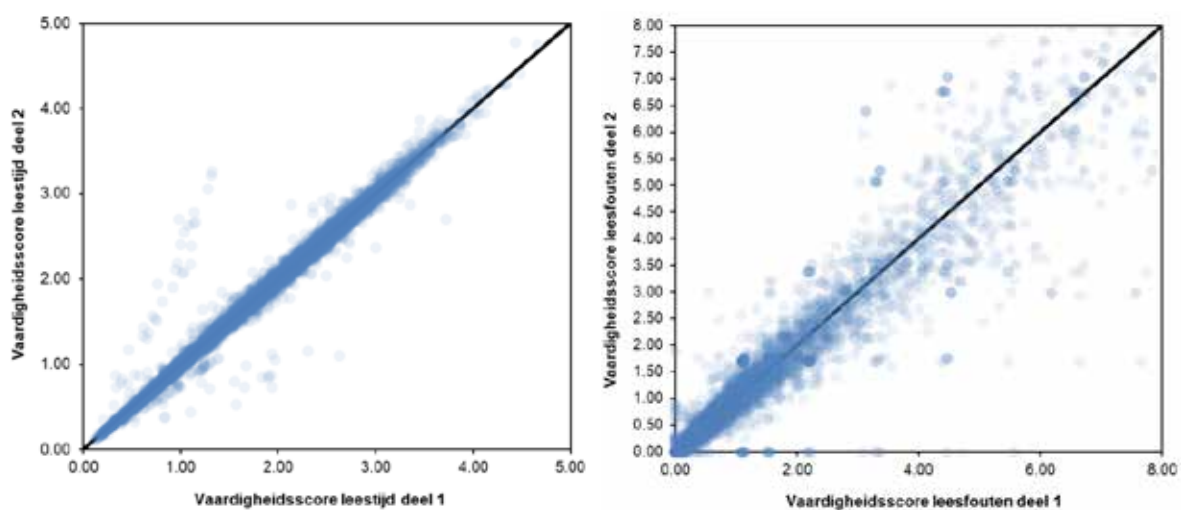
vaardigheidsschatting in een vervolgstap te vergelijken met de vaardigheidsschattingen van diezelfde leerlingen, maar dan gebaseerd op de AVI-kaart die zij als laatste gemaakt hebben. In figuur 4.4 is zowel voor de leesfouten (links) als de leestijden (rechts) een visuele weergave te vinden van de relatie tussen de twee sets met vaardigheidsschattingen. Bij de leestijden is er duidelijk sprake van een lineaire relatie tussen de sets; de grotere verschillen bij de hogere vaardigheidsscores is toe te schrijven aan meetfout (zie ook Hoofdstuk 5). Bij de leesfouten is het beeld minder duidelijk. Daar lijkt de ene set met vaardigheidsschattingen op het eerste gezicht nogal te verschillen van de andere set met vaardigheidsparameters, terwijl we onder aanname van het meetmodel een sterk lineair verband verwachten te vinden.

Figuur 4.4 Invariantie vaardigheidsscores bij afname van één AVI-kaart



Het verband tussen de twee sets met vaardigheidsscores voor leesfouten is nader onderzocht. In de dataset is te zien dat het aantal leesfouten dat leerlingen maken heel erg klein is en ook uit de literatuur is bekend dat juist de leestijd de discriminerende factor is als het gaat om het onderscheid tussen goede en zwakke leerlingen. In dit geval kan de afname van één enkele AVI-kaart mogelijk onvoldoende zijn om de vaardigheid van leerlingen nauwkeurig te schatten. Daarom is de eerder beschreven procedure herhaald, maar nu zijn de vaardigheidsscores van de leerlingen gebaseerd op drie AVI-kaarten in plaats van één. In figuur 4.5 zijn de uitkomsten te vinden. Voor de volledigheid zijn ook de leestijden opnieuw met elkaar vergeleken. Waar we eerder zagen dat er nauwelijks sprake was van een verband tussen de twee sets met vaardigheidsscores voor leesfouten, is in figuur 4.5 te zien dat dit verband er wel is als we meerdere AVI-kaarten afnemen. We kunnen concluderen dat meetfout een overheersende rol speelde in figuur 4.4 en dat er geen reden is om aan te nemen dat de meetmodellen niet bij de data passen als we dit meewegen bij de interpretatie.

Figuur 4.5 Invariantie vaardigheidsscores bij afname van drie AVI-kaarten



4.5 Normering

Zoals in paragraaf 2.3 is vermeld, zijn zowel de samenstelling van de leeskaarten als de normering van de AVI-toets gebaseerd op de vorige uitgave, zodat de prestatie-eisen die bij de verschillende AVI-niveaus worden gesteld in termen van technische leesvaardigheid door de tijd gelijk blijven. Vervolgens zijn ter controle de cumulatieve scoreverdelingen voor leesfouten en leestijden uitgerekend.

Ten aanzien van de vergelijkbaarheid van de prestatie-eisen die bij de oude en de nieuwe AVI-toets worden gesteld, is als belangrijkste uitgangspunt gehanteerd dat de verdeling van leerlingen op basis van frustratie-, instructie- en beheersingscores voor ten minste 90 procent overeen moeten komen. In tabel 4.17 is ter illustratie weergegeven hoe de normering van de nieuwe AVI-toets zich verhoudt tot de normering van de oude AVI-toets bij AVI-niveau M3. We zien dat de classificaties een grote gelijkenis vertonen; ruim 96 procent van de leerlingen wordt op basis van de nieuwe AVI-toets hetzelfde geclassificeerd als op basis van de oude AVI-toets. Voor de classificaties van leerlingen bij de overige AVI-niveaus geldt ook dat 90 procent van de leerlingen of meer hetzelfde wordt geclassificeerd. De tabellen waarin dit is weergegeven zijn te vinden in bijlage 1.

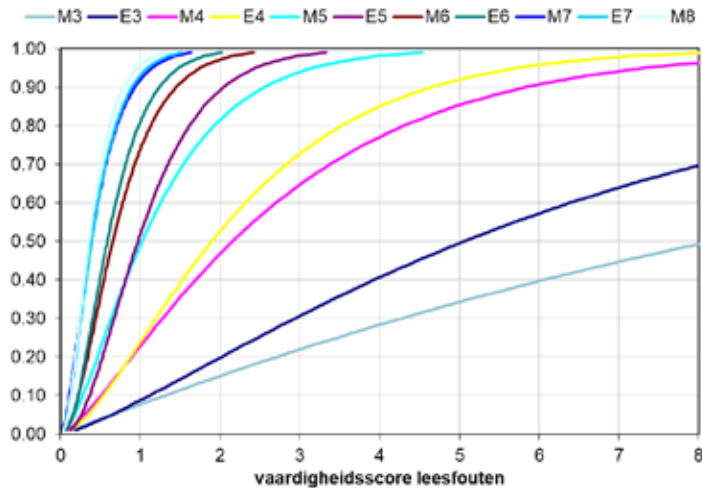
Tabel 4.17 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M3

		AVI-toets, uitgave 2017		
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	97,3	1,6	0,0
	percentage leerlingen met een score op instructieniveau	2,7	89,1	2,5
	percentage leerlingen met een score op beheersingsniveau	0,0	9,4	97,5
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.				96,5

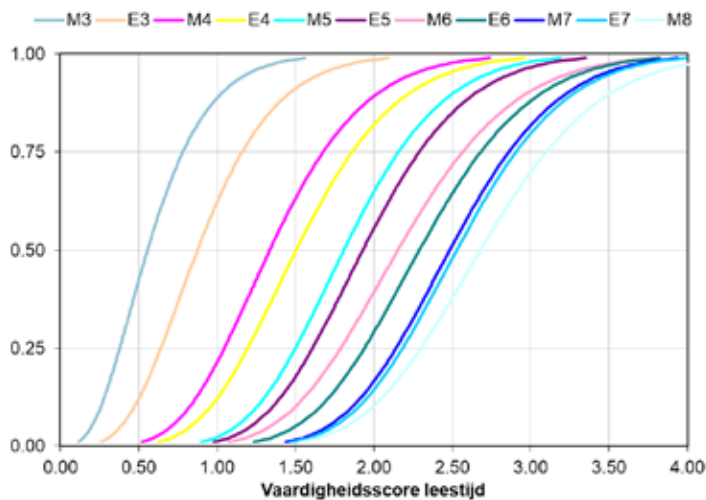
Naast de vergelijkbaarheid van de classificaties volgens de vorige en de nieuwe uitgave van de AVI-toets, is erop gelet dat we over de parallelversies heen precies dezelfde eisen stellen.

De cumulatieve scoreverdelingen voor leesfouten en leestijden zijn respectievelijk te vinden in figuren 4.6 en 4.7.

Figuur 4.6 Cumulatieve vaardigheidsverdelingen leesfouten van medio groep 3 tot medio groep 8



Figuur 4.7 Cumulatieve vaardigheidsverdelingen leestijden van medio groep 3 tot medio groep 8



De scoreverdelingen zijn logisch geordend, zowel als het om de leesfouten gaat als om de leestijden. Bij de leesfouten scoren de groep 8 leerlingen het laagst (let op: bij veel leesfouten is de vaardigheidsscore hoog) en de groep 3 leerlingen, zoals we mogen verwachten, het hoogst. Bij de leestijden is het beeld identiek, maar dan andersom. De groep 8 leerlingen scoren het hoogst en de groep 3 leerlingen het laagst. In figuur 4.6 valt op dat de variantie in het aantal leesfouten sterk afneemt over de tijd; waar er in de lagere leerjaren leerlingen zijn die heel veel fouten maken en leerlingen die heel weinig fouten maken, zijn de verschillen in het aantal leesfouten in de hogere leerjaren aanzienlijk minder groot. Bij de leestijden (figuur 4.7) is de variantie behoorlijk constant over de tijd. Al deze resultaten komen overeen met die van de vorige uitgave.

5 Betrouwbaarheid en meetnauwkeurigheid

5.1 Betrouwbaarheid

Omdat we beschikken over een zuivere schatter van de leessnelheid, kunnen we vrij gemakkelijk een schatting maken van de betrouwbaarheid van die schatter. We brengen even in herinnering dat deze zuivere schatter gegeven is door

$$(1) \quad \tilde{\theta}_v = \frac{M_v - 1}{\sum_j d_{vj} \sigma_j t_{vj}}$$

waarin M_v het totaal aantal woorden voorstelt dat leerling v heeft gelezen, dat wil zeggen: het totaal over alle kaarten die de leerling heeft gelezen. De designvariabele d_{vj} geeft aan of kaart j gelezen is door leerling v ($d_{vj} = 1$) of niet ($d_{vj} = 0$); σ_j is de moeilijkheidsparameter van de kaart en t_{vj} is de tijd die leerling v nodig had om kaart j te lezen. Dat deze schatter zuiver is betekent dat zijn verwachte waarde gelijk is aan de 'werkelijke' parameter θ_v . In een formule uitgedrukt betekent dit dat

$$(2) \quad E(\tilde{\theta}_v | \theta_v) = \theta_v.$$

Omdat de schatter een transformatie is van de geobserveerde leestijden, kunnen we die schatter ook interpreteren als geobserveerde score, en θ_v is de ware score in de zin van de Klassieke Test Theorie. Het verschil tussen de schatting en de ware score is dan de meetfout, en de variantie van de meetfout is niets anders dan het kwadraat van de standaardfout:

$$(3) \quad Var(\tilde{\theta}_v | \theta_v) = SE^2(\tilde{\theta}_v | \theta_v) = \frac{\tilde{\theta}_v^2}{M_v - 1}$$

De geobserveerde waarden in een populatie leerlingen hebben twee bronnen van variabiliteit: de meetfout binnen een persoon maar ook de variabiliteit van de ware scores tussen de personen zorgt voor verschillende observaties. Die twee bronnen van variabiliteit onderhouden de volgende relatie, die algemeen geldig is in de statistiek:

$$(4) \quad Var(\tilde{\theta}) = Var[E(\tilde{\theta} | \theta)] + E[Var(\tilde{\theta} | \theta)].$$

Omdat de schatter zuiver is (zie formule (2)), kunnen we (4) ook schrijven als

$$(5) \quad Var(\tilde{\theta}) = Var(\theta) + E[Var(\tilde{\theta} | \theta)]$$

Wat in woorden betekent: de geobserveerde variantie (linkerlid van (5)) is de som van de variantie van de ware scores (eerste term in het rechterlid) plus de gemiddelde meetfoutvariantie (tweede term in het rechterlid).

