

# Wetenschappelijke verantwoording LVS-toetsen DMT

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

[cito.nl](https://cito.nl)





# **Wetenschappelijke verantwoording LVS-toetsen DMT**

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

© Cito B.V. Arnhem (2018)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito B.V. 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.

# 1 Voorwoord

Bij de totstandkoming van de Drie-Minuten-Toets (DMT) 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 volgende Nederlandse en Vlaamse vakexperts die ons in verschillende stadia van het project hebben geadviseerd:

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



# Inhoud

<b>1</b>	<b>Inleiding</b>	<b>7</b>
<b>2</b>	<b>Uitgangspunten van de toetsconstructie</b>	<b>9</b>
2.1	Meetpretentie	9
2.2	Doelgroep	9
2.3	Gebruiksdoel	9
2.4	Theoretische inkadering	12
<b>3</b>	<b>Beschrijving van de toets</b>	<b>17</b>
3.1	Opbouw en structuur van de toets	17
3.2	Inhoudsverantwoording	18
3.2.1	Uitwerking domeinbeschrijving	19
3.2.2	Woordselectie en samenstelling van de leeskaarten	20
3.2.3	Gerealiseerde leeskaarten	23
3.2.4	Screening en try-out	24
3.2.5	Proeftoetsing	25
<b>4</b>	<b>Kalibratie en normering</b>	<b>27</b>
4.1	Dataverzameling	27
4.2	Representativiteit van de steekproef	28
4.3	Beschrijving van het meetmodel	40
4.4	Kalibratie	42
4.5	Normering	47
<b>5</b>	<b>Betrouwbaarheid en meetnauwkeurigheid</b>	<b>51</b>
5.1	Betrouwbaarheid	51
5.2	Nauwkeurigheid	53
<b>6</b>	<b>Validiteit</b>	<b>57</b>
6.1	Inhoudsvaliditeit	57
6.2	Begripsvaliditeit	57
6.3	Convergente en divergente validiteit	57
6.4	Verschillen tussen relevante subgroepen	61
<b>7</b>	<b>Samenvatting</b>	<b>63</b>
	<b>Literatuurlijst</b>	<b>65</b>
	<b>Bijlagen</b>	<b>69</b>





# 1 Inleiding

Deze wetenschappelijke verantwoording heeft betrekking op de DMT (Drie-Minuten-Toets). De DMT is bestemd voor groep 3 tot en met halverwege groep 8 van het primair en speciaal onderwijs. De DMT 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 DMT 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 DMT mogelijk op de volgende aspecten:

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

Het laatste aspect betreft alleen begripsvaliditeit en géén criteriumvaliditeit. Omdat de toetsen van het Cito Volgsysteem 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 DMT. 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 DMT 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 DMT. 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 de 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 DMT 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 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 langer over het lezen doen dan andere leerlingen.

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

Voor de toetsen van groep 3 tot en met groep 7 zijn zowel voor ‘midden leerjaar’ (half januari tot half februari) als voor ‘einde leerjaar’ (half mei tot en met eind juni) populatieparameters vastgesteld. Voor de toets in groep 8 zijn alleen populatieparameters bepaald voor midden leerjaar. De toetsen kunnen desgewenst ook op andere momenten in het schooljaar worden afgenomen, maar in dat geval is het moeilijker om uitspraken te doen over het niveau van een leerling ten opzichte van andere leerlingen in Nederland.

Er zijn twee uitgaven van de DMT 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 DMT 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 DMT diagnostische informatie, zodat inzicht geboden kan worden in het leesgedrag van de leerling. De verschillende doelen worden hieronder toegelicht.

#### **Niveaubepaling**

De resultaten van de toets geven informatie over het leesvaardigheidsniveau van leerlingen en groepen leerlingen. Iedere behaalde leesvaardigheidsscore kan daartoe normgericht geïnterpreteerd worden op

basis van de vaardigheidsverdeling in een referentiegroep (zie paragraaf 4.2 voor de beschrijving van de referentiegroep). De referentiegroep is op basis van de scores van de leerlingen in deze groep op twee manieren in vijf vaardigheidsniveaus verdeeld (zie figuur 2.1). De eerste manier, met de vaardigheidsniveaus I tot en met V, gaat uit van vijf groepen van ieder 20%. Op de rapportages wordt bij de indeling in I tot en met V, de hoogste groep nog onderverdeeld in twee groepen die ieder 10% leerlingen bevatten. Deze groepen worden van elkaar gescheiden door een stippellijn. De tweede indeling, die gebaseerd is op een indeling in kwartielen, levert de vaardigheidsniveaus A tot en met E op. De niveaugroepen A, B en C bestrijken elk een kwart van de populatie. Het vierde kwartiel wordt opgesplitst in twee subgroepen: D (15%) en E (10%).

De eerstgenoemde indeling in niveaugroepen is symmetrisch opgebouwd en heeft als voordeel ten opzichte van de indeling gebaseerd op kwartielen dat er een gemiddelde<sup>1</sup> groep onderscheiden wordt, namelijk vaardigheidsniveau III. Deze indeling blijkt in de praktijk intuïtiever aan te voelen en minder gevoelig te zijn voor verkeerde interpretaties. (Sommige leerkrachten beschouwen vaardigheidsniveau C ten onrechte als gemiddelde groep.) Om die reden wordt in de handleiding steeds eerst de indeling I tot en met V genoemd en pas daarna de indeling A tot en met E.

*Figuur 2.1 Onderscheiden niveaugroepen*

I – V		A – E	
20% hoogst scorende leerlingen	I 20%	A 25%	25% hoogst scorende leerlingen
20% boven het landelijk gemiddelde	II 20%	B 25%	25% ruim boven tot net boven het landelijk gemiddelde
20% landelijk gemiddelde	III 20%		
20% onder het landelijk gemiddelde	IV 20%	C 25%	25% net tot ruim onder het landelijk gemiddelde
20% laagst scorende leerlingen	V 20%	D 15%	15% ruim onder het landelijk gemiddelde
		E 10%	10% laagst scorende leerlingen

### **Progressiebepaling**

Het volgen van leerlingen in hun groei, ook wel aangeduid als progressiebepaling, is een van de belangrijkste functies van het Cito Volgsysteem 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?

<sup>1</sup> Het betreft hier geen gemiddelde in de statistische betekenis van het woord.

Het gehanteerde meetmodel (zie paragraaf 4.3) maakt het mogelijk om de scores van een leerling op verschillende afnamecombinaties van leeskaarten, op verschillende momenten onderling te vergelijken. De ruwe scores op de toetsen (de ruwe score is het totaal aantal goed gelezen woorden op de combinatie van gebruikte leeskaarten) zijn daartoe te transformeren in scores op één vaardigheidsschaal. Het aantal afnamemomenten per jaar wordt bepaald door het tempo waarin een vaardigheid gemiddeld gesproken binnen een leerjaar en over de gehele schoolperiode toeneemt. Meestal is er sprake van twee afnamemomenten per leerjaar ('medio' en 'einde' leerjaar, aangeduid als M en E). Dit geldt ook voor de DMT. Overigens is er geen bezwaar om de DMT (ook) op andere momenten in het jaar af te nemen. Voor andere afnamemomenten dan M of E zijn echter geen normeringstabellen gemaakt.

De groei van leerlingen kan in kaart worden gebracht door hun prestaties op de verschillende afnamemomenten te vergelijken. Die vergelijking kan op drie verschillende manieren worden gemaakt. Bij de eerste manier worden de vaardigheidsscores vergeleken, bijvoorbeeld: op tijdstip E3 had Dani vaardigheidsscore 20 en op tijdstip M4 had hij vaardigheidsscore 39. Bij de tweede manier vindt de vergelijking van resultaten plaats op basis van functioneringsniveaus. Een functioneringsniveau E4 betekent bijvoorbeeld dat de vaardigheidsscore van de leerling overeenkomt met de score van de gemiddelde leerling eind groep 4. Met behulp van deze informatie kan voor elke leerling een passend leertraject worden opgesteld. Ten slotte kunnen de prestaties van leerlingen op verschillende afnamemomenten ook worden vergeleken door te kijken naar vaardigheidsniveaus. Als een leerling op vijf achtereenvolgende afnamemomenten vaardigheidsniveau II scoort en op het eerstvolgende moment een IV, kan dit een reden zijn om de resultaten nader te analyseren.

Elke toets heeft een bepaalde meetfout. Als we rekening houden met die meetfout dan zou het best zo kunnen zijn dat Sanneke vaardigheidsniveau III heeft behaald op het eerste tijdstip en niet vaardigheidsniveau IV. Of dat we moeten concluderen dat het verschil in de prestaties op de twee tijdstippen statistisch niet significant is: Sanneke is voor-, noch achteruitgegaan. In alle gevallen speelt het betrouwbaarheidsinterval (BI) rondom de vaardigheidsscore een belangrijke rol. Dat geldt ook voor de indeling in vaardigheidsniveaus. Hoe het BI doorwerkt in de indeling in vaardigheidsniveaus en de verdere gevolgen daarvan wordt beschreven in hoofdstuk 5.

Op deze plaats beperken we ons tot de vraag of we de uitspraak kunnen doen dat een leerling of groep (werkelijk) 'gegroeid' is. De eenvoudigste manier is om te kijken of de BI's voor de twee tijdstippen overlappen. Als deze twee BI's niet overlappen dan is er sprake van een significant verschil in vaardigheid tussen beide tijdstippen. Overlappen ze wel, dan is er misschien helemaal geen verschil in vaardigheid.

We geven hier een voorbeeld. Bij de afname op het E3-moment behaalde Harun een vaardigheidsscore van 23 (totaalscore 82) met een score-interval (het 67%-betrouwbaarheidsinterval, zie ook paragraaf 5.2 van deze verantwoording) van 21 - 26. Bij de afname M4 behaalde hij een vaardigheidsscore van 33 (totaalscore 111); het bijbehorende betrouwbaarheidsinterval is 30 - 36. Aangezien de betrouwbaarheidsintervallen niet overlappen, kunnen we er zeker van zijn dat Haruns vaardigheid is toegenomen.

Uitspraken over vaardigheidsgroei dienen altijd met voorzichtigheid te worden gehanteerd. Hoe de leerkracht hier in de praktijk mee om moet gaan, wordt toegelicht in de handleiding bij het toetsmateriaal.

### **Diagnostiek**

De indeling in leeskaarten met verschillende woordtypen maakt het mogelijk om na te gaan hoe een leerling scoort op verschillende woordtypen. Met behulp van de categorieënanalyse (digitaal) en/of kaartenanalyse (handmatig) kan worden bepaald met welke woordtypen leerlingen vooral moeite hebben. Individuele leerlingen die moeite hebben met bepaalde woordtypen, of die steeds dezelfde soorten leesfouten maken, zullen wellicht baat hebben bij extra instructie en gerichte oefeningen.

Daarnaast kan door afname van de DMT worden nagegaan of de leerling achterblijft op de nauwkeurigheid van het lezen. Dit wordt kenbaar gemaakt aan de hand van een signalering.

Tot slot wordt de leerkracht of zorgaanbieder de mogelijkheid geboden om in een overzicht op de scoreformulieren te noteren welk type fouten de leerling heeft gemaakt om zo meer zicht te krijgen op het leesgedrag van de leerling. Maakt hij of zij bijvoorbeeld fouten in het verklanken van woorden of leest hij nog vaak spellend? 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 voor 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 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 DMT is een methodeonafhankelijk instrument voor het meten van de technische leesvaardigheid. Het maakt voor de afname van de DMT 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 (mondelijke) 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, pseudowoorden), 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 pseudowoorden; 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 (Schijf, 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 knooppfamilies worden verondersteld: letterknopen, foneemknopen en semantische kenmerknopen (Bosman, 2000). De activatie tussen de knooppfamilies 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 knooppfamilies zijn betrokken. In dit proces komt woordherkenning tot stand door onderlinge beïnvloeding van knopen van verschillende families, waaruit relatief stabiele terugkoppelingslussen ontstaan. Relaties tussen knopen uit verschillende knooppfamilies 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.

#### *Amalgaammodellen*

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 Drie-Minuten-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 DMT geoperationaliseerd worden, wordt beschreven in hoofdstuk 3.