De betrouwbaarheid van de geobserveerde scores is per definitie gelijk aan

$$(6) \quad Rel(\tilde{\theta}) = \frac{Var(\theta)}{Var(\tilde{\theta})} = 1 - \frac{E[Var(\tilde{\theta} | \theta)]}{Var(\tilde{\theta})}$$

Teller en noemer in de breuk in het rechterlid van (6) kunnen we consistent schatten als we beschikken over een steekproef. De geobserveerde variantie is gewoon de variantie van de schattingen in de steekproef, en de gemiddelde foutvariantie krijgen we door formule (3) toe te passen voor elke leerling en dan te middelen. Eenzelfde procedure kunnen we toepassen voor de leesfouten omdat in we ook daar beschikken over een zuivere schatting van de vaardigheid. Uiteraard moeten de formules (1) en (3) aangepast worden voor de poisson-likelihood.

Formule (6) hebben we uitgerekend voor de halfjaargroepen medio groep 3, eind groep 3, ..., medio groep 8 voor de afzonderlijke AVI-kaarten. In tabel 5.1 staan de resultaten voor de leestijden, in tabel 5.2 voor de leesfouten.

Tabel 5.1 *Betrouwbaarheid van de $\tilde{\theta}$ -schatting en test-heretest correlatie bij de AVI-toetskaarten voor de leessnelheid*

Halfjaargroep	AVI-kaart	Betrouwbaarheid $\tilde{\theta}$ -schatting	Betrouwbaarheid test-heretest
Medio groep 3	AVI-M3A	0,954	0,973
Medio groep 3	AVI-M3B	0,954	0,972
Medio groep 3	AVI-M3C	0,953	0,972
Eind groep 3	AVI-E3A	0,941	0,96
Eind groep 3	AVI-E3B	0,944	0,963
Eind groep 3	AVI-E3C	0,946	0,962
Medio groep 4	AVI-M4A	0,925	0,94
Medio groep 4	AVI-M4B	0,928	0,942
Medio groep 4	AVI-M4C	0,927	0,942
Eind groep 4	AVI-E4A	0,939	0,95
Eind groep 4	AVI-E4B	0,939	0,95
Eind groep 4	AVI-E4C	0,939	0,949
Medio groep 5	AVI-M5A	0,919	0,929
Medio groep 5	AVI-M5B	0,918	0,928
Medio groep 5	AVI-M5C	0,92	0,931
Eind groep 5	AVI-E5A	0,911	0,921
Eind groep 5	AVI-E5B	0,912	0,923
Eind groep 5	AVI-E5C	0,911	0,921
Medio groep 6	AVI-M6A	0,918	0,928
Medio groep 6	AVI-M6B	0,914	0,925
Medio groep 6	AVI-M6C	0,915	0,926
Eind groep 6	AVI-E6A	0,901	0,911
Eind groep 6	AVI-E6B	0,901	0,91
Eind groep 6	AVI-E6C	0,898	0,909
Medio groep 7	AVI-M7A	0,902	0,91
Medio groep 7	AVI-M7B	0,904	0,912
Medio groep 7	AVI-M7C	0,902	0,911
Eind groep 7	AVI-E7A	0,904	0,911
Eind groep 7	AVI-E7B	0,897	0,905
Eind groep 7	AVI-E7C	0,905	0,913
Medio groep 8	AVI-PIA	0,91	0,918
Medio groep 8	AVI-PIB	0,91	0,918
Medio groep 8	AVI-PIC	0,909	0,917

Tabel 5.2 Betrouwbaarheid van de $\tilde{\theta}$ -schatting en test-her-test correlatie bij de AVI-toetskaarten voor de leesfouten

Halfjaargroep	AVI-kaart	Betrouwbaarheid $\tilde{\theta}$ -schatting	Betrouwbaarheid test-her-test
Medio groep 3	AVI-M3A	0,788	0,789
Medio groep 3	AVI-M3B	0,779	0,781
Medio groep 3	AVI-M3C	0,792	0,791
Eind groep 3	AVI-E3A	0,753	0,754
Eind groep 3	AVI-E3B	0,815	0,815
Eind groep 3	AVI-E3C	0,789	0,791
Medio groep 4	AVI-M4A	0,709	0,708
Medio groep 4	AVI-M4B	0,713	0,714
Medio groep 4	AVI-M4C	0,724	0,723
Eind groep 4	AVI-E4A	0,719	0,719
Eind groep 4	AVI-E4B	0,782	0,784
Eind groep 4	AVI-E4C	0,735	0,734
Medio groep 5	AVI-M5A	0,694	0,691
Medio groep 5	AVI-M5B	0,719	0,72
Medio groep 5	AVI-M5C	0,684	0,683
Eind groep 5	AVI-E5A	0,655	0,653
Eind groep 5	AVI-E5B	0,696	0,699
Eind groep 5	AVI-E5C	0,667	0,669
Medio groep 6	AVI-M6A	0,601	0,606
Medio groep 6	AVI-M6B	0,623	0,623
Medio groep 6	AVI-M6C	0,669	0,671
Eind groep 6	AVI-E6A	0,634	0,636
Eind groep 6	AVI-E6B	0,565	0,565
Eind groep 6	AVI-E6C	0,564	0,568
Medio groep 7	AVI-M7A	0,667	0,663
Medio groep 7	AVI-M7B	0,714	0,718
Medio groep 7	AVI-M7C	0,694	0,693
Eind groep 7	AVI-E7A	0,699	0,697
Eind groep 7	AVI-E7B	0,612	0,616
Eind groep 7	AVI-E7C	0,669	0,669
Medio groep 8	AVI-PIA	0,694	0,693
Medio groep 8	AVI-PIB	0,693	0,695
Medio groep 8	AVI-PIC	0,735	0,736

De betrouwbaarheidsindices in tabel 5.1 voor de leestijden hebben hoge waarden (.90 of hoger). Voor toetsen van het type waar geen zware consequenties voor leerlingen aan verbonden zijn, zoals de AVI, geeft de Cotan aan dat een betrouwbaarheidscoëfficiënt lager dan 0.70 onvoldoende is, een betrouwbaarheidscoëfficiënt tussen 0.70 en 0.80 voldoende en een betrouwbaarheidscoëfficiënt hoger dan 0.80 goed (Evers et al., 2009, p. 33). Op grond van deze criteria is de betrouwbaarheid van de AVI-toets voor de leestijden goed te noemen.

Voor de leesfouten (tabel 5.2) scoren we een voldoende in de lagere groepen (.70 of hoger) en in de hogere groepen een onvoldoende. Merk op dat dit niet zo verwonderlijk is aangezien in de hogere groepen nauwelijks leesfouten worden gemaakt.

Er heeft geen test-hertest onderzoek plaatsgevonden. De afnamecontext van de AVI-toetsen leent zich daar niet goed voor. Het feit dat alle kaarten echter gekalibreerd zijn, maakt het mogelijk een hertest te simuleren. We hebben een dubbele afname gesimuleerd van 100.000 leerlingen. Daarbij hebben we enerzijds de vaardigheidsverdeling van alle leerlingen en anderzijds alle kaartparameters als uitgangspunt genomen. Steeds is een bepaalde vaardigheid aselekt uit de verdeling genomen en zijn twee bij deze vaardigheid horende afnames gesimuleerd. Uiteindelijk is de correlatie tussen deze 100.000 dubbele (virtuele) afnames berekend. Men kan deze simulatie beschouwen als een test-hertestonderzoek onder ideale condities. De tweede toetsafname is immers volledig onafhankelijk van de eerste toetsafname en wordt niet beïnvloed door de kennis die de leerling mogelijk verworven heeft in de eerste toetsafname. Daarnaast is er sprake van invloed van een test-hertest-interval: beide afnames worden gesimuleerd alsof zij op hetzelfde moment plaats zouden vinden. De resultaten zijn weergegeven in tabel 5.1 voor de leestijd en in tabel 5.2 voor de leesfouten. De uitkomsten komen goed overeen met eerder berekende coëfficiënten en leiden dan ook tot dezelfde conclusies met betrekking tot de betrouwbaarheid van de AVI-toets voor de leestijden en leesfouten.

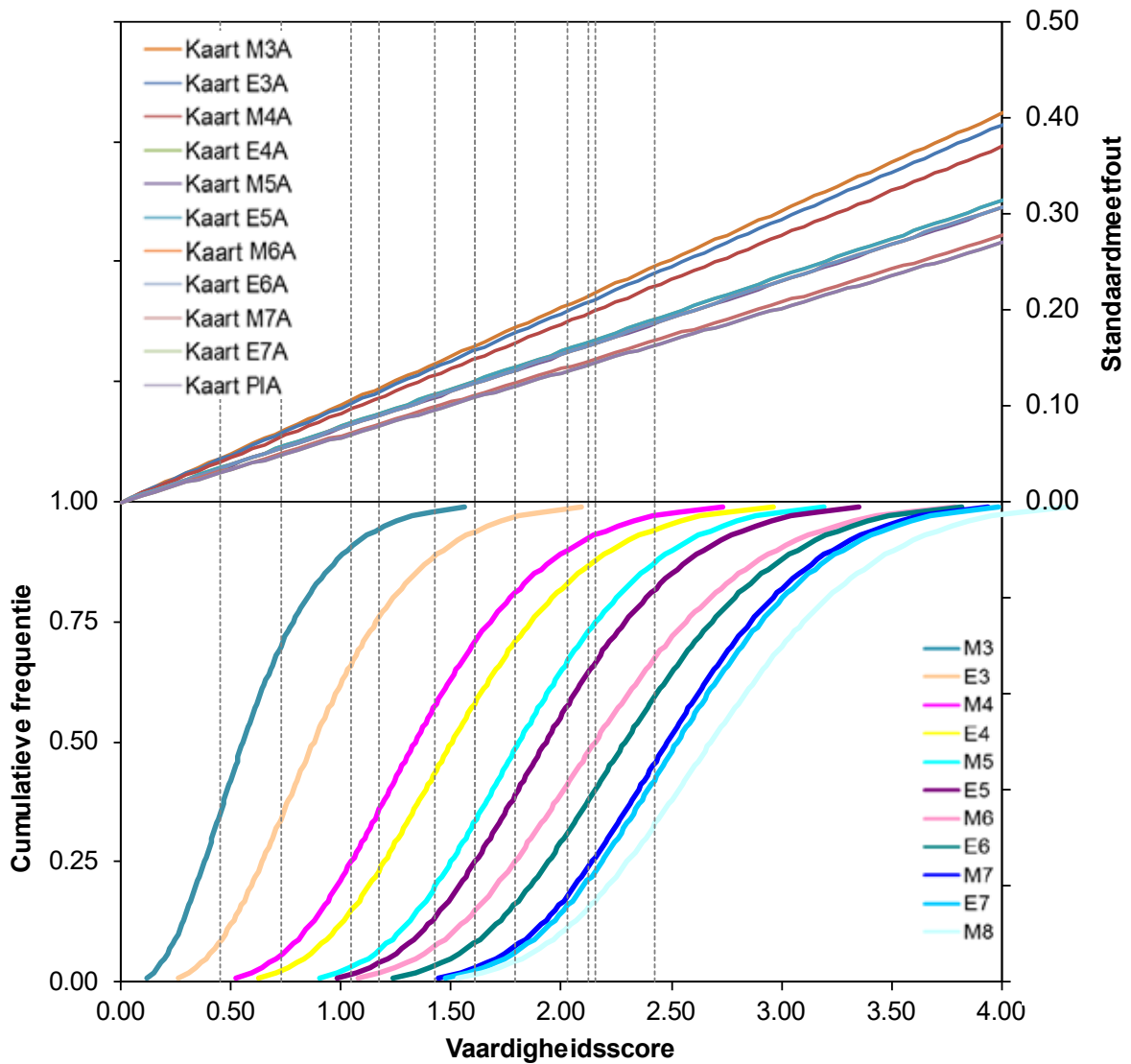
Concluderend kunnen we stellen dat alleen in de hogere groepen de (globale) betrouwbaarheid voor de leesfouten niet voldoende is. De betrouwbaarheid voor de leessnelheid, de vaardigheid die doorgaans de bepalende factor is voor het onderscheiden van sterke en zwakke technisch-lezers, is goed. De secundaire vaardigheid, de leesfouten, fungeert eigenlijk als een soort vangnet voor radende lezers. Dit zijn leerlingen die niet accuraat genoeg lezen voor de snelheid die ze eigenlijk zouden willen hebben. Ze hebben vaak een te snelle leesstijl ontwikkeld en raden de woorden die ze niet snel genoeg kunnen verklanken, waardoor de score op de primaire vaardigheid, de leessnelheid, niet valide is. Omdat het uiteindelijke AVI-niveau bepaald wordt op basis van de scores op zowel leestijden als leesfouten, is het van belang om na te gaan hoe betrouwbaar de combinatie van beide criteria is. Dit wordt aan het einde van de volgende paragraaf uitgewerkt.

5.2 Nauwkeurigheid

Lokale meetnauwkeurigheid (grafisch)

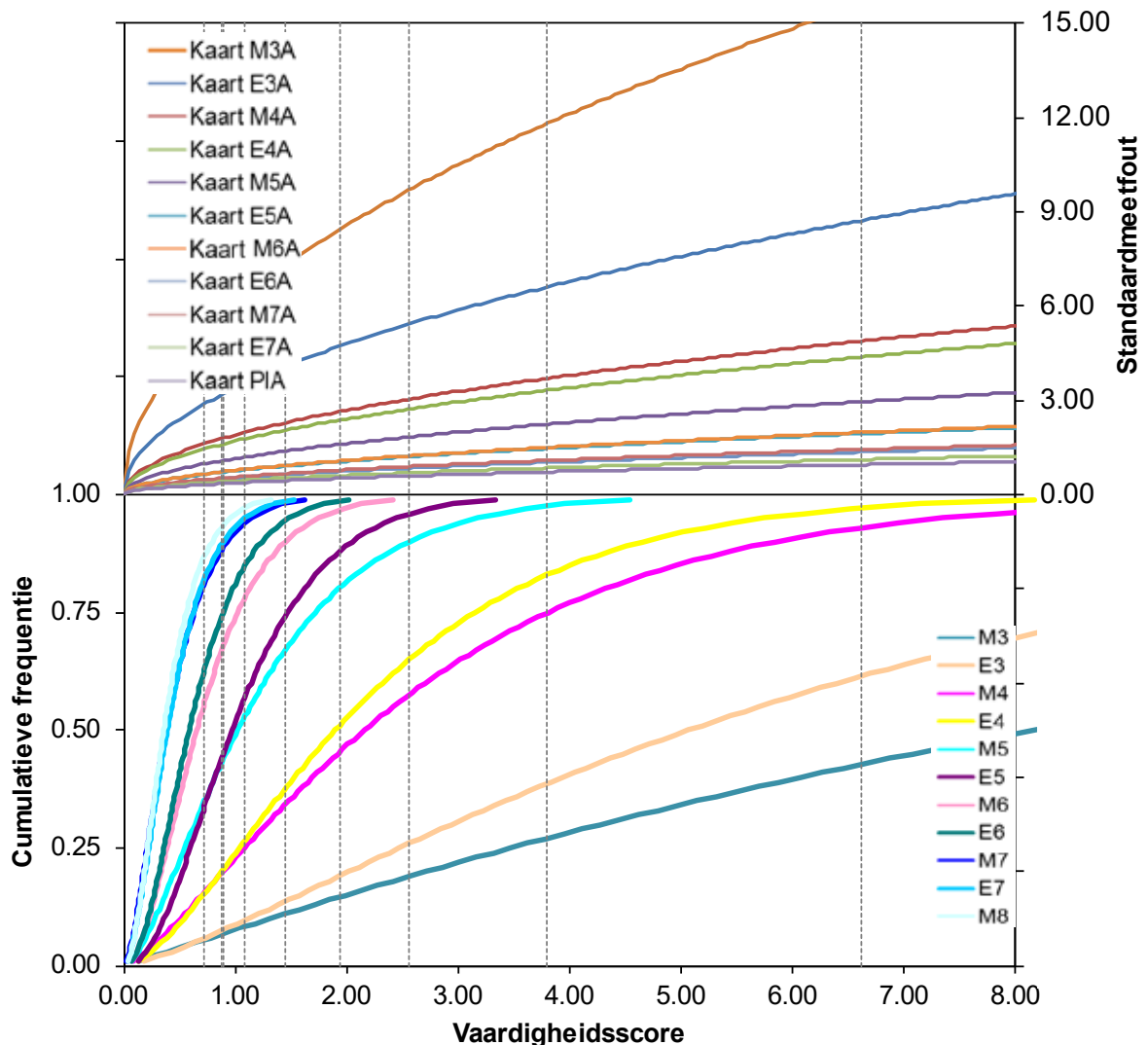
De hiervoor vermelde betrouwbaarheidcoëfficiënten hebben betrekking op de *globale* meetnauwkeurigheid van de toetskaarten, maar geven geen beeld van de *lokale* meetnauwkeurigheid van deze kaarten. Dat doen de figuren 5.1 en 5.2 wel. In figuur 5.1 staat voor ieder van de toetskaarten (versie A) de grootte van de meetfout afgebeeld behorend bij de leestijd. Het volstaat om alleen de meetfouten van de A-kaarten af te beelden, aangezien de A-, de B- en de C-kaarten parallel zijn. In de onderste helft van figuur 5.1 staan de cumulatieve vaardigheidsverdelingen voor de halfjaar momenten. In de figuren 5.1 en 5.2 is met behulp van verticale lijnen afgebeeld wat het vereiste vaardigheidsniveau is om een beheersingsscore te halen. Elke lijn vertegenwoordigt een volgend AVI-niveau. Bij figuur 5.2 ontbreken overigens de laagste twee AVI-niveaus AVI-M3 en AVI-E3, omdat ze buiten het bereik van de grafiek vallen.

Figuur 5.1 Grootte van de meetfout voor de leestijden op de AVI-toetskaarten AVI-M3A tot en met AVI-PlusA



De figuur maakt duidelijk dat de meetfout lineair toeneemt met het toenemen van de vaardigheid. De range van meetfouten op het niveau van vaardigheidsscore 1.5 bijvoorbeeld ligt tussen 0.1 en 0.15 in het geval van de toetskaarten M3 A tot en met M8. Algemeen genomen is de meetfout voor de leestijden op de moeilijkere kaarten dus iets kleiner dan die op de makkelijkere kaarten, maar ook voor laatstgenoemde kaarten blijft de meetfout beperkt. Vaardigheidsscore 1.5 correspondeert met percentiel 50 in de normverdeling op moment E4.

Figuur 5.2 Grootte van de meetfout voor de leesfouten op de AVI-toetskaarten AVI-M3A tot en met AVI-PlusA



De figuur maakt duidelijk dat de meetfout toeneemt met het toenemen van de vaardigheid (in het geval van leesfouten betekent een hoge vaardigheid een grotere geneigdheid tot het maken van leesfouten). De range van meetfouten op het niveau van vaardigheidsscore 3 bijvoorbeeld ligt tussen 0.5 en 10.5 in het geval van de toetskaarten AVI-M3A tot en met AVI-Plus. Algemeen genomen is de meetfout voor de leesfouten op de moeilijker kaarten dus kleiner dan die op de makkelijkere kaarten, maar ook voor laatstgenoemde kaarten blijft de meetfout beperkt. Vaardigheidsscore 5 correspondeert met percentiel 50 in de normverdeling op moment E3.

Lokale meetnauwkeurigheid (tabellarisch)

De betekenis van de meetnauwkeurigheid voor de beslissingen die met de toetsen genomen worden, staan in de onderstaande classificatie-/misclassificatietabellen. Deze tabellen laten het effect van de lokale meetnauwkeurigheid zien. Zo laat tabel 5.3 bijvoorbeeld zien dat 79% van de leerlingen die op toetskaart AVI-E5A met hun geschatte vaardigheidsscore qua leessnelheid in niveaugroep F (frustratie) vallen ook met hun werkelijke vaardigheidsscore in deze groep vallen. Anders gezegd, de kans dat een F-leerling qua leestijden (terecht) als een F-leerling wordt bestempeld is 79%. Verder laat de tabel zien dat ongeveer 21% van de leerlingen in niveaugroep F een vaardigheidsscore voor leestijden heeft die in werkelijkheid in scoregroep I (instructie) valt. Er zijn nauwelijks leerlingen in niveaugroep F die in ware-scoregroep B (beheersing) vallen ten aanzien van de leestijden.

Tabel 5.3 Classificatie-/misclassificatietabel voor de leestijden op de toetskaart AVI-E5A eind jaargroep 5

	Toetskaart AVI-E5A		
	Scoregroep waarin ware score valt		
Niveau groepen F, I en B	F	I	B
F	79,0	20,5	0,5
I	10,5	61,2	28,3
B	0,1	4,1	95,9

Tabel 5.4 is de classificatie-/misclassificatietabel voor de leesfouten voor de toetskaart AVI-E5A. Merk op dat de niveau groepen in tegenstelling tot de leestijden zijn omgedraaid, een leerling met een hoge vaardigheid maakt veel leesfouten. De kans dat een B-leerling op basis van de gemaakte leesfouten (terecht) als een B-leerling wordt bestempeld is op dit afnamemoment ongeveer 94%. Leerlingen in de niveau groep I en F worden minder goed geclassificeerd, dit zijn echter geen grote aantallen. Deze en ook de overige classificatie-/misclassificatietabellen staan in bijlage 2 voor de leestijden en in bijlage 3 voor de leesfouten.

Tabel 5.4 Classificatie-/misclassificatietabel voor de leesfouten op de toetskaart AVI-E5A eind jaargroep 5

	Toetskaart AVI-E5A		
	Scoregroep waarin ware score valt		
Niveau groepen B, I en F	B	I	F
B	94,4	5,6	0,0
I	41,8	56,2	2,0
F	2,2	65,5	32,3

In de onderzoeksliteratuur is weinig geschreven over de beoordeling van betrouwbaarheidstabellen. Wanneer een betrouwbaarheidstabel als goed of voldoende kan worden beschouwd is onduidelijk en wat verwacht mag worden onder ideale omstandigheden is, voor zover ons bekend, niet onderzocht. Daarom worden betrouwbaarheidstabellen vaak samengevat in één of meerdere indices. Wij gebruiken de *plus/minus 1 niveau-index* en de *Marginal Classification Accuracy*. De eerste maat is bedacht door Pilliner (1969). Hij stelt als ambitieniveau dat 95 procent van de leerlingen in een scoregroep in werkelijkheid ook in die scoregroep moet scoren, **of** één scoregroep daarboven **of** één scoregroep daaronder. Bij de door ons gebruikte driedeling (beheersing, instructie of frustatie) is deze index niet zo zinvol. De tweede maat is zinvoller en wordt op verschillende plekken in de literatuur beschreven. De maat laat zien hoe vaak de classificatie op basis van de geschatte vaardigheidsscore berekend aan de hand van geobserveerde waarden, percentueel gezien overeenstemt met de classificatie op basis van de gesimuleerde vaardigheidsscore. De berekening van de marginal classification accuracy (MCA) is als volgt: de aantallen goed geclassificeerde leerlingen in de classificatie-/misclassificatietabel worden gesommeerd (diagonaal cellen in de tabel) en gedeeld door het totaal aantal leerlingen in de tabel. Deze proportie goed geclassificeerde leerlingen wordt vervolgens vermenigvuldigd met 100. Het resultaat is het percentage

goed geclassificeerde leerlingen. Bij een ideale, maar in de praktijk niet te realiseren, toetsafname lijkt de *Marginal Classification Accuracy* rond 75% - 80% uit te komen. In de praktijk liggen de waarden vaak tussen 60% en 70%.