## 3 Beschrijving van de toets

### 3.1 Opbouw en structuur van de toets

Het toetspakket voor de DMT bestaat uit een handleiding en papieren toetsmaterialen (leeskaarten, score-formulieren, 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 DMT is een individuele toets. Het is de taak van de leerling om de losse woorden op een leeskaart gedurende één minuut 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 voor zorgaanbieders) drie leeskaarten opgenomen die elk gedurende één minuut hardop moeten worden voorgelezen. De drie leeskaarten bevatten elk 150 woorden. De woordtypen die zijn geselecteerd verschillen per leeskaart:

- Op leeskaart 1 staan woorden van het type 'km' (klinker(s) gevolgd door medeklinker, zoals bijvoorbeeld ijs), 'mk' (koe) en 'mkm' (vaas).
- Op leeskaart 2 staan woorden van het type 'mmkm' (spoor), 'mkmm' (kers), 'mmkmm' (plant), 'mmmkm' (sproet) en 'mkmmm' (liefst).
- Op leeskaart 3 staan woorden met twee, drie en vier lettergrepen, zoals 'groente', 'hartelijk' en 'schoolvakantie'.

De uitgave voor het onderwijs bevat twee parallelversies van elke leeskaart. Dat betekent dat er van elke leeskaart een A- en een B-versie is. De uitgave voor zorgaanbieders bevat slechts één versie van elke leeskaart. Dit is de C-versie. Versie A en B bevatten dezelfde woorden, maar in een andere volgorde. Zo wordt voorkomen dat leerlingen bij herhaald toetsen gedurende de basisschool de eerste woorden van iedere leeskaart onthouden en op voorhand kunnen opzeggen. De leeskaarten van versie C bevatten andere woorden dan versie A en B, maar zijn qua opzet volledig vergelijkbaar. Op deze manier kan voorkomen worden dat de leerlingen de inhoud van de leeskaarten al kennen door frequente afname in het onderwijs.

#### Afname

De DMT wordt individueel afgenomen door een toetsleider. Dat kan de leerkracht zijn, maar ook een intern begeleider of zorgaanbieder. De toetsafname begint met een instructie. Daarna legt de toetsleider de leerling een leeskaart voor en geeft het beginsignaal. Tegelijkertijd start de tijdwaarneming. De leerling krijgt één minuut de tijd om zoveel mogelijk woorden van de leeskaart hardop te lezen. De woorden moeten per kolom, van boven naar beneden, worden gelezen. In tabel 3.1 is te zien welke leeskaarten op welk momenten worden afgenomen.

Op het eerste afnamemoment (midden groep 3) hebben de meeste leerlingen nog geen meerlettergrepige woorden gelezen. Daarom is het raadzaam om bij hen alleen leeskaart 1 en 2 af te nemen. Vanaf eind groep 3 is het wel zinvol om alle drie de leeskaarten af te nemen. Vanaf het midden van jaargroep 5 kan overwogen worden om leeskaart 1 niet meer voor te leggen, met name bij de betere leerlingen, waarvan kan worden aangenomen dat het proces van het leren lezen zover gevorderd is, dat ze geen problemen meer hebben met de eenlettergrepige woorden zonder medeklinkercombinaties van leeskaart 1. Op de Cito Portal worden hier nadere richtlijnen voor gegeven.

Tabel 3.1 Afnameschema voor de leeskaarten van groep 3 tot en met groep 8

Afnamemoment	Periode	Leeskaart
Midden groep 3 (M3)	half januari tot half februari	1, 2
Eind groep 3 (E3)	half mei tot en met eind juni	1, 2, 3
Midden groep 4 (M4)	half januari tot half februari	1, 2, 3
Eind groep 4 (E4)	half mei tot en met eind juni	1, 2, 3
Midden groep 5, 6, 7 en 8 (M5, M6, M7, M8)	half januari tot half februari	(1,) 2, 3
Eind groep 5, 6 en 7 (E5, E6, E7)	half mei tot en met eind juni	(1,) 2, 3

Om de resultaten van de leerlingen te kunnen vergelijken met die van andere leerlingen uit het land is het belangrijk dat de toetsleider zich houdt aan de vermelde toetsperiodes. Tabel 3.1 betreft een afnameschema waarin voor elk afnamemoment is aangegeven wat de afnameperiode is en welke combinatie van leeskaarten aan de orde is.

### Scoring

Tijdens het lezen noteert de toetsleider op het bijbehorende scoreformulier welke woorden fout gelezen worden. De scorevoorschriften en notatiewijzen voor diverse foutensoorten zijn beschreven op de 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, behalve als zij daarna alsnog correct verklankt worden.

Op basis van het totaal aantal goed gelezen woorden, de totaalscore, wordt een inschatting gemaakt van de vaardigheid van de leerlingen. De leerkracht of intern begeleider kan de totaalscore invoeren in het Computerprogramma LOVS. De totaalscore wordt zo automatisch omgezet naar de bijbehorende vaardigheidsscore met een score-interval ofwel betrouwbaarheidsinterval. Een andere optie is om met behulp van de omzettingstabellen in de toetsmap of op Cito Portal de vaardigheidsscore bij de behaalde totaalscore op te zoeken.

### Verwerking resultaten en interpretatie

Na de toetsafname en de scoring kunnen de toetsresultaten verwerkt worden op speciaal ontwikkelde rapportageformulieren. 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. Ze zijn, met uitzondering van de kaartenanalyse, niet opgenomen in de uitgave voor zorgaanbieders omdat deze voor hen niet relevant zijn. In hoofdstuk 3 van de handleiding is een overzicht opgenomen van de belangrijkste rapportages. In hoofdstuk 4 van de handleiding wordt toegelicht hoe de rapportages op leerling- en groepsniveau geïnterpreteerd kunnen worden. In hoofdstuk 5 van de handleiding worden de analyses op schoolniveau (zoals dwarsdoorsnedes en trendanalyses) beschreven. De handleiding in de toetsmap gaat in op de inhoudelijke interpretatie van de rapportages. 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.

## 3.2 Inhoudsverantwoording

In het ontwikkelproces van de DMT zijn een aantal fasen te onderscheiden:

- Uitwerking van de domeinbeschrijving;
- Wordselectie en samenstelling van de leeskaarten
- Screening en try-out
- Proeftoetsing

- Kalibratie en normering
- Uitgave

De eerste vier fases, 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 DMT onder een zekere tijdsdruk afgenomen.

Om aan te kunnen sluiten bij het tweede element van de definitie – het herkennen van woorden als dragers van betekenis – is ervoor gekozen om in de DMT bestaande woorden op te nemen en geen pseudoworden. Een tweede reden om voor bestaande woorden te kiezen, heeft te maken met het doel van technisch lezen. Zoals hierboven al is vermeld, leren leerlingen technisch lezen om geschreven teksten te kunnen begrijpen en in deze teksten komen over het algemeen geen pseudoworden voor.

Met behulp van de DMT kan worden nagegaan hoe goed leerlingen *geïsoleerde* woorden van *uiteenlopende moeilijkheidsgraad* *hardop* kunnen verklanken. Waarom hiervoor gekozen is, wordt hieronder uitgelegd.

#### **Waarom 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. Bij het lezen van de DMT-leeskaarten kan dit compensatiemechanisme niet in werking treden. De DMT voorziet dan ook in een zuivere meting van de technische leesvaardigheid; ook in het geval van een meting bij zwakkere technisch-lezers, in zoverre dat deze leerlingen niet de kans krijgen 'zich beter voor te doen' door gebruik te maken van contextinformatie. Anderzijds is het lezen van geïsoleerde woorden natuurlijk een leestaak die wat verder van het lezen van alledag af staat dan het lezen van woorden in de context van een zin of tekst.

#### **Waarom woorden van uiteenlopende moeilijkheidsgraad?**

Woordspecifieke kenmerken blijken van invloed op de leesscores van leerlingen. Dit is al geïllustreerd in paragraaf 2.4.1 van deze verantwoording waarin onderzoek beschreven is dat uitgevoerd werd om de daar gepresenteerde theoretische leesmodellen empirisch te ondersteunen. De snelheid en accuratesse waarmee woorden worden gelezen is afhankelijk van factoren als:

- *Woordbekendheid*: betekenisvolle woorden worden sneller gelezen dan pseudoworden (per definitie onbekend) (zie bijvoorbeeld Taylor & Lupker, 2001).
- *Woordfrequentie*: naarmate woorden frequenter voorkomen in geschreven taal, worden ze sneller en accurater gelezen (zie bijvoorbeeld Brysbaert, Lange, & Van Wijnendaele, 2000 voor evidentie in de Nederlandse taal).
- *Woordlengte*: korte woorden worden over het algemeen sneller herkend dan langere woorden (zie bijvoorbeeld New, Ferrand, Pallier, & Brysbaert, 2006).

- *Voorstelbaarheid* (imageability): woorden met meer sensorische/motorische representatie, woorden waarbij je iets (een plaatje) kunt voorstellen, worden sneller gelezen dan laagvoorstelbare woorden (zie bijvoorbeeld Strain, Patterson, & Seidenberg, 1995).
- *Aantal 'orthografische burens'*: Dit aantal kan worden bepaald door één letter van een woord te veranderen, terwijl de andere letters op hun plaats blijven staan. Zo zijn de woorden: klap, klos en plas allemaal orthografische burens van het woord klas. Woorden met weinig 'burens' worden langzamer gelezen dan woorden met veel 'burens' (zie bijvoorbeeld Huntsman, & Lima, 2002). Bovenstaande factoren zijn aan elkaar gerelateerd en hebben een gezamenlijk effect op de leesbaarheid. Bijvoorbeeld: het effect van het aantal orthografische burens hangt samen met de woordlengte (kortere woorden hebben meer orthografische burens).

De scores op de drie leeskaarten geven informatie over de technische leesvaardigheid van leerlingen in het lezen van woorden van uiteenlopende woordtypen. De woorden op de onderscheiden leeskaarten verschillen in orthografische structuur en/of complexiteit, en daarmee in moeilijkheidsgraad. Een score op leeskaart 1 geeft aan hoe goed leerlingen eenlettergrepige woorden zonder medeklinkerclusters kunnen verklanken. Afname van leeskaart 2 geeft informatie over de mate waarin leerlingen letterclusters en spellingpatronen kunnen herkennen. Ten slotte tonen de resultaten op leeskaart 3 aan in hoeverre leerlingen meerlettergrepige woorden, zoals *pa-pe-gaai* kunnen lezen en betekenselementen, zoals in het woord *keuken-tafel-tje* weten af te leiden.

### 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 hardop te laten lezen, omdat snelheid en accuratesse dan direct in kaart kunnen worden gebracht en niet hoeven te worden afgeleid.

### 3.2.2 Wordselectie en samenstelling van de leeskaarten

Het selecteren van woorden en het samenstellen van leeskaarten op basis van de selectie bestond uit een procedure van acht stappen. Deze stappen zijn hieronder beschreven.

#### Stap 1: steekproeftrekking uit het BasiLex corpus

Voor de samenstelling van de leeskaarten is een aselechte steekproef getrokken van woorden uit het BasiLex-corpus (Tellings, Hulsbosch, Vermeer, & van den Bosch, 2014)<sup>2</sup>. BasiLex is een verzameling van geschreven teksten die Nederlandse kinderen in de basisschoolleeftijd vanaf groep 3 aangeboden krijgen. BasiLex is voor 42% gebaseerd op onderwijsmateriaal (lesboeken, toetsen), voor 38% op lees- en stripboeken en tijdschriften voor kinderen (onder andere Donald Duck) en voor 20% op media: websites die kinderen veel bezoeken (bijvoorbeeld van het Jeugdjournaal en van voetbalclubs) en ondertitels van veelbekeken kinderprogramma's (zoals Dance Academy, en programma's op Disney Channel). De (11,5 miljoen) woorden zijn onder meer gelabeld op frequentie, woordsoort, lengte en leeftijd (voor welke leeftijd is de tekst waaruit het woord komt bestemd).

#### Stap 2: selectie op basis van woordsoorten

Uit de steekproef van woorden (totaal 158.745) die wij ontvingen van de ontwikkelaars van BasiLex zijn eerst de zelfstandig naamwoorden en de adjectieven (bijvoeglijk naamwoorden) geselecteerd. Dit resulteerde in een verzameling van 92.586 inhoudswoorden. Er is gekozen voor deze inhoudswoorden, omdat toevoeging van een substantieel aantal overige woordsoorten, zoals werkwoorden en bijwoorden,

<sup>2</sup> Tellings, A., Hulsbosch, M., Vermeer, A. & van den Bosch, A. (2014). BasiLex: an 11.5 million word corpus of Dutch texts written for children. *Computational Linguistics in the Netherlands*, 4, 191-208.

ertoe zou kunnen leiden dat er onbedoeld halve zinnen konden ontstaan op de leeskaarten (bijvoorbeeld: meisjes aarzelen niet altijd). Een ander nadeel van overige woordsoorten is dat deze vaak te frequent voorkomen; denk aan lidwoorden, voegwoorden en voornaamwoorden. In de onderstaande tabel is de gewenste verdeling van woordsoorten voor elke leeskaart weergegeven. Elk vermeld percentage is daarbij omgeven door een bandbreedte.