De marginal classification accuracy indices zijn te vinden in tabel 5.5. Zowel voor de leestijden als de leesfouten zijn deze indices voor alle AVI-toetskaarten uitzonderlijk hoog. Kortom: leerlingen worden met de relevante kaarten voor de halfjaargroepen goed geclassificeerd.

Tabel 5.5 *Marginal classification accuracy bij de AVI-toetskaarten voor de leessnelheid en leesfouten*

Halfjaargroep	AVI-kaart	Marginal classification accuracy	
		leessnelheid	leesfouten
Medio groep 3	AVI-M3A	91,1	93,1
Medio groep 3	AVI-M3B	91,1	92,1
Medio groep 3	AVI-M3C	91	92,8
Eind groep 3	AVI-E3A	89,8	88,3
Eind groep 3	AVI-E3B	90,1	89,5
Eind groep 3	AVI-E3C	90,1	89,9
Medio groep 4	AVI-M4A	90,6	92,8
Medio groep 4	AVI-M4B	91	92,8
Medio groep 4	AVI-M4C	90,8	93,6
Eind groep 4	AVI-E4A	92	86,8
Eind groep 4	AVI-E4B	92,1	88,4
Eind groep 4	AVI-E4C	92	88
Medio groep 5	AVI-M5A	91,1	92,1
Medio groep 5	AVI-M5B	91,1	92,2
Medio groep 5	AVI-M5C	91,4	90,3
Eind groep 5	AVI-E5A	88,9	88,3
Eind groep 5	AVI-E5B	89	89,8
Eind groep 5	AVI-E5C	89	89,4
Medio groep 6	AVI-M6A	88,9	88
Medio groep 6	AVI-M6B	88,7	89,9
Medio groep 6	AVI-M6C	88,9	91,2
Eind groep 6	AVI-E6A	86	86,9
Eind groep 6	AVI-E6B	85,8	84,3
Eind groep 6	AVI-E6C	85,9	84,2
Medio groep 7	AVI-M7A	88	90,4
Medio groep 7	AVI-M7B	88	90
Medio groep 7	AVI-M7C	88	91,1
Eind groep 7	AVI-E7A	89	90,9
Eind groep 7	AVI-E7B	90,9	88,9
Eind groep 7	AVI-E7C	88,8	90,9
Medio groep 8	AVI-PlusA	86,8	89,1
Medio groep 8	AVI-PlusB	87,3	88,9
Medio groep 8	AVI-PlusC	86,6	89,6

Tot zover zijn de betrouwbaarheden van de leestijden en leesfouten afzonderlijk behandeld. Het uiteindelijk te bepalen AVI-niveau is echter een combinatie van de behaalde score voor leestijd én leesfouten. In tabel 5.6 is weergegeven hoe het uiteindelijke AVI-niveau tot stand komt. Als een leerling Beheersing (B) scoort voor zowel leestijd als voor leesfouten, dan is het uiteindelijke AVI-niveau ook Beheersing. Bij de

combinatie van leestijd en leesfouten is het laagste niveau leidinggevend. Zo leidt een score op Instructieniveau (I) voor leestijd en Frustratieniveau (F) voor leesfouten uiteindelijk tot een Frustratieniveau voor de betreffende kaart.

Tabel 5.6 Uiteindelijke AVI-score voor de combinatie van leestijden en leesfouten op een toetskaart

Leestijden ↓	Leesfouten		
	B	I	F
B	B	I	F
I	I	I	F
F	F	F	F

Om na te gaan wat de betrouwbaarheid is van het uiteindelijk toegekende AVI-niveau, is in tabel 5.7 een classificatie-/misclassificatietabel voor de combinatie van leestijden én leesfouten weergegeven voor één van de leeskaarten met een iets minder hoge betrouwbaarheid, namelijk AVI-E6B. De leerlingen eind groep 6 in de steekproef vormen hiervoor de basis (n=744). Hun vaardigheidsscores voor leestijd en leesfouten is als uitgangspunt (ware score) genomen. Merk op dat deze vaardigheidsscore gebaseerd is op de geobserveerde resultaten van drie kaartafnames (zie paragraaf 4.1), waardoor de vaardigheidsscore nauwkeurig geschat kan worden. Vervolgens is het uiteindelijke AVI-niveau (Beheersing, Instructie of Frustratie) bepaald voor kaart AVI-E6B. De afhankelijkheid tussen leessnelheid en fouten blijft hiermee behouden.

Met behulp van de ware vaardigheidsscores en de kaartparameters is vervolgens een leestijd (via Gamma-likelihood) en een aantal fouten (via Poisson-likelihood) gegenereerd voor alle leerlingen in de steekproef. De gesampelde tijden en fouten zijn omgezet naar nieuwe vaardigheidsscores, en vervolgens naar beheersings-, instructie- of frustratieniveaus op de afzonderlijke vaardigheden en tenslotte naar een uiteindelijk AVI-niveau. Hiermee is als het ware een tweede afname gesimuleerd. Ten slotte zijn classificatie-/misclassificatietabellen gemaakt om na te gaan hoe het ware, uiteindelijk behaalde AVI-niveau in termen van Beheersing, Instructie en Frustratie zich verhoudt tot het uiteindelijk AVI-niveau uit de simulatie.

De resultaten staan in tabel 5.7. Het totale percentage leerlingen dat vergelijkbaar wordt geclassificeerd (83,5) is uitzonderlijk hoog, zeker als we dit getal vergelijken met vermelde MCA's in tabel 5.5 (85,8 voor leestijden en 84,3 voor leesfouten). De classificatie-/misclassificatietabellen voor de overige kaarten zijn te vinden in bijlage 4. Uit deze bijlage komt naar voren dat de leerlingen op basis van de combinatie van de vaardigheden leestijd en leesfouten voor alle kaarten goed worden geclassificeerd, in ieder geval op beheersings- en frustratieniveau. Er moet een kleine uitzondering worden gemaakt voor het instructie-niveau. Voor de niveaus AVI-E3 én voor AVI-M6 tot en met AVI-E7 geldt dat leerlingen soms wat minder nauwkeurig worden geclassificeerd, omdat het percentage leerlingen dat terecht het instructieniveau krijgt toegekend daar soms lager is dan 50 procent. Voor de overige AVI-niveaus geldt dat de leerlingen op instructieniveau wel goed worden geclassificeerd. Het feit dat het totale percentage leerlingen dat in de simulatie vergelijkbaar wordt geclassificeerd zo hoog is, is een belangrijke vaststelling, omdat leerlingen in de praktijk ook altijd op basis van beide vaardigheden een AVI-niveau krijgen toegekend. Het uiteindelijke AVI-niveau (berekend op basis van de scores voor leestijd én leesfouten) vormt overigens ook het uitgangspunt voor analyses ten behoeve van tabel 4.17 (classificatie van leerlingen bij de oude en nieuwe AVI-kaarten) en ten behoeve van tabel 6.1 (vereiste technische leesvaardigheid voor een beheersingscore per leeskaart).

Tabel 5.7 Classificatie-/misclassificatietabel voor het uiteindelijk AVI-niveau (gebaseerd op de combinatie van leestijden en leesfouten) voor leeskaart AVI-E6B, eind jaargroep 6

	Scoregroep waarin ware score valt			
	B	I	F	
Niveau groepen B, I en F (sample) ↓				
B	93,9	6,1	0,0	
I	35,7	54,5	9,8	
F	2,8	18,1	79,2	
Percentage leerlingen dat vergelijkbaar wordt geclassificeerd				83,5

6 Validiteit

Er kunnen verschillende soorten validiteit worden onderscheiden: inhoudsvaliditeit (content validity), begripsvaliditeit (construct validity) en criteriumvaliditeit. De inhoudsvaliditeit heeft betrekking op de vraag in hoeverre de opgaven in een toets een representatieve steekproef vormen van de te toetsen kennis en/of vaardigheid. Dit wordt behandeld in paragraaf 6.1. De begripsvaliditeit heeft betrekking op de vraag in hoeverre de toets het door de constructeur beoogde kenmerk van de leerling (onderliggende trek, vaardigheid) meet. Het antwoord op deze vraag komt aan de orde in paragraaf 6.2. Om een oordeel te kunnen geven over de begripsvaliditeit, moet men ook over informatie beschikken met betrekking tot soortgenotenonderzoek (convergente en divergente validiteit en prestaties van verschillende subgroepen). Over deze aspecten wordt achtereenvolgens gerapporteerd in de paragrafen 6.3 en 6.4. De criteriumvaliditeit betreft de voorspellende waarde van de toets. Omdat de AVI-toets echter niet bedoeld is voor voorspellend gebruik, is criteriumvaliditeit niet van toepassing.

6.1 Inhoudsvaliditeit

Beoogd is de AVI-toets onafhankelijk samen te stellen van de verschillende onderwijsmethoden voor technisch lezen. De teksten zijn speciaal voor de toets geschreven door professionele kinderboekenauteurs die zich voor elk AVI-niveau hebben gebaseerd op een set tekstkenmerken. Daarna zijn de kaarten aangepast na opmerkingen van Vlaamse en Nederlandse vakexperts, leerkrachten en na analyse van de try-out en de proeftoetsresultaten. Voor de definitieve leeskaarten geldt dat ze, ook na de herzieningen, voldeden aan de vooraf vastgestelde tekstkenmerken. Dit alles vormt een degelijke basis voor de inhoudsvaliditeit van de AVI-toets.

6.2 Begripsvaliditeit

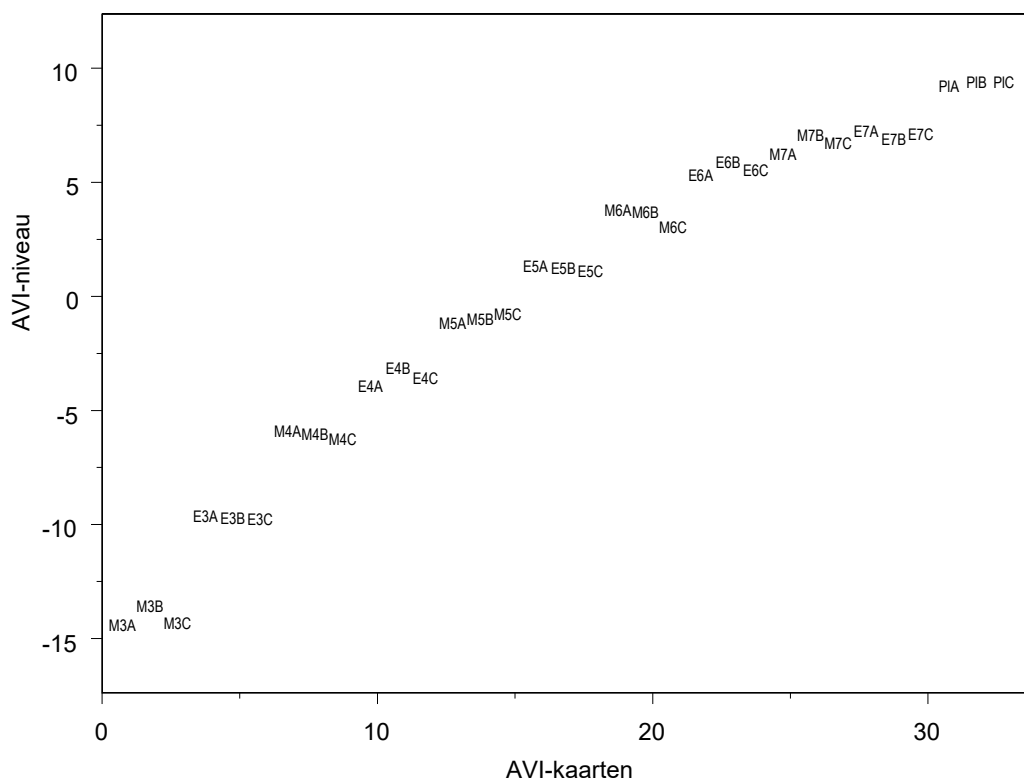
In paragraaf 2.1 werd al beschreven dat technische leesvaardigheid bij voorkeur wordt gedefinieerd in termen van meetnauwkeurigheid én snelheid en dat er daarom voor is gekozen om de AVI-toets onder een zekere tijdsdruk af te nemen. In paragraaf 3.2.1 worden de overige aspecten die kenmerkend zijn voor de opzet van de AVI-toets verantwoord. In de eerste plaats gaat het daarbij om de keuze voor teksten in plaats van geïsoleerde woorden, om de keuze voor teksten van uiteenlopende moeilijkheidsgraad in plaats van teksten die allemaal even moeilijk zijn en ten slotte om de keuze voor hardop lezen.

In hoofdstuk 4 hebben we laten zien dat de prestaties op de leeskaarten aansluiten bij de niveauontwikkeling die leerlingen doormaken van groep 3 tot en met groep 8. In combinatie met de manier waarop het construct Technisch Lezen bij de AVI-toets is geoperationaliseerd, vormt dit een degelijke basis voor de begripsvaliditeit van de AVI-toets.

Structuur

Een van de aspecten die kenmerkend is voor de opzet van de AVI-toets, is de oplopende technische moeilijkheidsgraad van de leeskaarten. Dit wordt enerzijds gewaarborgd doordat de kaarten aansluiten bij vooraf geformuleerde tekstkenmerken. Figuur 6.1 laat zien dat de normering van de leeskaarten in termen van tijd én fouten hierbij aansluit. In de figuur staan op de horizontale as de leeskaarten weergegeven van laag naar hoog. De hoeveelheid technische leesvaardigheid die nodig is om een kaart te lezen op beheersingsniveau is weergegeven op de verticale as. De figuur laat zien dat er steeds meer technische leesvaardigheid vereist is om een leeskaart van een bepaald AVI-niveau te beheersen en dat de toename, geheel volgens verwachting, afvlakt bij de hogere AVI-niveaus.

Figuur 6.1 Vereiste technische leesvaardigheid (beheersingsscore) per leeskaart



Ten slotte is berekend hoeveel procent van de leerlingen een beheersingsscore op een AVI-leeskaart haalt, terwijl de kaart op het onderliggende niveau niet wordt beheerst. Gezien de betrouwbaarheid van de leeskaarten (zie hoofdstuk 5), zou dit percentage beperkt moeten zijn. De resultaten zijn weergegeven in tabel 6.1. Uit de tabel komt naar voren dat het percentage leerlingen dat een beheersingsscore op een AVI-leeskaart haalt, terwijl de kaart op het onderliggende niveau niet wordt beheerst, voor de meeste AVI-niveaus lager is dan vijf procent. Dit is, gezien de betrouwbaarheid van de leeskaarten en het feit dat toename in leesvaardigheid afvlakt, bij de hogere AVI-niveaus, volgens verwachting.

Tabel 6.1 Percentage leerlingen dat een beheersingsscore haalt terwijl de kaart op het onderliggende niveau niet wordt beheerst

AVI-beheersingsniveau	Percentage leerlingen dat een kaart hoger beheerst
AVI-M3	0
AVI-E3	0,8
AVI-M4	3,6
AVI-E4	2,7
AVI-M5	2,8
AVI-E5	2,6
AVI-M6	5,8
AVI-E6	7,5
AVI-M7	9,1
AVI-E7	4,2
AVI-Plus	4,4

6.3 Convergente en divergente validiteit

De mate waarin de AVI-toets gerelateerd is aan soortgenootinstrumenten als de DMT (Cito, 2017) en de EMT (Een-Minuu-Test) en niet-soortgenootinstrumenten is een belangrijke aanwijzing voor de begripsvaliditeit van de toets. In de tabellen 6.2 tot en met 6.7 worden correlaties gerapporteerd tussen scores op de AVI-toets (uitgesplitst naar tijd en fouten¹), de DMT, de EMT en de toetsen Begrijpend lezen en Woordenschat die in het kader van het Cito Volgsysteem primair en speciaal onderwijs zijn ontwikkeld. Alle bovengenoemde toetsen zijn positief beoordeeld door de Cotan of de Expertgroep PO. Zoals in hoofdstuk 4 is beschreven, maakten de EMT en de DMT deel uit van het normeringsonderzoek. De scores van de deelnemende leerlingen op de toetsen Begrijpend lezen en Woordenschat zijn in een later stadium, middels dataretour via het Computerprogramma LOVS, opgevraagd. In de tabellen zijn de cellen leeg gelaten waarbij de correlaties betrekking hadden op minder dan 50 leerlingen. Tabel 6.2 tot en met 6.7 laten zien dat de scores op AVI-toets voor leestijd, de DMT en de EMT volgens verwachting onderling sterk samenhangen. Het aantal gemaakte fouten bij de AVI-toets hangt negatief samen met de DMT. Blijkbaar maken de leerlingen die goed scoren op de DMT minder fouten bij de AVI-toets en andersom, wat eveneens in de lijn der verwachting ligt, omdat leerlingen over het algemeen weinig fouten maken bij de DMT en er bovendien meestal weinig verschil is tussen het aantal leesfouten dat hoog en laag scorende leerlingen op de DMT maken. Dit kan als een bewijs voor convergente validiteit worden opgevat.

De samenhang tussen de AVI-leestijd en de toetsen Begrijpend lezen en Woordenschat is over het algemeen matig, omdat bij deze laatste twee toetsen deels dezelfde en deels andere vaardigheden dan technisch lezen een rol spelen. Dit kan worden aangevoerd als bewijs voor divergente validiteit. Het aantal gemaakte fouten op de AVI-toets hangt negatief samen met de scores op de toetsen Begrijpend lezen en Woordenschat. Net zoals bij de DMT geldt dat leerlingen die blijkbaar goed scoren op de toetsen Begrijpend lezen en Woordenschat minder fouten maken bij de AVI-toets en andersom, wat eveneens in de lijn der verwachting ligt.

Tabel 6.2 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 3 (n = 843)

	DMT m.	DMT e.	AVI t. m.	AVI t. e.	AVI f. m.	AVI f. e.	EMT m.	BL e.	WS m.	WS e.
DMT medio	1,000									
DMT eind	0,847	1,000								
AVI-tijd medio	0,848	0,901	1,000							
AVI-tijd eind	0,747	0,930	0,917	1,000						
AVI-fouten medio	-0,479	-0,539	-0,562	-0,531	1,000					
AVI-fouten eind	-0,509	-0,657	-0,593	-0,627	0,869	1,000				
EMT medio	0,786	0,711	0,743	0,674	-0,339	-0,462	1,000			
Begr lezen eind	0,388	0,426	0,408	0,459	-0,460	-0,635		1,000		
Woordenschat medio	0,021	-0,055	-0,057	-0,060	-0,093	-0,209		0,308	1,000	
Woordenschat eind	0,157	0,086	0,025	0,096	-0,170	-0,231		0,398	0,932	1,000

¹ De correlaties van AVI-fouten met AVI-tijd en met alle andere toetsen zijn negatief doordat sprake is van een omgekeerde norm: hoe hoger de 'score' (= het aantal fouten), hoe lager de vaardigheid. Bij alle andere toetsen geldt: hoe hoger de score, hoe hoger de vaardigheid.

Tabel 6.3 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 4 (n = 1003)

	DMT m.	DMT e.	AVI t. m.	AVI t. e.	AVI f. m.	AVI f. e.	EMT m.	BL m.	BL e.	WS m.	WS e.
DMT medio	1,000										
DMT eind	0,965	1,000									
AVI-tijd medio	0,875	0,871	1,000								
AVI-tijd eind	0,871	0,920	0,963	1,000							
AVI-fouten medio	-0,667	-0,651	-0,506	-0,586	1,000						
AVI-fouten eind	-0,662	-0,685	-0,509	-0,590	0,959	1,000					
EMT medio	0,953	0,981	0,898	0,933	-0,636	-0,663	1,000				
Begr lezen medio	0,505	0,522	0,538	0,588	-0,494	-0,446		1,000			
Begr lezen eind	0,444	0,511	0,497	0,530	-0,336	-0,398		0,789	1,000		
Woordenschat medio	0,266	0,349	0,431	0,536	-0,302	-0,194		0,599	0,581	1,000	
Woordenschat eind	0,378	0,489	0,452	0,556	-0,367	-0,361		0,666	0,643	0,812	1,000

Tabel 6.4 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 5 (n = 919)

	DMT m.	DMT e.	AVI t. m.	AVI t. e.	AVI f. m.	AVI f. e.	EMT m.	BL m.	BL e.	WS m.
DMT medio	1,000									
DMT eind	0,980	1,000								
AVI-tijd medio	0,904	0,889	1,000							
AVI-tijd eind	0,912	0,934	0,928	1,000						
AVI-fouten medio	-0,701	-0,696	-0,601	-0,614	1,000					
AVI-fouten eind	-0,698	-0,704	-0,580	-0,603	0,991	1,000				
EMT medio	0,922	0,912	0,821	0,764	-0,541	-0,556	1,000			
Begr lezen medio	0,497	0,489	0,658	0,556	-0,491	-0,450		1,000		
Begr lezen eind	0,380	0,352	0,516	0,374	-0,438	-0,445		0,852	1,000	
Woordenschat medio	0,327	0,380	0,483	0,426	-0,420	-0,457		0,744	0,840	1,000

Tabel 6.5 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 6 (n = 935)

	DMT m.	DMT e.	AVI t. m.	AVI t. e.	AVI f. m.	AVI f. e.	EMT m.	BL m.	WS m.	WS e.
DMT medio	1,000									
DMT eind	0,985	1,000								
AVI-tijd medio	0,856	0,890	1,000							
AVI-tijd eind	0,867	0,927	0,940	1,000						
AVI-fouten medio	-0,708	-0,704	-0,593	-0,577	1,000					
AVI-fouten eind	-0,747	-0,745	-0,624	-0,611	0,996	1,000				
EMT medio	0,973	0,936	0,868	0,810	-0,656	-0,694	1,000			
Begr lezen medio	0,482	0,583	0,686	0,712	-0,455	-0,479		1,000		
Woordenschat medio	0,339	0,373	0,502	0,455	-0,376	-0,394		0,630	1,000	
Woordenschat eind	0,257	0,343	0,344	0,423	-0,171	-0,215		0,714	0,815	1,000

Tabel 6.6 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 7 (n = 881)