Tabel 3.2 Beoogde verdeling van woordsoorten voor kaart 1, 2 en 3

	Woordsoorten		
	Zelfstandig naamwoord	Bijvoeglijk naamwoord	Overige woordsoorten
Na te streven percentage per kaart	75% (+/-10%)	Ongeveer 20% (+/-5%)	Ongeveer 7% (+/-3%)

### Stap 3: selectie op basis van woordfrequentie

De frequentie van de woorden in het bestand met inhoudswoorden liep uiteen van 1 tot 541.605. In de uiteindelijke selectie zijn alleen woorden opgenomen met een frequentie van  $> 10$  en  $\leq 5.000$ .

Ruim 13.000 woorden (13.359) voldeden aan dit criterium. Woorden met een frequentie lager dan 10 zijn uitgesloten, omdat het te veel afhangt van het toeval of leerlingen deze woorden al eens zijn tegengekomen, wat negatieve gevolgen heeft voor de betrouwbaarheid van de toets. Bovendien mogen anderstalige leerlingen met een kleinere woordenschat in de Nederlandse taal niet benadeeld worden. Wanneer de woorden relatief onbekend zijn doordat ze nauwelijks voorkomen in de geschreven taal op de basisschool, wordt de leestaak voor deze groep leerlingen moeilijker. Slechts 269 woorden hadden een frequentie die groter is dan 5000. Deze woorden komen zo vaak voor en zijn zo bekend, dat ze geen representatieve meting van de technische leesvaardigheid opleveren. De set van woorden met een frequentie tussen 11 en 5000 biedt een range aan laag- en hoogfrequente woorden. Het opnemen van relatief laagfrequente woorden is belangrijk, omdat daarmee gemeten kan worden of de leerling minder bekende woorden kan verklanken. Het opnemen van hoogfrequente woorden brengt goed de mate van automatisering van een leerling in kaart.

### Stap 4: aselechte trekking uit de gereduceerde steekproef

Uit de resulterende set van 13.000 inhoudswoorden is aselekt een groot aantal woorden getrokken (ongeveer 200 woorden voor leeskaart 1, 250 woorden voor leeskaart 2 en 300 woorden voor leeskaart 3). De willekeurige selectie van de woorden is eerst gescreend op geschiktheid, waarbij ouderwetse woorden (bijvoorbeeld belhamel) en ongewenste woorden zoals scheldwoorden zijn verwijderd. Ook woorden die weliswaar aan het frequentiecriterium voldeden, maar waarschijnlijk een lage spreiding hebben over de verschillende bronnen (bijvoorbeeld het woord kattenkop met een frequentie van 16) zijn niet opgenomen. Ten slotte zijn woorden geschrapt die niet in een woordenboek staan (bijvoorbeeld zzoemm), die meer dan vier lettergrepen hebben (en die dus niet voldoen aan de vooraf gedefinieerde woordstructuren, zie paragraaf 3.1 en die qua betekenis te moeilijk zijn voor groep 4 (bijvoorbeeld gedwee).

### Stap 5: verdeling van woorden volgens de kaartstructuur

Voor elke kaart is een kaartstructuur ontworpen. Volgens deze structuur is elke kaart opgebouwd uit 15 units. Een unit is een set van 10 woorden. Elke unit is zo samengesteld dat de verschillende woordstructuren die typerend zijn voor de leeskaart gelijkmatig verdeeld aan de orde komen. De opbouw in units maakt het mogelijk om de prestaties van leerlingen onderling goed te vergelijken.

Bij de vorige uitgave van de DMT (Cito, 2009) stonden de woorden die over het algemeen als makkelijk werden getypeerd vooraan en de woorden die als moeilijk werden ervaren, achteraan de kaart. Zo begon leeskaart 3 bijvoorbeeld met tweelettergrepige woorden, daarna kwamen drie- en ten slotte vierlettergrepige woorden. Snelle lezers kregen daardoor gemiddeld langere en dus moeilijkere woorden te lezen dan langzame lezers. Dit leidt tot een ongewenste extra 'handicap' voor snelle lezers. Het betekent dat 10 goed

gelezen woorden méér aan het begin van de leeskaart (van 40 naar 50 woorden) iets anders inhoudt dan 10 woorden goed gelezen méér aan het einde (van 80 naar 90 woorden). Door te werken met units kan dit probleem worden ondervangen.

In tabel 3.3 is de vooraf door ons beoogde verdeling van woorden over de verschillende woordstructuren weergegeven. De gehanteerde verdeling is representatief voor de verdeling van woordstructuren in het totale BasiLex-corpus. De woorden die na stap 4 waren overgebleven, zijn op basis van hun structuur ingedeeld in vooraf gedefinieerde woordtypes. Daarna is per woordtype het benodigde aantal woorden per unit geselecteerd.

Tabel 3.3 *Beoogde verdeling van woordtypes/woordstructuren per unit voor leeskaart 1, leeskaart 2 en leeskaart 3*

Lees-kaart	Woordtype	Voorbeeld	Aantal per unit	Percentage
1	km en mk <sup>3</sup>	as en fee	1	10
	mkm – korte klinker	vak	3	30
	mkm – lange klank	peen	3	30
	mkm – tweetekenklank	fout	3	30
2	mmk; mmkm	groei/bloem/krab/kring	4	40
	kmm; mkmm	arm/tand/beurt/schelp	3	30
	mmkmm	slang	2	20
	mmmkm; mmmkmm; kmmm; mkmmm; mmkmmm; kmmmm; mkmmmm	strik/strand/eerst/ barst/slechts/ angst/herfst	1	10
3	twee lettergrepen	geluid	5	50
	drie lettergrepen	dromerig	3	30
	vier lettergrepen	keukentafel	2	20

Na controle op voldoende variatie per unit (bijvoorbeeld niet drie woorden met -d in één unit) zijn nog enkele woorden tussen units uitgewisseld. Vervolgens is de volgorde van de woorden binnen een unit aselekt bepaald en is gecontroleerd op ongewenste volgorde-effecten (bijvoorbeeld twee woorden die op elkaar lijken direct na elkaar, of opeenvolgende woorden die een zin zouden kunnen vormen). Ten slotte is de volgorde van de units aselekt bepaald.

#### **Stap 6: controle op verdeling leesmoelijkheden**

Na stap 5 is voor leeskaart 2 en 3 nagegaan of in de verzameling geselecteerde woorden de diverse leesmoelijkheden, zoals weergegeven in tabel 3.4, voldoende aan bod kwamen. Voor beide kaarten is een richtlijn opgesteld voor de verdeling van leesmoelijkheden over de kaarten. Deze luidde als volgt:

- Leeskaart 2: in elke unit dient minstens 1 van de 3 onderscheiden leesmoelijkheden (a, b of c) voor te komen, waarbij de subcategorieën worden afgewisseld in de opeenvolgende units.
- Leeskaart 3: in elke unit dient minstens 1 van de 6 onderscheiden leesmoelijkheden (a tot en met f) voor te komen, waarbij de subcategorieën worden afgewisseld in de opeenvolgende units.

Bij units die niet voldeden aan de bovenstaande richtlijn, zijn woorden omgewisseld.

<sup>3</sup> De helft van de units bevat een km-woord en helft van de units bevat een mk-woord.



Tabel 3.4 Overzicht van leesmoelijkheden voor kaart 2 en 3

Leeskaart	Leesmoelijkheid	Omschrijving (subcategorieën)
1	-	n.v.t.
2	a	woorden met sch-; woorden met -ng/-nk
	b	woorden met -aai/-ooi/-oei ; woorden met -ieuw/-eeuw/-uw
	c	woorden met -d/-b; woorden met -ch(t)
3	a	woorden met een open lettergreep; woorden met een gesloten lettergreep
	b	woorden met een voorvoegsel (be-/ge-/ver-); woorden met een achtervoegsel (-lijk/-ig); meervoudsvormen; verkleinwoorden
	c	samengestelde woorden
	d	woorden met th-; woorden met ch-; woorden met i uitgesproken als ie; woorden met -x(-); woorden met (-)y(-)
	e	woorden met -tie; woorden met -isch; woorden met -ik; woorden met -eren/-elen/-enen; woorden met -eal/-eel/-ieel; woorden met -el/-er/-em
	f	leenwoorden (hoogfrequent)

#### Stap 7: controle op overlap vorige uitgave DMT

Vervolgens is nagegaan hoeveel van de geselecteerde woorden ook in de vorige uitgave van de DMT waren opgenomen. Omdat het aantal beschikbare woorden in de steekproef van het Basilexcorpus dat voldeed aan alle criteria voor bepaalde woordtypen beperkt was, bleek het niet mogelijk om leeskaarten samen te stellen die geen enkele overlap vertoonden met de vorige uitgave. Besloten is om de overlap te beperken hooguit tot 10 à 15 woorden. Bij kaarten waaruit de overlap uit meer woorden bestond, zijn woorden omgewisseld. Na afloop van stap 6 waren de parallelversies A en C van kaart 1, 2 en 3 gereed.

#### Stap 8: samenstelling parallelversie B van kaart 1, 2 en 3

Op basis van parallelversie A is parallelversie B samengesteld. Versie B bevat dezelfde woorden als versie A, maar de volgorde van units en de volgorde van de woorden binnen de units is verschillend. Dat laatste betekent dat de 10 woorden van een unit wel bij elkaar blijven, maar in een andere volgorde staan.

#### 3.2.3 Gerealiseerde leeskaarten

De samenstelling van de gerealiseerde leeskaarten wordt hierna beschreven aan de hand van kenmerken op het niveau van de leeskaarten en het niveau van de units.

De verdeling van de verschillende woordsoorten per leeskaart is weergegeven in tabel 3.5. In deze tabel worden de volgende afkortingen gebruikt: ZN = zelfstandig naamwoord, ADJ = bijvoeglijk naamwoord, VNW = voornaamwoord, TW= telwoord, BW= bijwoord, VZ= voorzetsel, TSW = tussenwerpsel, VG = voegwoord, WW = werkwoord. De tabel laat zien dat de gerealiseerde verdeling overeenkomt met de beoogde verdeling zoals weergegeven in tabel 3.2.

Tabel 3.5 Verdeling van woordsoorten in leeskaart 1, leeskaart 2 en leeskaart 3

Leeskaart	Woordsoorten												totaal
	ZN	% ZN	ADJ	% ADJ	VNW	TW	BW	VZ	TSW	VG	WW	% overige woordsoorten	
1 A/B	111	74%	30	20%	3	2	4	0	0	0	0	6%	150
1 C	113	75%	28	19%	0	4	4	0	1	0	0	6%	150
2 A/B	115	77%	26	17%	1	1	4	3	0	0	0	6%	150
2 C	120	80%	18	12%	1	4	7	0	0	0	0	8%	150
3 A/B	105	70%	30	20%	1	1	5	2	0	1	5	10%	150
3 C	105	70%	32	21%	1	1	4	2	0	1	4	9%	150

In tabel 3.6 is de gerealiseerde verdeling van woordtypes/woordstructuren voor leeskaart 1, leeskaart 2 en leeskaart 3 weergegeven. De verdeling is geldig voor versie A, B en C. De gerealiseerde verdeling per unit komt overeen met de beoogde verdeling zoals weergegeven in tabel 3.3.

Tabel 3.6 Gerealiseerde verdeling van woordtypes/woordstructuren per leeskaart

Leeskaart	Woordtype	Voorbeeld	Aantal woorden per leeskaart
1 A/B/C	km en mk	as en fee	15
	mkm – korte klinker	vak	45
	mkm - lange klank	peen	45
	mkm - tweeteekenklank	fout	45
2 A/B/C	mmk; mmkm	groei/bloem/krab/kring	60
	kmm; mkmm	arm/tand/beurt/schelp	45
	mmkmm	slang	30
	mmmkm; mmmkmm; kmmm; mkmmm; mmkmmm; kmmmm; mkmmmm	strik/strand/eerst/barst/ slechts/angst/herfst	15
3 A/B/C	twee lettergrepen	geluid	75
	drie lettergrepen	dromerig	45
	vier lettergrepen	keukentafel	30

### 3.2.4 Screening en try-out

Nadat de leeskaarten waren samengesteld zijn ze voorgelegd aan een aantal Nederlandse en Vlaamse vak-experts. 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 woorden die werden vervangen door andere overeenkwamen qua woordtype en leesbaarheid.

### 3.2.5 Proeftoetsing

De herziene leeskaarten zijn in een proeftoets voorgelegd aan leerlingen in de eerste twee leerjaren van het primair onderwijs. Periode van afname was het E-moment (mei/juni 2015). De leeskaarten zijn niet de hogere groepen afgenomen, omdat eventuele struikelblokken als onverwacht vaak fout gelezen woorden het beste in de laagste leerjaren opgespoord konden worden.

Figuur 3.1 geeft gedetailleerd weer hoe het proeftoetsdesign er voor jaargroep 3 en 4 uitzag. De verschillende versies van de DMT leeskaarten zijn volgens een “balanced incomplete block” principe aan subgroepen toegewezen. Dit betekent dat alle versies even vaak, in alle mogelijke combinaties, zijn afgenomen.

Figuur 3.1 Proeftoetsdesign jaargroep 3 en 4

Subgroep	DMT1A	DMT2A	DMT3A	DMT1B	DMT2B	DMT3B	DMT1C	DMT2C	DMT3C
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									

Er zijn in totaal 270 leerlingen geworven per jaargroep (10 leerlingen × 27 subgroepen), zodat van elke leeskaart van de DMT minstens 90 waarnemingen beschikbaar waren per jaargroep (180 in totaal). Deze doelstelling is ruimschoots gerealiseerd. Tabel 3.7 laat zien hoe vaak elke leeskaart is gemaakt.

Tabel 3.7 Aantal leerlingen in de proeftoets uitgesplitst naar afnamemoment en leeskaart

Afname	DMT1A	DMT1B	DMT1C	DMT2A	DMT2B	DMT2C	DMT3A	DMT3B	DMT3C	Totaal aantal waarne- mingen
E3	107	98	98	101	100	102	101	101	101	
E4	107	93	96	103	97	96	99	99	98	
Totaal	214	191	194	204	197	198	200	200	199	1797

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 samenstelling van de leeskaarten licht gewijzigd. Zo is op leeskaart 2C het woord 'bloot' bijvoorbeeld vervangen door het woord 'bril', omdat het eerste woord volgens de toetsleiders tot veel hilariteit leidde. Daarnaast zijn bijvoorbeeld de woorden 'tuinhek' en 'magisch' die respectievelijk op leeskaart A/B en C stonden, omgewisseld, zodat de moeilijkheidsgraad van beide versies beter in evenwicht was. Net zoals bij de try-out is ervoor gezorgd dat woorden die werden vervangen door andere overeenkwamen voor wat betreft woordtype en leesmoelijkheden.

## 4 Kalibratie en normering

### 4.1 Dataverzameling

In schooljaar 2015/2016 is een landelijk normeringsonderzoek in het reguliere basisonderwijs uitgevoerd ten behoeve van de normering van de DMT. 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 (medio afname) en in mei/juni (einde afname). De leerlingen in groep 8 zijn alleen in januari/februari getoetst.

Er is naar gestreefd om bij de medio en einde afname in een jaargroep dezelfde leerlingen te toetsen, zodat de voortgang van deze leerlingen in beeld kon worden gebracht. Bij de toetsafname kregen leerlingen drie leeskaarten van de DMT voorgelegd, met uitzondering van de leerlingen in groep 3. Zij lazen bij de medio afname alleen kaart 1 en 2, omdat de meeste leerlingen halverwege groep 3 nog niet toe zijn aan het lezen van meerlettergrepige woorden. Daarnaast zijn met het oog op de validering van de DMT enkele andere toetsen afgenomen. Het ging om de Eén-Minuut-Test (Pearson, 1999), de AVI-toets (Cito, 2017) en een nieuw geconstrueerde leeskaart K4 (zie voor toelichting paragraaf 4.4) met 50 woorden van het type KM, MK en MKM, 50 woorden van het type MMK, MMKM, MKMM, MMKMM, MMMKM, MKMMM en MKMMMM, en 50 woorden met twee, drie of vier lettergrepen. De woordtypen werden door elkaar aangeboden en waren afkomstig uit de (nieuw ontwikkelde) DMT. Figuur 4.1 geeft gedetailleerd weer hoe het design eruit zag voor jaargroep 4 tot en met 7. Het design voor jaargroep 3 en 8 was gebaseerd op onderstaand design, maar week op enkele punten af. Voor jaargroep 3 gold dat alle versies van kaart 3 zijn weggelaten bij de medio-afname. Voor jaargroep 8 gold dat de eindafname kwam te vervallen.

Figuur 4.1 Onderzoeksdisegn jaargroepen 3 tot en met 8

Subgroep	medio											eind											
	DMT1A	DMT2A	DMT3A <sup>1</sup>	DMT1B	DMT2B	DMT3B <sup>1</sup>	DMT1C	DMT2C	DMT3C <sup>1</sup>	K4	AVI	EMT	DMT1A	DMT2A	DMT3A	DMT1B	DMT2B	DMT3B	DMT1C	DMT2C	DMT3C	K4	AVI
1																							
2																							
3																							
4																							
5																							
6																							
7																							
8																							
9																							
10																							
11																							
12																							
13																							
14																							
15																							
16																							
17																							
18																							
19																							
20																							
21																							
22																							
23																							
24																							
25																							
26																							
27																							

<sup>1</sup> Deze leeskaart is bij de medio afname niet in groep 3 voorgelegd.

In figuur 4.1 is te zien dat elke jaargroep tijdens het normeringsonderzoek is opgedeeld in 27 subgroepen. Alle subgroepen maakten de AVI-toets (Cito, 2017) en K4 en één op de zes leerlingen maakte de EMT (Brus en Voeten, 1979). Daarnaast kregen de leerlingen drie leeskaarten van de DMT voorgelegd. De verschillende versies van de DMT leeskaarten zijn daarbij volgens een “balanced incomplete block” principe aan subgroepen toegewezen. Dit betekent dat alle versies even vaak, in alle mogelijke combinaties, zijn afgenomen, en dat leerlingen bij de tweede afname (einde) nooit dezelfde versie maakten als bij de eerste afname (medio). 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, aan de materialen en natuurlijk aan de toetsprocedure, zodat iedereen de toetsen op dezelfde manier zou afnemen. Bij beide afnames waren ongeveer 45 toetsleiders actief.

Bij het normeringsonderzoek is er naar gestreefd om elke subgroep te laten bestaan uit 20 leerlingen. Per afnamemoment zouden dan 540 leerlingen deelnemen (20 leerlingen  $\times$  27 subgroepen) en in totaal zou elke leeskaart van de DMT 1980 keer worden afgenomen (9 subgroepen  $\times$  20 leerlingen  $\times$  11 afname-momenten (M3 t/m M8)). Deze doelstelling is ruimschoots gerealiseerd. Tabel 4.1 laat zien hoeveel leerlingen bij elke afname hebben deelgenomen en hoe vaak elke leeskaart is gemaakt. Bij alle afname-momenten is het gerealiseerde leerlingaantal groter dan het streefaantal van 540 en ook het aantal waarnemingen per leeskaart is steeds groter dan het streefaantal van 1980.

Tabel 4.1 Aantal leerlingen in de steekproef uitgesplitst naar afnamemoment en leeskaart

Afname	DMT1A	DMT1B	DMT1C	DMT2A	DMT2B	DMT2C	DMT3A	DMT3B	DMT3C	Totaal
m3	193	187	184	187	190	187	---	---	---	564
e3	211	227	217	222	215	218	220	216	219	655
m4	182	174	176	175	179	178	169	179	184	532
e4	258	283	279	271	277	272	272	273	275	820
m5	193	194	198	197	195	193	195	194	196	585
e5	241	241	238	240	242	238	241	239	240	720
m6	200	191	185	192	192	192	194	194	188	576
e6	250	244	242	243	247	246	246	243	247	736
m7	200	196	194	201	196	193	195	197	198	590
e7	221	225	225	220	224	227	226	223	222	671
m8	200	204	200	202	201	201	199	202	203	604
Totaal	2349	2366	2338	2350	2358	2345	2157	2160	2172	7053

## 4.2 Representativiteit van de steekproef

In deze paragraaf besteden we aandacht aan de kenmerken die relevant zijn om de relatie tussen de normeringssteekproeven en de populatie in kaart te brengen. Tijdens de dataverzameling is informatie verzameld over *regio*, *urbanisatiegraad*, *schooltype* en *seks*. De verschillende 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 één van de ouders of beide ouders heeft of hebben een opleiding gehad uit categorie 3
  - 0.3 beide ouders of de ouder die belast is met de dagelijkse verzorging heeft of hebben een opleiding uit categorie 2 gehad
  - 1.2 één van de ouders heeft een opleiding gehad uit categorie 1 en de ander een opleiding uit categorie 1 óf 2.

In 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](http://data.duo.nl)) en CBS ([statline.cbs.nl](http://statline.cbs.nl)).

De representativiteitsanalyse bestond uit twee 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  $\chi^2$ -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  $\varphi$  van de  $\chi^2$ -toetsing. Een  $\varphi$  van .10 kan aangemerkt worden als een klein effect, een  $\varphi$  van .30 als een gemiddeld effect, en een  $\varphi$  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 was de toetsscore (totaalscore over alle afgenomen leeskaarten) de afhankelijke variabele en waren 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 564 leerlingen. Tabel 4.1 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 achterstandsleerlingen. Van deze variabelen is regio echter de enige die er echt toe doet bij deze normgroep. De noordelijke provincies presteren beter 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 ( $\varphi = .08$ ). Bovendien scoren jongens niet significant hoger of lager dan meisjes op de DMT ( $z = -1.00$ ,  $d = -0.09$ ).

Tabel 4.1 Resultaten representativiteits- en regressieanalyse medio groep 3

Variabele	Categorie	Percentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	54	9.57	6.98	3	0.07	0.11	---	---
	Oost	22.23	110	19.50					-2.57	-0.44
	West	48.22	266	47.16					-2.67	-0.42
	Zuid	19.63	134	23.76					-2.16	-0.37
Urbanisatie	Platteland	54.29	361	64.01	21.46	1	0.00	0.20	---	---
	Stad	45.71	203	35.99					0.62	0.06
Schooltype	[0, .15)	81.23	498	88.30	18.48	1	0.00	0.18	---	---
	[.15, 1]	18.77	66	11.70					-1.85	-0.25
Sekse	Jongen	50.46	252	46.67	3.11	1	0.08	0.08	---	---
	Meisje	49.54	288	53.33					-1.00	-0.09

### Representativiteit normgroep 2: eind groep 3

Voor afnamemoment *einde 3* bestond de steekproef uit 655 leerlingen. In tabel 4.2 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 achterstandsleerlingen. 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. Bij deze normgroep scoren jongens hoger dan meisjes ( $z = -2.94$ ,  $d = -0.25$ ) en daarnaast is er sprake van een prestatieverschil tussen de scholen met relatief veel en relatief weinig achterstandsleerlingen ( $z = -3.78$ ,  $d = -0.64$ ).

Tabel 4.2 Resultaten representativiteits- en regressieanalyse eind groep 3

Variabele	Categorie	Percentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	44	6.72	14.88	3	0.00	0.15	---	---
	Oost	22.23	173	26.41					0.12	0.02
	West	48.22	295	45.04					0.08	0.01
	Zuid	19.63	143	21.83					-1.08	-0.20
Urbanisatie	Platteland	54.29	512	78.17	150.49	1	0.00	0.48	---	---
	Stad	45.71	143	21.83					-0.36	-0.04
Schooltype	[0, .15)	81.23	594	90.69	38.42	1	0.00	0.24	---	---
	[.15, 1]	18.77	61	9.31					-3.78	-0.64
Sekse	Jongen	50.46	268	46.69	3.26	1	0.07	0.08	---	---
	Meisje	49.54	306	53.31					-2.94	-0.25



### Representativiteit normgroep 3: medio groep 4

Voor afnamemoment *medio 4* bestond de steekproef uit 532 leerlingen. In tabel 4.3 is te zien dat de verdeling van leerlingen over de onderscheiden groepen grote gelijkenis 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. De scholen met relatief veel achterstandsleerlingen zijn representatief vertegenwoordigd in de normeringssteekproef. Prestatieverschillen tussen leerlingen van deze normgroep kunnen we enigszins verklaren vanuit de variabele regio.