	DMT m.	DMT e.	AVI t. m.	AVI t. e.	AVI f. m.	AVI f. e.	EMT m.	BL m.	BL e.	WS m.	WS e.
DMT medio	1,000										
DMT eind	0,991	1,000									
AVI-tijd medio	0,807	0,831	1,000								
AVI-tijd eind	0,841	0,870	0,932	1,000							
AVI-fouten medio	-0,600	-0,588	-0,472	-0,485	1,000						
AVI-fouten eind	-0,665	-0,663	-0,571	-0,559	0,986	1,000					
EMT medio	0,974	0,961	0,800	0,844	-0,513	-0,563	1,000				
Begr lezen medio	0,275	0,372	0,610	0,579	-0,276	-0,379		1,000			
Begr lezen eind	0,352	0,411	0,666	0,572	-0,336	-0,435		0,894	1,000		
Woordenschat medio		0,272		0,459		-0,120		0,652		1,000	
Woordenschat eind		0,253		0,454		-0,064		0,570		0,772	1,000

Tabel 6.7 Correlaties tussen de AVI-toets, de DMT en andere toetsen in groep 8 (n = 604)

	DMT m.	AVI t. m.	AVI f. m.	EMT m.	BL m.	WS m.
DMT medio	1,000					
AVI-tijd medio	0,825	1,000				
AVI-fouten medio	-0,640	-0,509	1,000			
EMT medio	0,987	0,862	-0,683	1,000		
Begr lezen medio	0,413	0,578	-0,339		1,000	
Woordenschat medio	0,125	0,325	-0,183		0,656	1,000

6.4 Verschillen tussen relevante subgroepen

Er zijn effectschattingen uitgevoerd over alle normgroepen als geheel voor de variabelen sekse, leerlinggewicht (zie paragraaf 4.2) en wel of geen dyslexie. Bij het maken van effectschattingen is Cohens D gebruikt. De interpretatie van de effectgroottes is volgens Cohen (1988). De resultaten van de effectschattingen zijn opgenomen in tabel 6.8. De tabel laat zien dat voor de variabelen sekse en leerlinggewicht geen betekenisvolle verschillen worden gevonden. Voor de variabele wel of geen dyslexie worden kleine tot matige positieve effecten gevonden ten gunste van leerlingen bij wie geen dyslexie is vastgesteld. Deze bevindingen sluiten aan bij theoretische verwachtingen ten aanzien van de prestaties van de verschillende subgroepen (Schijf, 2009).

Tabel 6.8 Resultaten effectschattingen voor sekse, leerlinggewicht en wel of geen dyslexie

	AVI-tijden medio	AVI-tijden eind	AVI-fouten medio	AVI-fouten eind
sekse	-0,05	-0,01	0,01	0,00
leerlinggewicht	0,09	-0,02	0,09	-0,01
wel of geen dyslexie	0,17	0,42	0,23	0,22

7 Samenvatting

In dit samenvattende hoofdstuk geven we kort weer wat in de voorafgaande hoofdstukken is besproken. In hoofdstuk 1 wordt aangegeven dat de AVI-toets uit het Cito Volgsysteem primair en speciaal onderwijs is bedoeld om het vaardigheidsniveau en de ontwikkeling van leerlingen op het gebied van technisch lezen in kaart te brengen.

In hoofdstuk 2 wordt een gedetailleerde beschrijving gegeven van de uitgangspunten van de toetsconstructie. Na elkaar komen aan de orde: de meetpretentie van de AVI-toets, de doelgroep en de gebruiksdoelen. Vervolgens wordt technisch lezen gedefinieerd als een vaardigheid die 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. Daarna wordt uitgelegd hoe leerlingen het technisch lezen onder de knie krijgen en welke aspecten en principes een rol spelen bij technisch lezen.

Hoofdstuk 3 beschrijft de opbouw en structuur van de AVI-toets en gaat in op de afname en de rapportage. Vervolgens wordt een uitwerking van de domeinbeschrijving gegeven waarin wordt uitgelegd hoe de definitie van technisch lezen zoals gegeven in hoofdstuk 2, is geoperationaliseerd. Aansluitend hierop wordt uitgelegd hoe de leeskaarten zijn samengesteld. Ten slotte wordt aangetoond dat de definitieve leeskaarten voldeden de vooraf vastgestelde tekstkenmerken.

In hoofdstuk 4 wordt verslag gedaan van de kalibratie en normering. In de eerste plaats wordt beschreven dat bij de dataverzameling gebruik is gemaakt van onderzoeksdesigns waardoor uitgebalanceerde sets leeskaarten aan leerlingen konden worden voorgelegd. Vervolgens komt de representativiteit van de steekproef aan de orde. De analyses tonen aan dat de steekproeven voor regio en urbanisatiegraad en soms ook voor schooltype afwijken van de landelijke verdelingen. Soms lijkt er een relatie bestaan tussen één van deze variabelen en de prestaties van leerlingen, maar een duidelijk patroon over de afnamemomenten ontbreekt. Omdat de normering van de AVI-toets gebaseerd is op de vorige uitgave (en niet op de populatieverdeling van de normeringssteekproef), is het ontbreken van volledige representativiteit echter geen probleem. Daarna volgt een beschrijving van het gehanteerde meetmodel voor leesfouten en leestijden. Ook wordt gerapporteerd hoe op basis van deze meetmodellen de kalibratie heeft plaatsgevonden. Ten slotte is weergegeven hoe de normering tot stand is gekomen en wordt aangetoond dat de AVI-toets gedurende de gehele basisschoolperiode goed in staat is om groepen leerlingen met een verschillend leesvaardigheidsniveau van elkaar te onderscheiden.

In hoofdstuk 5 staat de betrouwbaarheid van de toets centraal. Eerst wordt uitgelegd welke procedure is gevolgd om de betrouwbaarheid van de schatter voor leestijden en leessnelheid te berekenen. Vervolgens worden betrouwbaarheidsindices voor leestijden en leesfouten gepresenteerd. Deze indices zijn met 0,91 en hoger voor de leestijden zeer goed te noemen in relatie tot het beoogde gebruik van de toets. Verder zijn in dit hoofdstuk figuren opgenomen die inzicht geven in de lokale betrouwbaarheid. Hieruit komt naar voren dat de meetfout voor leestijden het kleinst lijkt te zijn op de moeilijker kaarten, maar dat de leesfout ook op de makkelijker kaarten beperkt blijft. Vervolgens wordt de marginal classification accuracy gepresenteerd voor zowel leestijden als leesfouten. Deze waarden liggen voor alle kaarten boven de 80, wat hoog te noemen is. Ten slotte wordt de betekenis van de meetnauwkeurigheid voor de beslissingen die met de toetsen genomen worden, in de classificatie-/misclassificatietabellen gepresenteerd, eerst voor leestijden en leesfouten apart en daarna voor de combinatie van beide vaardigheden (leestijden en leesfouten). De tabellen waarin beide vaardigheden worden gecombineerd geven het beste beeld van de betrouwbaarheid, omdat bij het bepalen van het AVI-niveau beide vaardigheden een rol spelen. Uit de tabellen komt naar voren dat de leerlingen met de relevante kaarten voor de halfjaargroepen goed geclassificeerd worden op beheersingsniveau en frustratieniveau en bij de meeste kaarten ook op instructieniveau.

In hoofdstuk 6 wordt als eerste de soortgenootvaliditeit gerapporteerd. Als belangrijke schakel in de bewijsvoering wordt de hoge correlatie van de AVI-leestijden met de DMT en de EMT aangevoerd. De negatieve correlatie van het aantal fouten op de AVI-toets en de DMT en EMT draagt hiertoe bij omdat het voor de hand ligt dat leerlingen die veel fouten maken op de AVI-toets minder goed presteren op de DMT en de EMT en andersom. Op basis hiervan is de convergente validiteit van de AVI-toets toets hoog te noemen. De correlaties tussen de AVI-leestijden enerzijds en de toetsen Begrijpend lezen en Woordenschat anderzijds bleken lager dan de correlatie tussen de toetsen technisch lezen onderling. Dit kan als bewijs van divergente validiteit worden opgevat.

Vervolgens wordt aangetoond dat de normering van de AVI-toets in termen van tijd en fouten aansluit bij de toenemende technische leesmoeilijkheid van de kaarten: er is steeds meer technische leesvaardigheid nodig om een AVI-kaart van een hoger AVI-niveau te lezen. Daarnaast laten analyses zien dat het percentage leerlingen dat een beheersingsscore haalt voor een AVI-niveau terwijl het onderliggende niveau niet wordt beheerst, in overeenstemming is met de waarden die gezien de betrouwbaarheid van de toets verwacht mogen worden en met het feit dat de vereiste vaardigheids groei afvlakt bij de hogere AVI-niveaus. Dit kan eveneens worden aangevoerd als bewijs van de validiteit van de AVI-toets. Ten slotte wordt aangetoond dat de prestaties van relevante subgroepen op de AVI-toets aansluiten bij de theoretische verwachtingen ten aanzien van de verschillende subgroepen.

Literatuurlijst

- Allington, R.L. (1984). Oral reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 829-864). New York: Longmans.
- Armbruster, B.B., & Wilkinson, I.A.G. (1991). Silent reading, oral reading, and learning from text. *The Reading Teacher*, 45, 154-155.
- Bosman, A.M.T. (2000). Lezen en spellen volgens het fonologisch coherentiemodel. In H. Kuster en H. Wieberink (red.), *'Ongestoord voort'. Over omgaan met dyslexie in het onderwijs* (pp. 34-47). Ede: Stichting Schoolbegeleidingsdienst De Zuid-Vallei.
- Bosman, A.M.T., and Van Orden, G.C. (2003). Het fonologisch coherentiemodel voor lezen en spellen. *Pedagogische Studiën*, 80, 391-406.
- Boxtel, H. van & B.T. Hemker (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.
- Brus, B.Th. en Voeten, M.J.M. (1973). *Eén-Minuu-Test vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout Testmateriaal BV.
- Cito (2009). *Cito Leerling- en onderwijsvolgsysteem Primair onderwijs. DMT en AVI. Groep 3 tot en met 8*. Arnhem: Cito.
- Cito (2017). *Cito Volgsysteem Primair en speciaal onderwijs. AVI. Groep 3 tot en met 8*. Arnhem: Cito.
- Cito (2017). *Cito Volgsysteem Primair en speciaal onderwijs. DMT. Groep 3 tot en met 8*. Arnhem: Cito.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). New York: Academic Press.
- Coltheart, M. (2006). Dual route and connectionist models of reading: an overview. *London Review of Education*, 4, 5-17.
- Coltheart, M., and Rastle, K. (1994). Serial processing in reading aloud: evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1197-1211.
- Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed processing approaches. *Psychological Review*, 100, 589-608.
- Coltheart, M., Rastle, K., Perry, C., Ziegler, J., and Langdon, R. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.
- Curtis, M.E., and Glaser, R. (1983). Reading theory and the assessment of reading achievement. *Journal of Educational Measurement*, 20, 133-147.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.