Tabel 4.3 Resultaten representativiteits- en regressieanalyse medio groep 4

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	35	6.58	19.79	3	0.00	0.19	---	---
	Oost	22.23	115	21.62					1.22	0.24
	West	48.22	241	45.30					2.19	0.41
	Zuid	19.63	141	26.50					1.38	0.27
Urbanisatie	Platteland	54.29	334	62.78	15.46	1	0.00	0.17	---	---
	Stad	45.71	198	37.22					-1.80	-0.18
Schooltype	[0, .15)	81.23	448	84.21	3.10	1	0.08	0.08	---	---
	[.15, 1]	18.77	84	15.79					1.24	0.15
Sekse	Jongen	50.46	258	50.99	0.06	1	0.81	0.01	---	---
	Meisje	49.54	248	49.01					-0.49	-0.04

### Representativiteit normgroep 4: einde groep 4

Voor afnamemoment *einde 4* bestond de steekproef uit 820 leerlingen. In tabel 4.4 is te zien dat plattelandsgebieden, de zuidelijke provincies en de scholen met relatief weinig achterstandsleerlingen 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 bij deze normgroep echter gering. Jongens 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 ( $z = -3,40$ ,  $d = -0,39$ ).

Tabel 4.4 Resultaten representativiteits- en regressieanalyse eind groep 4

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	54	6.59	82.85	3	0.00	0.32	---	---
	Oost	22.23	214	26.10					-0.83	-0.13
	West	48.22	304	37.07					0.24	0.04
	Zuid	19.63	248	30.24					0.68	0.12
Urbanisatie	Platteland	54.29	582	70.98	92.00	1	0.00	0.34	---	---
	Stad	45.71	238	29.02					-3.40	-0.39
Schooltype	[0, .15)	81.23	742	90.49	46.10	1	0.00	0.24	---	---
	[.15, 1]	18.77	78	9.51					1.87	0.29
Sekse	Jongen	50.46	360	49.59	0.22	1	0.64	0.02	---	---
	Meisje	49.54	366	50.41					-0.07	-0.01

#### Representativiteit normgroep 5: medio groep 5

Voor afnamemoment *medio 5* bestond de steekproef uit 585 leerlingen. In tabel 4.5 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 achterstandsl leerlingen is in balans met de aantallen die we vinden in de populatie. Prestatieverschillen tussen leerlingen van deze normgroep 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 4.5 Resultaten representativiteits- en regressieanalyse medio groep 5

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	49	8.38	32.39	3	0.00	0.24	---	---
	Oost	22.23	96	16.41					0.17	0.03
	West	48.22	275	47.01					0.35	0.06
	Zuid	19.63	165	28.21					0.00	0.00
Urbanisatie	Platteland	54.29	379	64.79	25.97	1	0.00	0.21	---	---
	Stad	45.71	206	35.21					0.87	0.08
Schooltype	[0, .15)	81.23	470	80.34	0.30	1	0.58	0.02	---	---
	[.15, 1]	18.77	115	19.66					-1.11	-0.12
Sekse	Jongen	50.46	251	45.31	5.89	1	0.02	0.10	---	---
	Meisje	49.54	303	54.69					-1.40	-0.12

### Representativiteit normgroep 6: einde groep 5

Voor afnamemoment *einde 5* bestond de steekproef uit 720 leerlingen. In tabel 4.6 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 van deze normgroep. 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 achterstandsleerlingen nu iets uit balans ten opzichte van de populatie.

Tabel 4.6 Resultaten representativiteits- en regressieanalyse eind groep 5

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Percentuele verdeling in de steekproef	$\chi^2$	$df$	$p$	$\phi$	$z$	$d$
Regio	Noord	9.91	56	7.78	130.39	3	0.00	0.43	---	---
	Oost	22.23	149	20.69					-1.15	-0.18
	West	48.22	254	35.28					-0.77	-0.12
	Zuid	19.63	261	36.25					-1.11	-0.18
Urbanisatie	Platteland	54.29	469	65.14	34.15	1	0.00	0.22	---	---
	Stad	45.71	251	34.86					-0.35	-0.04
Schooltype	[0, .15)	81.23	632	87.78	20.25	1	0.00	0.17	---	---
	[.15, 1]	18.77	88	12.22					-1.24	-0.17
Sekse	Jongen	50.46	339	48.99	0.60	1	0.44	0.03	---	---
	Meisje	49.54	353	51.01					0.05	0.00

### Representativiteit normgroep 7: medio groep 6

Voor afnamemoment *medio 6* bestond de steekproef uit 576 leerlingen. In tabel 4.7 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 en schooltype. De plattelandsgebieden en de scholen met relatief weinig achterstandsleerlingen waren oververtegenwoordigd in de steekproef. De variabele regio is een voorspeller van de toetsscore bij deze normgroep. De overige variabelen hebben geen effect op de prestaties van leerlingen.

Tabel 4.7 Resultaten representativiteits- en regressieanalyse medio groep 6

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	45	7.81	5.73	3	0.13	0.10	---	---
	Oost	22.23	117	20.31					1.59	0.28
	West	48.22	286	49.65					2.09	0.35
	Zuid	19.63	128	22.22					1.69	0.31
Urbanisatie	Platteland	54.29	357	61.98	13.72	1	0.00	0.15	---	---
	Stad	45.71	219	38.02					0.71	0.07
Schooltype	[0, .15)	81.23	508	88.19	18.32	1	0.00	0.18	---	---
	[.15, 1]	18.77	68	11.81					0.86	0.11
Sekse	Jongen	50.46	278	50.45	0.00	1	1.00	0.00	---	---
	Meisje	49.54	273	49.55					0.90	0.08

#### Representativiteit normgroep 8: einde groep 6

Voor afnamemoment *einde 6* bestond de steekproef uit 736 leerlingen. In tabel 4.8 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 deze normgroep kunnen de prestaties van leerlingen voorspeld worden vanuit urbanisatiegraad; leerlingen in verstedelijkte gebieden presteren zwakker dan leerlingen op het platteland ( $z = -2,06$ ,  $d = -0,21$ )

Tabel 4.8 Resultaten representativiteits- en regressieanalyse eind groep 6

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	df	p	$\varphi$	z	d
Regio	Noord	9.91	52	7.07	89.40	3	0.00	0.35	---	---
	Oost	22.23	157	21.33					-0.43	-0.07
	West	48.22	283	38.45					1.13	0.18
	Zuid	19.63	244	33.15					0.46	0.08
Urbanisatie	Platteland	54.29	448	60.87	12.84	1	0.00	0.13	---	---
	Stad	45.71	288	39.13					-2.06	-0.21
Schooltype	[0, .15)	81.23	640	86.96	15.83	1	0.00	0.15	---	---
	[.15, 1]	18.77	96	13.04					0.64	0.09
Sekse	Jongen	50.46	348	50.51	0.00	1	0.98	0.00	---	---
	Meisje	49.54	341	49.49					0.63	0.05

### Representativiteit normgroep 9: medio groep 7

Voor afnamemoment *medio* 7 bestond de steekproef uit 590 leerlingen. De uitkomsten van de representativiteitsanalyse zijn te vinden in tabel 4.9. 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  $\phi$ . 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 van deze normgroep. Bij de overige variabelen bestaan er geen significante verschillen tussen de onderscheiden groepen leerlingen.

Tabel 4.9 Resultaten representativiteits- en regressieanalyse medio groep 7

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	df	p	$\phi$	z	d
Regio	Noord	9.91	42	7.12	34.51	3	0.00	0.24	---	---
	Oost	22.23	89	15.08					2.36	0.45
	West	48.22	301	51.02					1.77	0.30
	Zuid	19.63	158	26.78					3.05	0.55
Urbanisatie	Platteland	54.29	362	61.36	11.87	1	0.00	0.14	---	---
	Stad	45.71	228	38.64					-0.66	-0.06
Schooltype	[0, .15)	81.23	517	87.63	15.84	1	0.00	0.16	---	---
	[.15, 1]	18.77	73	12.37					-0.56	-0.07
Sekse	Jongen	50.46	266	47.58	1.85	1	0.17	0.06	---	---
	Meisje	49.54	293	52.42					-0.43	-0.04

### Representativiteit normgroep 10: einde groep 7

Voor afnamemoment *einde* 7 bestond de steekproef uit 671 leerlingen. Tabel 4.10 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  $\phi$  wijst op gemiddelde effecten. Vrijwel geen van de variabelen blijkt van invloed op de prestaties van leerlingen van deze normgroep. Alleen urbanisatiegraad is een uitzondering. Leerlingen in verstedelijkte gebieden scoren significant lager dan de leerlingen op het platteland ( $z = -2.44$ ;  $d = -0.25$ ).

Tabel 4.10 Resultaten representativiteits- en regressieanalyse eind groep 7

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	$df$	$p$	$\varphi$	$z$	$d$
Regio	Noord	9.91	32	4.77	85.00	3	0.00	0.36	---	---
	Oost	22.23	137	20.42					-0.70	-0.14
	West	48.22	281	41.88					0.03	0.01
	Zuid	19.63	221	32.94					1.33	0.27
Urbanisatie	Platteland	54.29	428	63.79	24.38	1	0.00	0.19	---	---
	Stad	45.71	243	36.21					-2.44	-0.25
Schooltype	[0, .15)	81.23	569	84.80	5.61	1	0.02	0.09	---	---
	[.15, 1]	18.77	102	15.20					1.08	0.15
Sekse	Jongen	50.46	307	49.52	0.22	1	0.64	0.02	---	---
	Meisje	49.54	313	50.48					-1.37	-0.11

#### Representativiteit normgroep 11: medio groep 8

Voor afnamemoment *medio 8* bestond de steekproef uit 604 leerlingen. Tabel 4.11 laat de uitkomsten van de representativiteitsanalyse zien. Er is sprake van een grote gelijkenis 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. Net als bij normgroep 10 bestaat er bij normgroep 11 ook geen verband tussen de achtergrondvariabelen en de prestaties van leerlingen. De eerder gevonden uitzondering bij urbanisatiegraad vinden we op afnamemoment *medio 8* niet terug.

Tabel 4.11 Resultaten representativiteits- en regressieanalyse medio groep 8

Variabele	Categorie	Procentuele verdeling in de populatie	N steekproef	Procentuele verdeling in de steekproef	$\chi^2$	$df$	$p$	$\varphi$	$z$	$d$
Regio	Noord	9.91	45	7.45	20.69	3	0.00	0.19	---	---
	Oost	22.23	108	17.88					1.36	0.25
	West	48.22	295	48.84					1.00	0.17
	Zuid	19.63	156	25.83					1.02	0.18
Urbanisatie	Platteland	54.29	384	63.58	20.99	1	0.00	0.19	---	---
	Stad	45.71	220	36.42					0.16	0.02
Schooltype	[0, .15)	81.23	508	84.11	3.28	1	0.07	0.07	---	---
	[.15, 1]	18.77	96	15.89					-0.51	-0.06
Sekse	Jongen	50.46	272	47.89	1.50	1	0.22	0.05	---	---
	Meisje	49.54	296	52.11					-0.85	-0.07

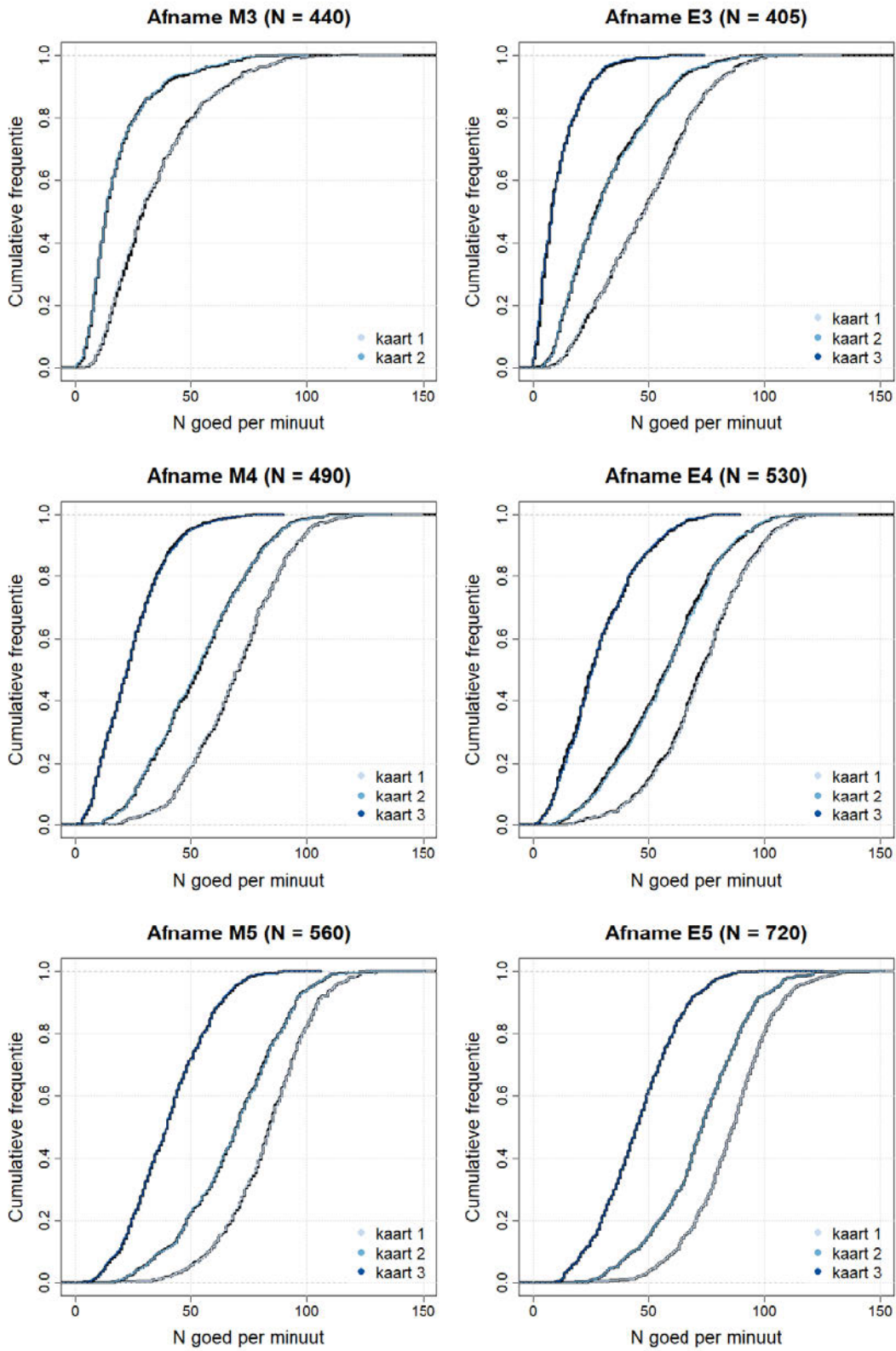
Samenvattend leveren de representativiteitsanalyses het volgende beeld op:

De normeringssteekproeven laten een grote onderlinge gelijkenis 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, 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.

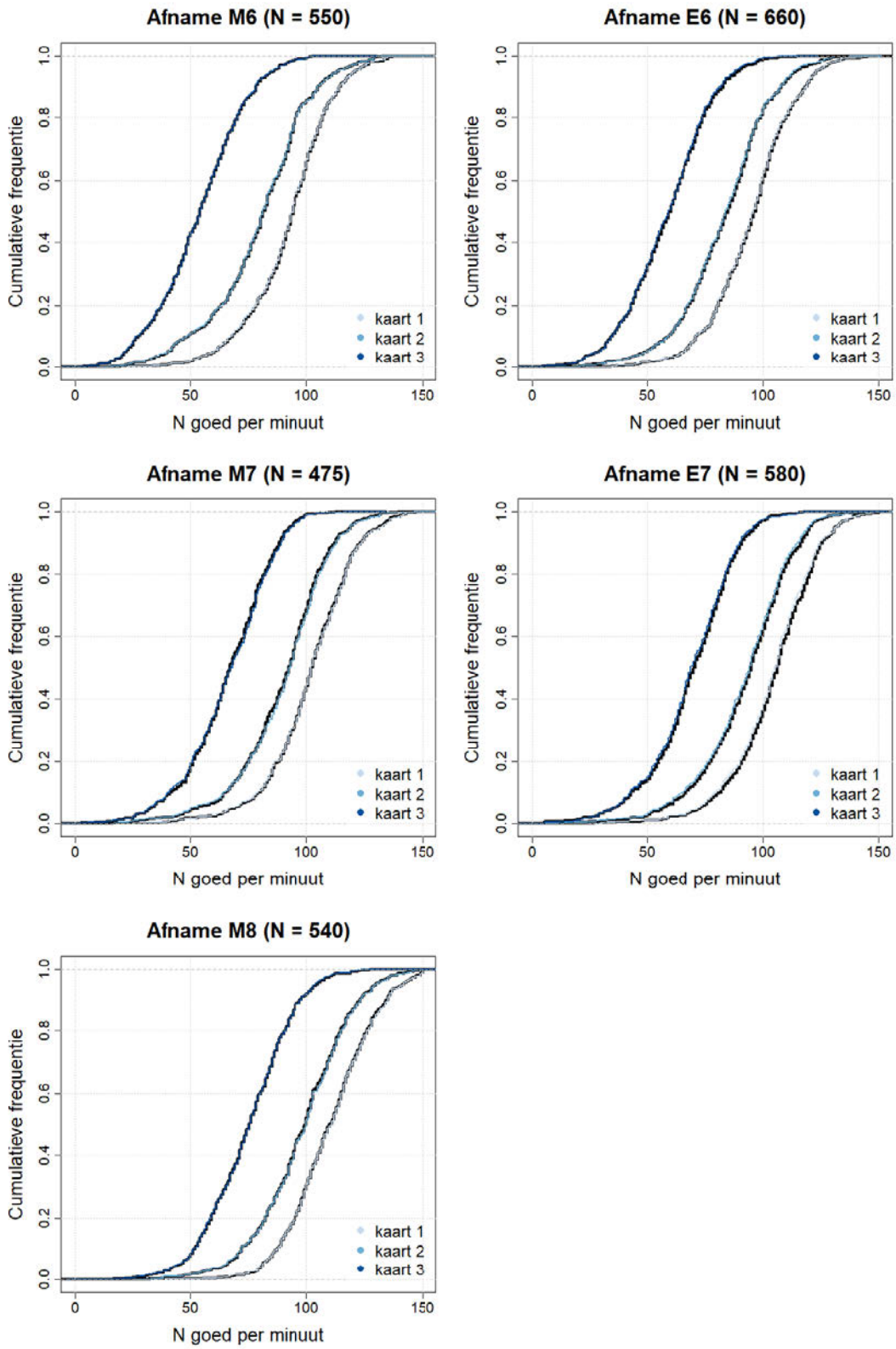
Op grond van de uitkomsten van de representativiteitsanalyses is het twijfelachtig of we op basis van de beschikbare gegevens een landelijk representatieve normering op kunnen stellen voor de DMT. Een onder- of overrepresentatie van een bepaalde groep leerlingen is niet altijd problematisch, maar als de prestaties van de leerlingen in die groep verschillen van de prestaties van de leerlingen in andere groepen kan dat leiden tot *bias* in de normen. In geval van de DMT is er af en toe sprake van een onder- of overrepresentatie én een betekenisvol prestatieverschil. Daarom is besloten om via *sampling* per afnamemoment een representatieve steekproef te trekken uit de beschikbare data ( $N$  minimaal 400) en de scoreverdelingen van de *subsamples* te vergelijken met de scoreverdelingen van de volledige steekproeven. Figuur 4.2 laat per afnamemoment de cumulatieve scoreverdelingen zien van de volledige steekproeven (zwarte lijnen) en de *subsamples* (blauwe lijnen) per leeskaart. De zwarte en de blauwe lijnen overlappen volledig, waardoor in de afbeelding geen zes lijnen te zien zijn (drie zwarte voor kaart 1 t/m 3 en 3 blauwe voor kaart 1 t/m 3) maar 3 blauw-zwarte lijnen. Dit betekent dat het voor de normering weinig uitmaakt of we ons baseren op de volledige steekproeven of de *subsamples*. Daarom is ervoor gekozen om bij de normering alle data te gebruiken.

Figuur 4.2 *Vergelijking tussen de cumulatieve scoreverdelingen van de volledige steekproeven en de subsamples*





Figuur 4.1 Vervolg



### 4.3 Beschrijving van het meetmodel

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

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

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  $\lambda$  afhankelijk is van de persoon en van de taak. Toegepast op de DMT leidt dit tot het volgende model:

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

waarin  $\lambda_{vj}$  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  $\tau_j$  is geen parameter maar een bekende constante die de omvang van de taak uitdrukt. Bij de DMT is dit de toegestane tijd (1 minuut per leeskaart). De parameter  $\sigma_j$  drukt de moeilijkheid van de taak uit: hoe groter deze parameter hoe makkelijker de taak, en de grootte  $\theta_v$  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 ( $\theta_v$ ) die gebaseerd is op de observatie die we van deze leerling hebben wanneer hij leeskaart  $j$  leest: de score  $s_{vj}$ .

De kansverdeling van de scores is afhankelijk van de Poisson-parameter  $\lambda$ , die op zijn beurt weer samengesteld is uit drie componenten. Als de drie componenten gegeven zijn ligt hun product vast, maar als het product gegeven is liggen de drie componenten niet uniek vast; we kunnen namelijk altijd een component delen door een positieve constante  $c$  en een van de twee 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 twee *normalisaties* invoeren. Voor de DMT hebben we dit als volgt gedaan:

- 1 We kiezen een eenheid van tijd. Voor de DMT is deze eenheid de minuut. Omdat de toegestane leestijd voor alle leeskaarten een minuut is, geldt in deze toepassing dus dat  $\tau_j = 1$  voor alle taken.
- 2 De DMT bestaat uit drie verschillende leeskaarten en voor elke leeskaart moet een  $\sigma$ -parameter worden geschat. Hierbij is de restrictie opgelegd dat het product van deze parameters gelijk moet zijn aan 1.

Een echte of denkbeeldige leeskaart waarvan de  $\sigma$ -parameter gelijk is aan 1, noemen we een *standaardkaart*. De interpretatie van de vaardigheid  $\theta_v$  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  $\lambda$ . Als leerling  $v$  leeskaart  $j$  leest, dan is zijn score  $S_{vj}$  een toevalsvariabele, en door (1) en (2) te combineren krijgen we dus

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

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

- $S_{vj}$  is een aantal correct gelezen woorden; zijn verwachte waarde is dus ook een aantal correct gelezen woorden.
- $\tau_j$  is de toegestane leestijd uitgedrukt in minuten.
- $\sigma_j$  vatten we op als een onbenoemd getal, een soort correctiefactor die controleert voor de verschillende moeilijkheid van de woorden op de leeskaarten. De standaardkaart heeft per definitie een  $\sigma$ -waarde gelijk aan 1.

- Om het product van de rechterzijde in (3) in dezelfde eenheid uit te drukken als de linkerzijde, moeten we  $\theta_v$  benoemen als *het aantal correct gelezen woorden per tijdseenheid (minuut) op een standaardkaart*.

### Lokale stochastische onafhankelijkheid

Wat tot hiertoe beschreven is, is de modellering van de uitkomsten wanneer een leerling één enkele leeskaart 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  $\theta_v$  blijft onveranderd bij het lezen van meerdere kaarten op één en hetzelfde afnamemoment.
- 2 Bij elke leeskaart is de (Poisson-)verdeling van de score alleen afhankelijk van de eigenschappen van de kaart en de leestijd (de  $\sigma$ - en de  $\tau$ -parameter) en van de vaardigheid ( $\theta_v$ ). En niet van de score die de leerling op een van de andere leeskaarten 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 leeskaart 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 normeringsonderzoek in totaal  $k$  leeskaarten zijn gebruikt, en dat elke leerling een aantal van deze leeskaarten 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 leeskaarten. 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  $\delta_v$  een behoorlijk complexe structuur heeft: hij is afhankelijk van het dataverzamelingsdesign (de specifieke kaarten die de leerling heeft gelezen), van de toegestane leestijd op deze kaarten (die in het geval van DMT wel constant is, maar in het algemeen kan variëren over de taken) en van de onbekende moeilijkheidsparameters  $\sigma$  (zie formule (4)).