- Goldsmith-Phillips, J. (1989). Word and context in reading development: a test of the interactive-compensatory hypothesis. *Journal of Educational Psychology*, *81*, 299-305.
- Harm, M.W., and Seidenberg, M.S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological Review*, *106*, 491-528.
- Harm, M.W., and Seidenberg, M.S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00054.
- Kim (Yoon), Y.H., and Goetz, E.T. (1994). Context effects on word recognition and reading comprehension of poor and good readers: A test of the interactive-compensatory hypothesis. *Reading Research Quarterly*, *29*, 178-188.
- Krom, R., Jongen, I., Verhelst, N., Kamphuis, F., & Kleintjes, F. (2010). *Wetenschappelijke verantwoording DMT en AVI*. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Leu, D.J., Jr., DeGross, L.-J.C., and Simons, H.D. (1986). Predictable texts and interactive-compensatory hypotheses: evaluating individual differences in reading ability, context use, and comprehension. *Journal of Educational Psychology*, *78*, 347-352.
- McLeod, P., Plaut, D.C., and Shallice, T. (2001). Connectionist modelling of word recognition. *Synthese*, *129*, 173-183.
- Nicholson, T. (1991). Do children read words better in context or in lists? A classic study revisited. *Journal of Educational Psychology*, *83*, 444-450.
- Nicholson, T., Lillas, C., and Rzoska, M.A. (1988). Have we been misled by miscues? *The Reading Teacher*, *42*, 6-10.
- Paap, K.R., and Noel, R.W. (1991). Dual-route models of print to sound: still a good horse race. *Psychological Research*, *53*, 13-24.
- Pearson (1999). *Eén-Minuu-Test*. Brus, B.Th. en Voeten, M.J.M. Amsterdam: Pearson Benelux BV.
- Perry, C., Ziegler, J.C., and Zorzi, M. (2007). Nested modeling and strong inference resting in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *27*, 301-333.
- Perry, C., Ziegler, J.C., and Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, *61*, 106-151.
- Perry, C., Ziegler, J.C., and Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cognitive Science*, *37*, 800-828.
- Peterson R.L., Pennington B.F., Olson R.K. (2013). Subtypes of developmental dyslexia: Testing the predictions of the dual-route and connectionist frameworks. *Cognition*, 20-38. (10.1016/j.cognition.2012.08.007.).

- Pritchard, S.C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual-process models with human data. *Journal of Experimental Psychology: Human Perception and Performance*, 38(5), 1268-1288.
- Pilliner, A. (1969). *Estimation of number of grades to be awarded in an examination by consideration of its reliability coefficient*. Edinburgh: The Godfrey Thomson Unit for Educational Research.
- Rapcsak, S.Z., Henry, M.L., Teague, S.L., Carnahan, S.D., and Beeson, P.M. (2007). Do dual-route models accurately predict reading and spelling performance in individuals with acquired alexia and agraphia? *Neuropsychologia*, 45, 2519-2524.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Nielsen & Lydiche.
- Schijf, G.M. (2009). *Lees- en spellingvaardigheden van brugklassers* (proefschrift). Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam.
- Schumm, J.S. & Baldwin, R.S. (1989). Cue system usage in oral and silent reading. *Journal of Reading Behavior*, 21, 141-154.
- Seidenberg, M.S. (2007). Connectionist models of reading. In M.G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 235-250). Oxford: Oxford University Press.
- Seidenberg, M.S., and McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Søvik, N., Arntzen, O., and Samuelstuen, M. (2000). Eye-movement parameters and reading speed. *Reading and Writing: An Interdisciplinary Journal*, 13, 237-255.
- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, Vol. 16, 1*, 32-71.
- Stanovich, K.E., West, R.F., and Freeman, D.J. (1981). A longitudinal study of sentence context effects in second grade children: tests of an interactive-compensatory model. *Journal of Experimental Child Psychology*, 32, 185-199.
- Stanovich, K.E., and Stanovich P.J. (1995). How research might inform the debate about early reading acquisition. *Journal of Research in Reading*, 18, 87-105.
- Staphorsius, G. en Verhelst, N. (1997). Indexering van de leestechiek. *Pedagogische Studiën*, 74, 154-164.
- Stone, G.O., Vanhoy, M., and Van Orden, G.C. (1997). Perception is a two-way street: feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, 36, 337-359.
- Struiksmā, A.J.C., Leij, A. van der en Vieijra, J.P.M. (2009). *Diagnostiek van technisch lezen en aanvankelijk spellen*. Amsterdam: VU Uitgeverij.
- Van Orden, G.C., Johnston, J.C., and Hale, B.L. (1988). Word identification in reading precedes from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 371-386.

Van Orden, G.C., Pennington, B.F., and Stone, G.O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, 97, 488-522.

Verhelst, N.D., and Kamphuis, F.H. (2009). *A Poisson-Gamma model for speed tests*. Measurement and Research Department Reports 2009-2. Arnhem: Cito.

Verhelst, N.D., and Kamphuis, F.H. (in druk). *A Gamma-Gamma model for speed tests*. Measurement and Research Department Reports 2010-1. Arnhem: Cito.

Verhoeven, L. en Aarnoutse, C. (2000). *Tussendoelen beginnende geletterdheid*. Nijmegen: Expertisecentrum Nederlands.

Verhoeven, L.T.W. en Leeuwe, J.F.J. van (2003). *Ontwikkeling van decodeervaardigheid in het basisonderwijs*. *Pedagogische Studiën*, vol. 80, 4, 257-271.

Wentink, H. (1997). *From graphemes to syllables*. Academisch proefschrift, Universiteit van Nijmegen.

Ziegler, J.C., Perry, C., and Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology*, 12, 413-430.

Ziegler, J.C., Bertrand, D., Lété, B., & Grainger, J. (2013). Orthographic and Phonological Contributions to Reading Development: Tracking Developmental Trajectories Using Masked Priming. *Developmental Psychology*. Advance online publication. doi: 10.1037/a0035187.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131-1161.

Bijlagen

Bijlage 1

Tabel 1 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M3

		AVI-toets, uitgave 2017		
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	97,3	1,6	0,0
	percentage leerlingen met een score op instructieniveau	2,7	89,1	2,5
	percentage leerlingen met een score op beheersingsniveau	0,0	9,4	97,5
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.				96,5

Tabel 2 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau E3

		AVI-toets, uitgave 2017		
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	94,8	1,4	0,0
	percentage leerlingen met een score op instructieniveau	5,2	98,6	13,3
	percentage leerlingen met een score op beheersingsniveau	0,0	0,0	86,7
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.				90,2

Tabel 3 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M4

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	85,9	0,0	0,0	
	percentage leerlingen met een score op instructieniveau	14,1	93,2	3,6	
	percentage leerlingen met een score op beheersingsniveau	0,0	6,8	96,4	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					94,7

Tabel 4 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau E4

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	80,4	2,0	0,0	
	percentage leerlingen met een score op instructieniveau	19,6	77,9	0,0	
	percentage leerlingen met een score op beheersingsniveau	0,0	20,1	100,0	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					93,3

Tabel 5 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M5

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	95,2	18,8	0,0	
	percentage leerlingen met een score op instructieniveau	4,8	76,8	3,8	
	percentage leerlingen met een score op beheersingsniveau	0,0	4,3	96,2	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					93,7

Tabel 6 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau E5

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	70,0	0,0	0,0	
	percentage leerlingen met een score op instructieniveau	30,0	67,5	0,0	
	percentage leerlingen met een score op beheersingsniveau	0,0	32,5	100,0	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					90,6

Tabel 7 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M6

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	97,8	15,0	0,0	
	percentage leerlingen met een score op instructieniveau	2,2	85,0	10,3	
	percentage leerlingen met een score op beheersingsniveau	0,0	0,0	89,7	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					90,5

Tabel 8 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau E6

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	96,0	12,3	0,0	
	percentage leerlingen met een score op instructieniveau	4,0	87,7	11,5	
	percentage leerlingen met een score op beheersingsniveau	0,0	0,0	88,5	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					89,7

Tabel 9 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau M7

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	93,2	0,0	0,0	
	percentage leerlingen met een score op instructieniveau	6,8	100,0	13,2	
	percentage leerlingen met een score op beheersingsniveau	0,0	0,0	86,8	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					89,5

Tabel 10 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau E7

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	60,8	0,0	0,0	
	percentage leerlingen met een score op instructieniveau	39,2	52,4	0,0	
	percentage leerlingen met een score op beheersingsniveau	0,0	47,6	100,0	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					89,5

Tabel 11 Classificatie van leerlingen bij de oude en nieuwe AVI-toets voor AVI-niveau AVI-Plus

		AVI-toets, uitgave 2017			
		percentage leerlingen met een score op frustratieniveau	percentage leerlingen met een score op instructieniveau	percentage leerlingen met een score op beheersingsniveau	
AVI-toets, uitgave 2009	percentage leerlingen met een score op frustratieniveau	88,0	0,0	0,0	
	percentage leerlingen met een score op instructieniveau	12,0	61,7	0,0	
	percentage leerlingen met een score op beheersingsniveau	0,0	38,3	100,0	
Percentage leerlingen dat bij beide uitgaves van de AVI-toets vergelijkbaar wordt geclassificeerd.					89,7

Bijlage 2

Classificatie-/misclassificatietabellen voor de leestijden op de AVI-toetskaarten

Niveau- groep ↓	Toetskaart AVI-M3A			Niveau- groep ↓	Toetskaart AVI-M3B			Niveau- groep ↓	Toetskaart AVI-M3C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	90,7	9,2	0,1	F	90,7	9,2	0,1	F	90,6	9,3	0,1
I	13,9	65,8	20,3	I	13,2	66,6	20,2	I	13,6	66,3	20,1
B	0,1	3,7	96,2	B	0,0	3,8	96,2	B	0,0	3,8	96,1

Niveau- groep ↓	Toetskaart AVI-E3A			Niveau- groep ↓	Toetskaart AVI-E3B			Niveau- groep ↓	Toetskaart AVI-E3C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	88,1	11,7	0,2	F	88,4	11,5	0,1	F	88,2	11,6	0,2
I	13,1	64,7	22,2	I	13,3	64,3	22,4	I	13,7	64,3	22,0
B	0,1	4,2	95,8	B	0,0	4,0	95,9	B	0,0	4,0	95,9

Niveau- groep ↓	Toetskaart AVI-M4A			Niveau- groep ↓	Toetskaart AVI-M4B			Niveau- groep ↓	Toetskaart AVI-M4C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	81,6	18,4	0,1	F	81,8	18,1	0,0	F	82,2	17,8	0,0
I	7,8	69,5	22,7	I	8,3	69,5	22,2	I	8,6	68,3	23,2
B	0,0	3,6	96,4	B	0,0	3,4	96,6	B	0,0	3,5	96,5

Niveau- groep ↓	Toetskaart AVI-E4A			Niveau- groep ↓	Toetskaart AVI-E4B			Niveau- groep ↓	Toetskaart AVI-E4C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	82,0	18,0	0,0	F	83,0	17,0	0,0	F	82,8	17,1	0,1
I	7,7	71,7	20,6	I	7,7	71,9	20,4	I	7,7	71,6	20,7
B	0,0	3,0	97,0	B	0,0	3,0	97,0	B	0,0	3,1	96,9

Niveau- groep ↓	Toetskaart AVI-M5A			Niveau- groep ↓	Toetskaart AVI-M5B			Niveau- groep ↓	Toetskaart AVI-M5C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	78,4	21,3	0,3	F	78,1	21,6	0,3	F	78,3	21,3	0,5
I	9,9	63,0	27,1	I	9,3	62,5	28,1	I	9,5	62,8	27,7
B	0,0	3,1	96,9	B	0,0	3,1	96,9	B	0,0	2,9	97,1

	Toetskaart AVI-E5A				Toetskaart AVI-E5B				Toetskaart AVI-E5C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	79,0	20,5	0,5	F	79,3	20,1	0,6	F	78,5	20,8	0,7
I	10,5	61,2	28,3	I	11,0	60,6	28,4	I	11,0	60,2	28,8
B	0,1	4,1	95,9	B	0,0	4,0	95,9	B	0,1	3,9	96,1

	Toetskaart AVI-M6A				Toetskaart AVI-M6B				Toetskaart AVI-M6C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	81,5	16,5	2,1	F	81,3	16,5	2,2	F	81,3	16,4	2,3
I	17,3	47,9	34,8	I	17,0	48,1	34,9	I	17,6	47,1	35,3
B	0,2	3,6	96,2	B	0,3	3,8	95,9	B	0,3	3,6	96,1

	Toetskaart AVI-E6A				Toetskaart AVI-E6B				Toetskaart AVI-E6C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	79,1	19,7	1,2	F	79,5	19,2	1,3	F	78,8	19,7	1,6
I	13,0	57,3	29,7	I	14,4	53,2	32,3	I	14,0	53,7	32,3
B	0,1	5,4	94,5	B	0,2	5,2	94,6	B	0,2	5,1	94,6

	Toetskaart AVI-M7A				Toetskaart AVI-M7B				Toetskaart AVI-M7C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	77,9	19,4	2,6	F	77,8	19,7	2,4	F	78,7	18,6	2,7
I	15,0	48,2	36,8	I	15,1	48,4	36,5	I	15,1	47,8	37,0
B	0,3	3,9	95,8	B	0,3	3,9	95,8	B	0,3	4,0	95,8

	Toetskaart AVI-E7A				Toetskaart AVI-E7B				Toetskaart AVI-E7C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	76,7	22,1	1,1	F	68,8	31,2	0,0	F	76,5	22,5	1,0
I	11,4	57,2	31,4	I	2,4	76,3	21,3	I	10,5	59,3	30,1
B	0,1	3,7	96,2	B	0,0	4,0	96,0	B	0,1	3,9	96,0

	Toetskaart AVI-PlusA				Toetskaart AVI-PlusB				Toetskaart AVI-PlusC		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	F	I	B		F	I	B		F	I	B
F	79,5	20,3	0,2	F	78,7	21,2	0,1	F	78,9	20,9	0,2
I	9,2	67,7	23,1	I	7,7	71,2	21,1	I	9,6	67,6	22,8
B	0,0	5,6	94,4	B	0,0	5,7	94,3	B	0,0	5,7	94,2

Bijlage 3

Classificatie-/misclassificatietabellen voor de leesfouten op de AVI-toetskaarten

Niveau- groep ↓	Toetskaart AVI-M3A			Niveau- groep ↓	Toetskaart AVI-M3B			Niveau- groep ↓	Toetskaart AVI-M3C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	97,3	2,7	0,1	B	96,6	3,4	0,0	B	97,0	2,9	0,0
I	38,8	54,1	7,1	I	36,9	56,0	7,1	I	38,2	54,3	7,5
F	3,3	40,8	55,9	F	2,4	41,6	56,0	F	2,5	39,4	58,0

Niveau- groep ↓	Toetskaart AVI-E3A			Niveau- groep ↓	Toetskaart AVI-E3B			Niveau- groep ↓	Toetskaart AVI-E3C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	94,2	5,7	0,1	B	94,9	5,0	0,0	B	95,2	4,7	0,0
I	34,3	59,2	6,5	I	30,1	63,0	6,9	I	33,0	60,4	6,6
F	1,6	43,4	55,0	F	0,6	36,6	62,8	F	1,1	40,3	58,5

Niveau- groep ↓	Toetskaart AVI-M4A			Niveau- groep ↓	Toetskaart AVI-M4B			Niveau- groep ↓	Toetskaart AVI-M4C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	97,3	2,7	0,0	B	97,1	2,9	0,0	B	97,7	2,3	0,1
I	48,6	45,5	5,9	I	45,0	50,9	4,1	I	49,8	44,0	6,3
F	6,6	50,4	43,0	F	4,1	55,9	40,1	F	8,0	45,3	46,6

Niveau- groep ↓	Toetskaart AVI-E4A			Niveau- groep ↓	Toetskaart AVI-E4B			Niveau- groep ↓	Toetskaart AVI-E4C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	93,3	6,6	0,1	B	94,4	5,6	0,0	B	94,4	5,5	0,1
I	34,6	59,7	5,7	I	32,3	62,2	5,6	I	36,9	56,8	6,3
F	1,9	46,9	51,2	F	0,9	42,2	56,9	F	2,5	46,3	51,2

Niveau- groep ↓	Toetskaart AVI-M5A			Niveau- groep ↓	Toetskaart AVI-M5B			Niveau- groep ↓	Toetskaart AVI-M5C		
	Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt				Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	96,8	3,2	0,0	B	96,9	3,1	0,0	B	95,7	4,3	0,0
I	45,8	50,8	3,4	I	44,4	50,9	4,7	I	43,1	53,6	3,3
F	4,1	56,3	39,5	F	4,9	51,8	43,3	F	3,2	58,4	38,4

	Toetskaart AVI-E5A				Toetskaart AVI-E5B				Toetskaart AVI-E5C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	94,4	5,6	0,0	B	95,2	4,7	0,0	B	95,5	4,5	0,0
I	41,8	56,2	2,0	I	39,4	58,2	2,4	I	46,0	50,7	3,3
F	2,2	65,5	32,3	F	3,0	59,5	37,6	F	5,2	59,7	35,1

	Toetskaart AVI-M6A				Toetskaart AVI-M6B				Toetskaart AVI-M6C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	94,5	5,5	0,0	B	95,9	4,1	0,0	B	96,7	3,3	0,1
I	47,4	50,3	2,3	I	49,8	47,5	2,7	I	50,9	44,3	4,8
F	7,4	65,0	27,7	F	8,9	63,4	27,7	F	9,7	55,3	35,0

	Toetskaart AVI-E6A				Toetskaart AVI-E6B				Toetskaart AVI-E6C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	93,7	6,2	0,0	B	91,7	8,2	0,0	B	91,5	8,5	0,0
I	42,2	55,1	2,8	I	43,4	54,2	2,4	I	42,4	55,1	2,5
F	4,0	63,5	32,5	F	5,1	69,2	25,7	F	4,7	68,4	27,0

	Toetskaart AVI-M7A				Toetskaart AVI-M7B				Toetskaart AVI-M7C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	96,0	3,9	0,1	B	95,3	4,6	0,0	B	96,3	3,7	0,0
I	47,3	48,3	4,4	I	38,5	58,3	3,1	I	43,5	52,5	3,9
F	6,7	55,7	37,6	F	1,7	56,4	41,8	F	3,7	57,0	39,3

	Toetskaart AVI-E7A				Toetskaart AVI-E7B				Toetskaart AVI-E7C		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	96,0	4,0	0,0	B	95,0	4,9	0,0	B	96,1	3,8	0,0
I	42,2	54,6	3,1	I	46,6	50,4	3,0	I	45,6	51,5	2,9
F	3,1	60,5	36,3	F	6,1	64,0	29,9	F	5,9	60,0	34,1

	Toetskaart AVI-PlusA				Toetskaart AVI-PlusB				Toetskaart AVI-PlusC		
Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt			Niveau-groep ↓	Scoregroep waarin de ware score valt		
	B	I	F		B	I	F		B	I	F
B	95,5	4,4	0,1	B	95,4	4,5	0,1	B	95,7	4,2	0,1
I	46,2	47,9	5,9	I	45,3	48,4	6,3	I	42,6	51,3	6,1
F	7,4	49,7	42,9	F	7,2	49,5	43,3	F	4,9	47,2	47,8

Bijlage 4

Classificatie-/misclassificatietabellen voor de combinatie van leesfouten en leestijden op de AVI-leeskaarten

AVI-M3A					AVI-M3B					AVI-M3C				
	scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F	
↓					↓					↓				
B	96,1	3,9	0		B	97,4	2,6	0		B	96,3	3,7	0	
I	20	62,9	17,1		I	21,7	65,2	13		I	27,4	56,5	16,1	
F	0	3,5	96,5		F	0	7,3	92,7		F	0,6	9,7	89,7	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				92,1	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				92,3	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				90,2

AVI-E3A					AVI-E3B					AVI-E3C				
	scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F	
↓					↓					↓				
B	96,5	3,5	0		B	96,3	3,5	0,3		B	97,3	2,5	0,3	
I	28	52,7	19,4		I	29,5	51,3	19,2		I	34	46,6	19,4	
F	0,7	8,1	91,2		F	0	10,3	89,7		F	0	5,6	94,4	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88,5	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88,9	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88

AVI-M4A					AVI-M4B					AVI-M4C				
	scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F	
↓					↓					↓				
B	96,7	3,3	0		B	95,8	4,2	0		B	96,2	3,8	0	
I	24,5	66	9,4		I	23,1	68,5	8,3		I	26,7	62,9	10,5	
F	0	26,7	73,3		F	1,5	19,1	79,4		F	1,3	22,7	76	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88,1	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88,8	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				87,6

AVI-E4A					AVI-E4B					AVI-E4C							
		scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score					
score-groep van de sample		B	I	F		score-groep van de sample		B	I	F		score-groep van de sample		B	I	F	
↓						↓						↓					
B		95,5	4,5	0		B		94,7	5,3	0		B		94,9	5,1	0	
I		24,4	69	6,5		I		23,8	64,4	11,9		I		26,5	62,3	11,3	
F		0,9	20,5	78,6		F		0	21,9	78,1		F		0	25,7	74,3	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd					87,5	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					86,3	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					85,7

AVI-M5A					AVI-M5B					AVI-M5C							
		scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score					
score-groep van de sample		B	I	F		score-groep van de sample		B	I	F		score-groep van de sample		B	I	F	
↓						↓						↓					
B		97,4	2,6	0		B		96,1	3,9	0		B		97,2	2,8	0	
I		35,5	59,2	5,3		I		30,2	57,1	12,7		I		34,2	56,6	9,2	
F		0	18,3	81,7		F		0	20,3	79,7		F		1,4	21,1	77,5	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd					90,4	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					89,9	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					89,4

AVI-E5A					AVI-E5B					AVI-E5C							
		scoregroep van de ware score					scoregroep van de ware score					scoregroep van de ware score					
score-groep van de sample		B	I	F		score-groep van de sample		B	I	F		score-groep van de sample		B	I	F	
↓						↓						↓					
B		93,5	6,3	0,2		B		95,3	4,5	0,2		B		95,7	4,3	0	
I		32,2	58,9	8,9		I		27	64,3	8,7		I		38,5	51,4	10,1	
F		0	15,2	84,8		F		0,9	17,6	81,5		F		1	17,5	81,6	
percentage leerlingen dat vergelijkbaar wordt geclassificeerd					85,3	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					87,8	percentage leerlingen dat vergelijkbaar wordt geclassificeerd					84,6

AVI-M6A					AVI-M6B					AVI-M6C							
		scoregroep van de ware score						scoregroep van de ware score						scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F				
↓					↓					↓							
B	95,4	4,6	0		B	96,6	3,2	0,2		B	97,2	2,8	0				
I	46,1	47,4	6,6		I	47,9	38	14,1		I	46,7	44	9,3				
F	0	14,1	85,9		F	1	19,8	79,2		F	2	14,9	83,2				
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				87,8	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				86,8	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				88,3			

AVI-E6A					AVI-E6B					AVI-E6C							
		scoregroep van de ware score						scoregroep van de ware score						scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F				
↓					↓					↓							
B	94,1	5,9	0		B	93,9	6,1	0		B	92,5	7,3	0,2				
I	36	52,7	11,3		I	35,7	54,5	9,8		I	35,7	49,4	14,9				
F	1,4	21	77,5		F	2,8	18,1	79,2		F	1,6	16,7	81,7				
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				82,7	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				83,5	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				81,7			

AVI-M7A					AVI-M7B					AVI-M7C							
		scoregroep van de ware score						scoregroep van de ware score						scoregroep van de ware score			
score-groep van de sample	B	I	F		score-groep van de sample	B	I	F		score-groep van de sample	B	I	F				
↓					↓					↓							
B	95,2	4,3	0,5		B	92,7	7,3	0		B	95,9	3,8	0,2				
I	38,8	50,6	10,6		I	43,7	48,3	8		I	38,6	51,8	9,6				
F	3,7	16	80,2		F	2,4	14,3	83,3		F	1,2	18,3	80,5				
percentage leerlingen dat vergelijkbaar wordt geclassificeerd				87	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				85	percentage leerlingen dat vergelijkbaar wordt geclassificeerd				87,8			

Cito

Amsterdamseweg 13
6814 CM Arnhem
Postbus 1034
6801 MG Arnhem
T (026) 352 11 11

Fotografie: Ron Steemers