### Schatting van de individuele vaardigheid $\theta_v$

De eerste stap is het schatten van de taakparameters  $\sigma_i$ . 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  $\delta_v$  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

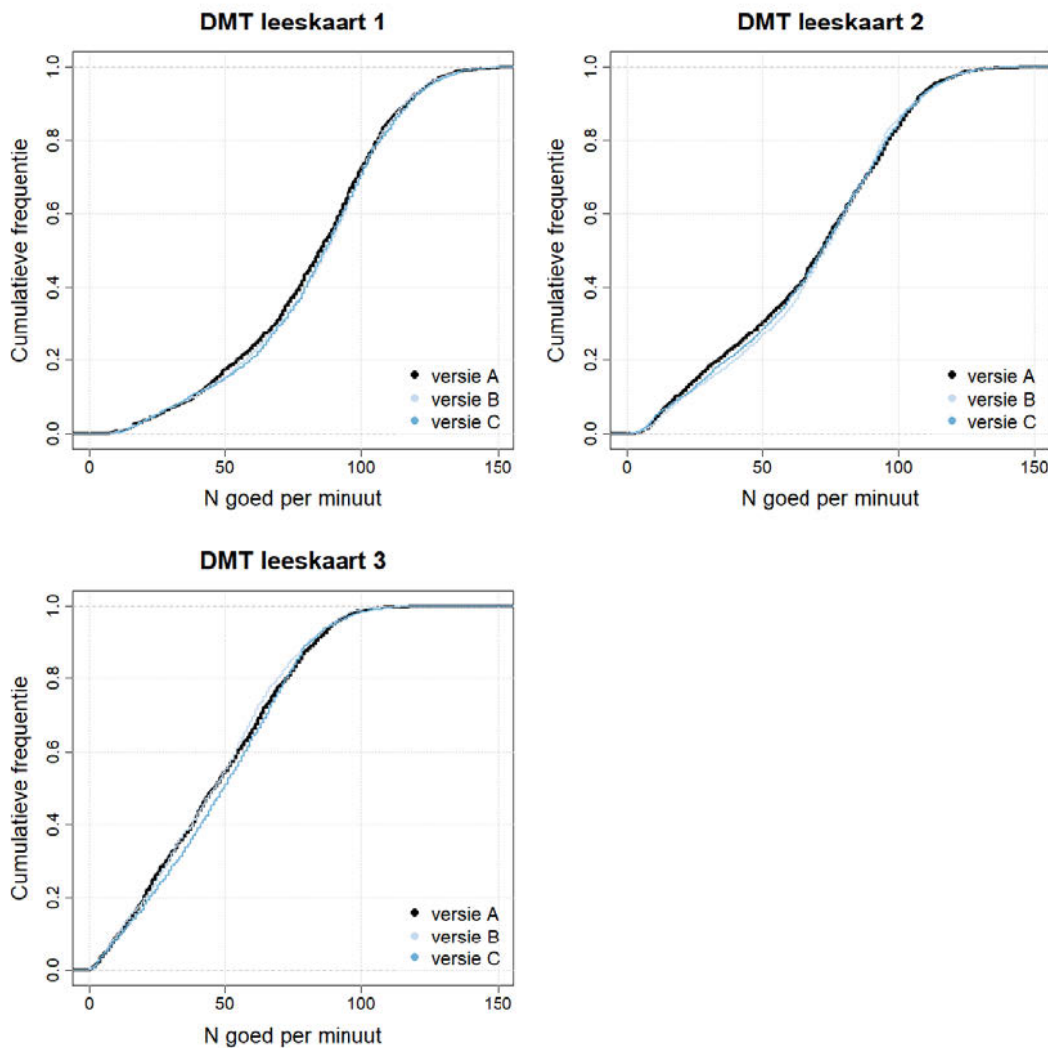
$$(7) \quad SE(\hat{\theta}_v) = \frac{\sqrt{s_v}}{\delta_v}$$

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

#### 4.4 Kalibratie

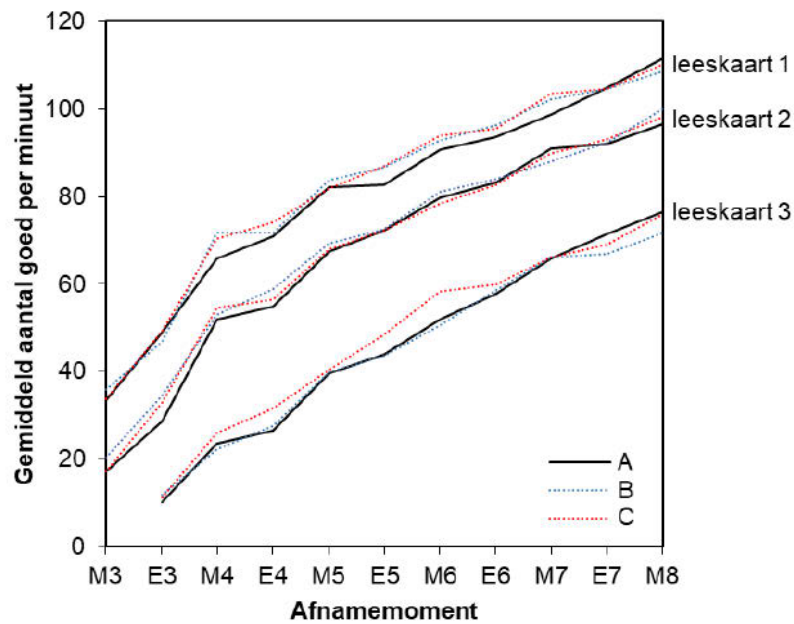
Voorafgaand aan het schatten van het Poisson-model is de paralleliteit van de leeskaarten onderzocht. Vanuit de inhoud bekeken is er geen reden om aan te nemen dat de leeskaartversies verschillende psychometrische eigenschappen hebben, omdat ze zijn samengesteld volgens hetzelfde stappenplan (zie paragraaf 3.2.2). Desalniettemin is een empirische controle belangrijk. De cumulatieve scoreverdelingen van de totale steekproef op de verschillende leeskaartversies zijn met elkaar vergeleken en daarnaast is (als extra controle) per afnamemoment gekeken in hoeverre er sprake is van gelijke gemiddelden. In figuur 4.3 zijn de cumulatieve scoreverdelingen te vinden.

Figuur 4.3 Vergelijking tussen de cumulatieve scoreverdelingen van de volledige steekproef (medio 3 tot en met medio 8) op de verschillende leeskaartversies



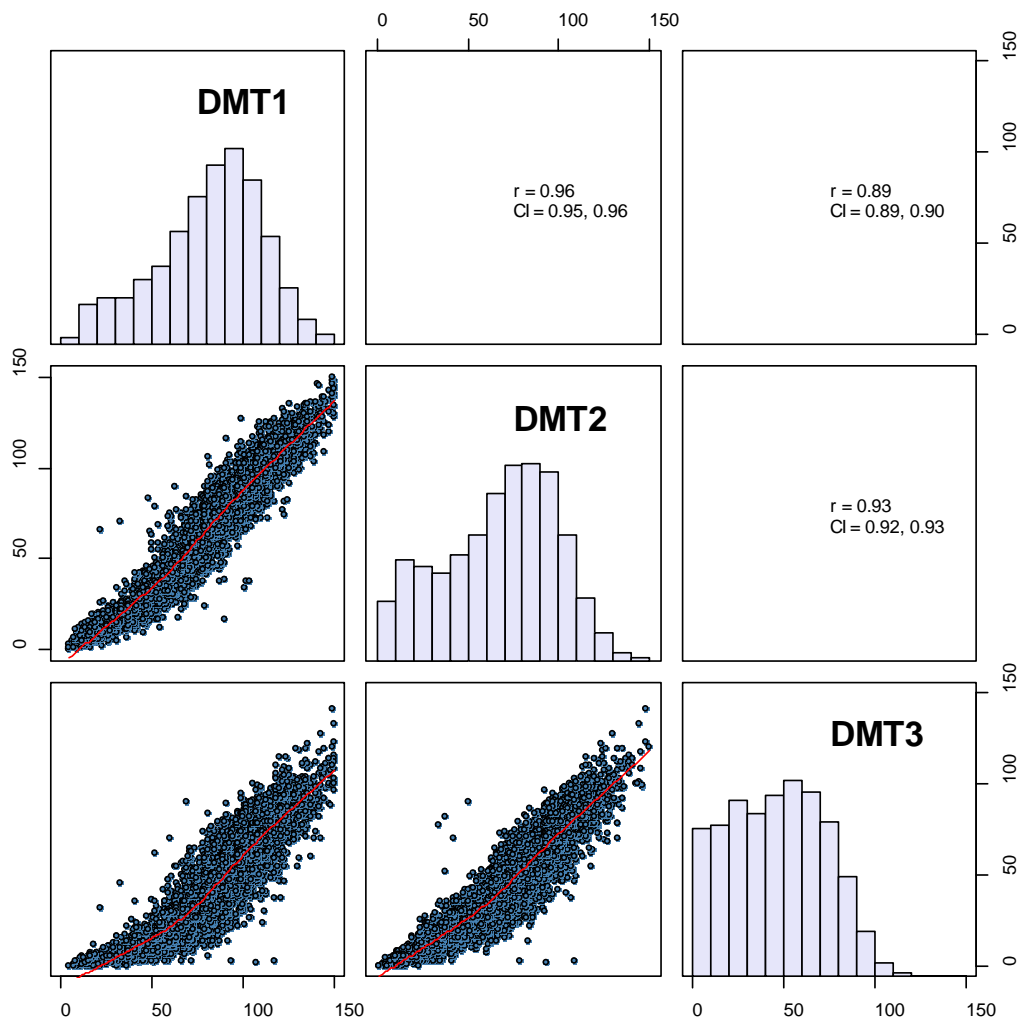
We zien dat de cumulatieve scoreverdelingen nagenoeg gelijk zijn. Dit betekent dat de leerlingen in de steekproef vergelijkbaar presteren op de verschillende versies van de leeskaarten. Figuur 4.4 bevestigt dit beeld. De leerlingen scoren op de verschillende afnamemomenten gemiddeld gezien vergelijkbaar; het gemiddelde gestandaardiseerde verschil (Cohen's  $d$ ) is gelijk aan .04. De kleine prestatieverschillen op de leeskaartversies zijn hoogstwaarschijnlijk dus toe te schrijven aan de steekproef of de metingen en niet het gevolg van verschillen in psychometrische eigenschappen. In de kalibratie is dan ook aangenomen dat de verschillende versies van de leeskaarten parallel zijn.

Figuur 4.4 Vergelijking gemiddelde score op de verschillende leeskaartversies uitgesplitst naar afnamemoment



In een vervolgstap is nagegaan hoe de prestaties van leerlingen op de verschillende leeskaarten met elkaar samenhangen. Het meetmodel veronderstelt dat de prestaties op leeskaart X lineair samenhangen met de prestaties op de leeskaart Y. In figuur 4.5 is te zien dat deze aanname in geval van de DMT niet helemaal opgaat; er is sprake van een knik. Leerlingen met een lage score op leeskaart 1 scoren in vergelijking met leerlingen met een hogere score (onder aanname van lineariteit) extreem laag op leeskaarten 2 en 3. Daarmee is het twijfelachtig of een lineair model een verdedigbare beschrijving kan geven van de prestaties van leerlingen op de DMT. Het is goed mogelijk dat we de prestaties van bepaalde leerlingen bij toepassing van een dergelijk model systematisch onder- of overschatten. Om deze reden is hier bij het schatten van het meetmodel zoals dat beschreven is in paragraaf 4.3 nadrukkelijk op gelet. Merk op dat de correlaties tussen de leeskaarten wel zeer hoog zijn, variërend van .89 tussen leeskaart 1 en 3 tot .96 tussen leeskaart 1 en 2.

Figuur 4.5 Samenhang tussen de prestaties van leerlingen op de verschillende leeskaarten



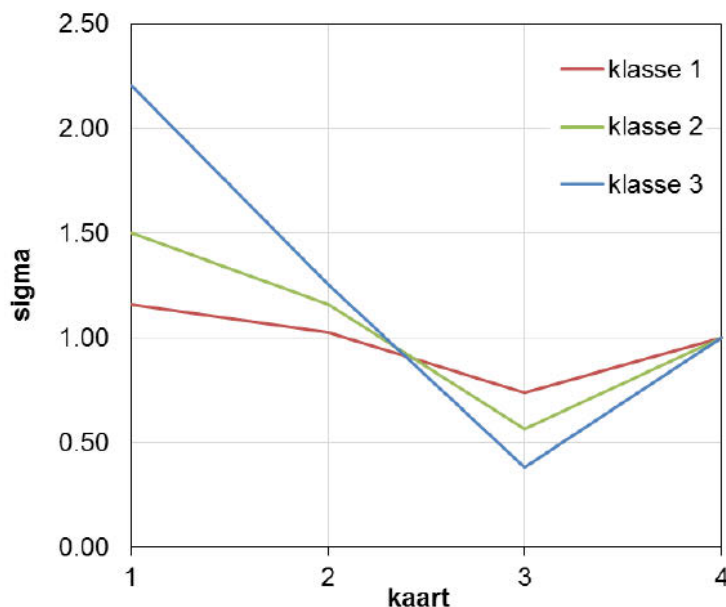
De schatting van het meetmodel heeft plaatsgevonden met behulp van een speciaal ontwikkeld computerprogramma. Het programma biedt de mogelijkheid om een meetmodel met latente klassen te kiezen. In een latent klasse model heeft elke klasse zijn eigen parameters. In plaats van drie parameters (één voor elke leeskaart) schatten we in een latent klasse model dus  $c \times 3$  sets met parameters, afhankelijk van het aantal klassen  $c$  dat we kiezen. Verkennende analyses lieten zien dat een model zonder latente klassen geen adequate beschrijving geeft van de DMT data. Zowel de kaartparameters  $\sigma$  als de persoonsparameters  $\theta$  waren niet invariant. De kaartparameters werden op basis van alleen de bovenbouwdata (groep 6 tot en met 8) anders geschat dan op basis van alleen de onderbouwdata (groep 3 tot en met 5) en voor de schatting van de persoonsparameters bleek het uit te maken op welke leeskaarten de persoonsparameters gebaseerd worden. Dit zijn twee belangrijke schendingen van het model (zie bijvoorbeeld Hambleton, Swaminathan, & Rogers, 1992) en daarom is het model in een vervolgstap verfijnd met latente klassen. Er zijn modellen geschat met 2, 3, 4 en 5 latente klassen. Het model met 3 latente klassen bleek uiteindelijk de beste beschrijving van de data te geven als we ook interpreteerbaarheid, spaarzaamheid en de afnameprocedure (minimaal een combinatie van twee leeskaarten) in ogenschouw nemen.

Bij gebruik van een latent klassenmodel is het de vraag hoe we over de prestaties van leerlingen rapporteren. In principe worden in een 3-lassenmodel voor elke leerling 3 vaardigheidsscores geschat; één voor elke klasse. Daarnaast kunnen we uitrekenen wat de kans is dat een leerling in een bepaalde klasse valt. We kunnen al deze informatie rapporteren aan de leerkracht en de leerling, maar dat is niet erg praktisch.

Een alternatief is om de vaardigheidsscore binnen elke klasse te vertalen naar één en dezelfde leeskaart (een zogeheten referentiekaart of standaardkaart met een  $\sigma$ -waarde gelijk aan 1) en dan uitsluitend de vaardigheidsscore van de meest waarschijnlijke klasse te rapporteren of de vaardigheidsscores per leerling te combineren tot een gewogen gemiddelde op basis van de latente klassen kansen. In geval van de DMT is gekozen voor de laatstgenoemde optie. De valideringsleeskaart (K4, zie paragraaf 4.1) is gebruikt als referentiekaart en de vaardigheidsscores die gerapporteerd worden aan de leraar en leerling hebben de vorm van een gewogen gemiddelde. De passing van het meetmodel is vanuit verschillende invalshoeken geëvalueerd. Eerst is gekeken naar de kaartparameters en het percentage leerlingen in elke klasse. Vervolgens is onderzocht of de persoonsparameters bij toepassing van het gekozen meetmodel invariant zijn.

In figuur 4.6 zijn de kaartparameters te vinden. We zien dat de  $\sigma$ -waarde van de valideringskaart voorafgaand aan het schatten van het model vastgezet is op 1. De  $\sigma$ -waarden van de overige leeskaarten zijn vrij geschat. Zoals we mogen verwachten op basis van woordtype is het voor leerlingen bij leeskaart 3 aanzienlijk moeilijker om veel woorden correct te decoderen in 1 minuut dan bij leeskaart 1. Daarnaast zien we dat de afstand in moeilijkheidsgraad gerelateerd is aan het klasselidmaatschap. Voor leerlingen in klasse 1 maakt het niet extreem veel uit welke leeskaart zij voorgelegd krijgen; zij zijn goed in staat om eenvoudigere woorden snel en correct te decoderen en kunnen “net zo goed” uit de voeten met complexere één- en tweelettergrepige woorden. Dat geldt niet voor leerlingen in klasse 2 en 3. Zij behalen relatief hoge scores op leeskaart 1, maar bij de leeskaarten met complexere woorden komen zij minder ver. Van de leerlingen die hebben deelgenomen aan het normeringsonderzoek valt in totaal 57 procent in klasse 1 (bij een classificatie op basis van de hoogste klassen kansen), 27 procent in klasse 2 en 16 procent in klasse 3.

Figuur 4.6  $\sigma$ -waarden DMT leeskaarten uitgesplitst naar klasse

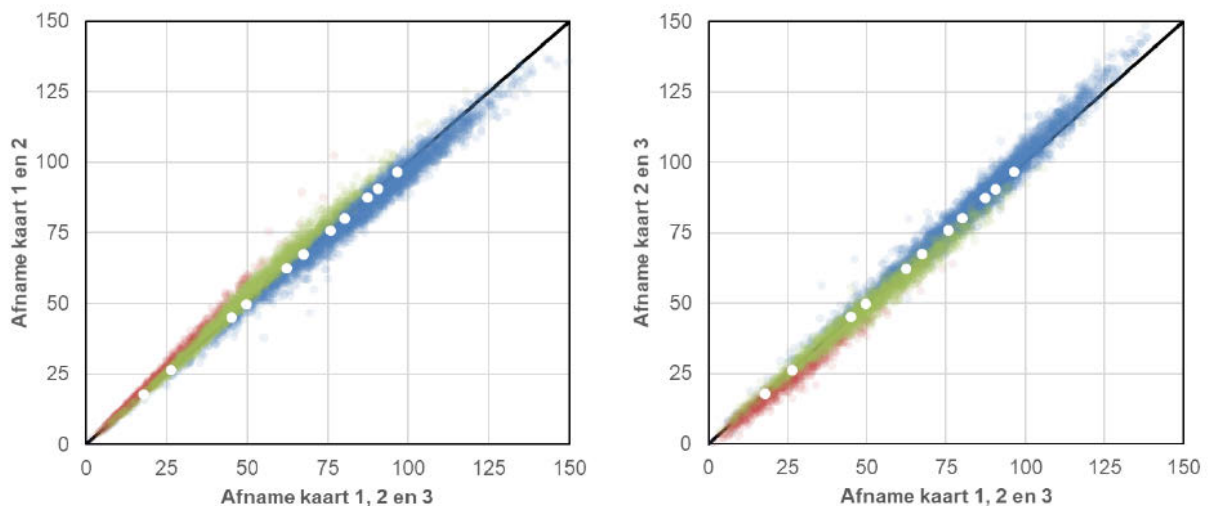


Als we een uitsplitsing maken naar afnamemoment blijken de klassen behoorlijk goed te relateren aan theoretische modellen over de ontwikkeling van (technische) leesvaardigheid. In de periode van het aanvankelijk lezen vallen de meeste leerlingen in klasse 3 (> 60 procent) en is klasse 1 relatief klein (< 15 procent). Naarmate de leesontwikkeling vordert, wordt het aantal leerlingen in klasse 1 groter en neemt het aantal leerlingen in respectievelijk klasse 3 en 2 af. Aan het einde van de basisschoolperiode zit meer dan 95 procent van de leerlingen in klasse 1. Een dergelijk patroon ligt voor de hand als we bedenken dat leerlingen in groep 3 beginnen met het decoderen van éénlettergrepige woorden met een eenvoudige MK, KM of MKM structuur (leeskaart 1) en daarna geleidelijk in aanraking worden gebracht met niet-klankzuivere één, twee, drie of vierlettergrepige woorden (leeskaart 2 en 3).



Om een zo nauwkeurig mogelijk schatting van de vaardigheid te verkrijgen is het wenselijk dat zoveel mogelijk alle leeskaarten worden afgenomen bij alle leerlingen. De afnameprocedure (zie Hoofdstuk 2) staat evenwel twee uitzonderingen toe. Omdat leerlingen halverwege groep 3 pas net gestart zijn met leren lezen wordt leeskaart 3 niet op dit afnamemoment aan leerlingen voorgelegd. Vanaf afnamemoment medio 5 kan er, met het oog op afnametijd, voor worden gekozen om leeskaart 1 te laten vervallen. Uitgaande van deze afnameregels is nagegaan in hoeverre de vaardigheid van leerlingen hetzelfde geschat wordt op basis van verschillende combinaties van leeskaarten. In figuur 4.7 worden de vaardigheidsscores zoals geschat op basis van een volledige afname (kaart 1, 2 en 3) in verband gebracht met de vaardigheidsscores zoals geschat op basis van een gedeeltelijke afname (leeskaart 1 en 2 of leeskaart 2 en 3). De kleuren in figuur 4.7 verwijzen naar de drie klassen (rood = klasse 1, groen = klasse 2 en blauw = klasse 3) en de witte markeringen representeren de gemiddelde score op de verschillende afnamemomenten, beginnend bij medio 3 en eindigend bij medio 8. We zien dat er sprake is van een lineaire relatie tussen de sets met vaardigheidsscores. Dit betekent dat er geen reden is om aan te nemen dat het model niet bij de data past.

*Figuur 4.7 Invariantie vaardigheidsscores bij toepassing van een 3-klasse Poisson-model*



#### 4.5 Normering

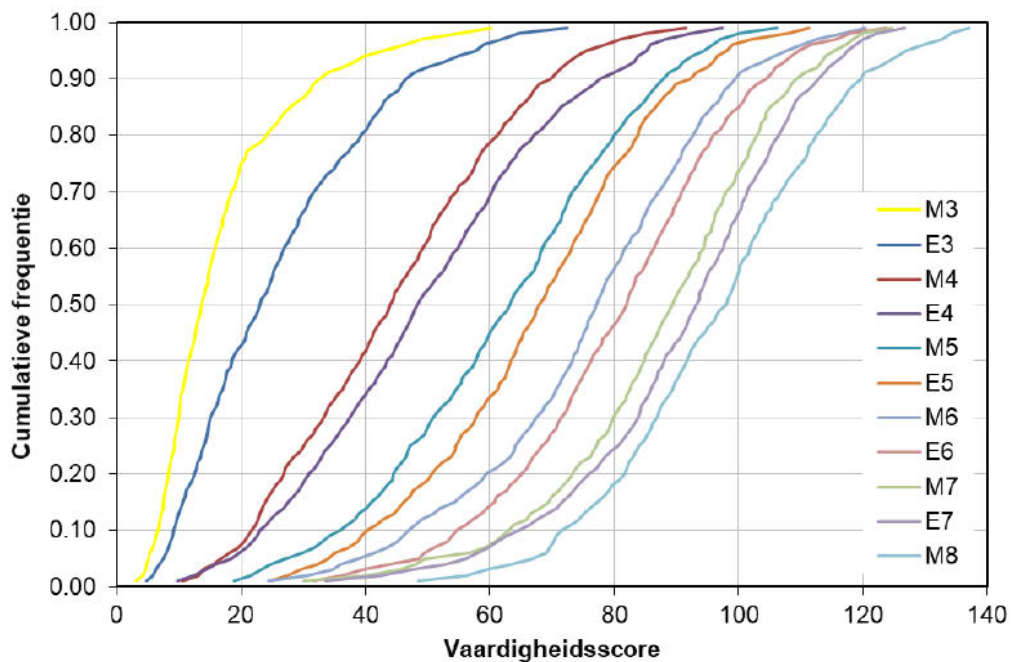
Na afloop van de afname van de DMT volgt er naast een vaardigheidsscore ook een indicatie van het relatieve niveau. We volgen daarbij de gebruikelijke Cito niveau-indelingen A-E en I-V. De grenswaarden hebben we gebaseerd op de verwachte scores van de leerlingen op de referentiekaart (K4), waarbij er per leerling een weging voor klasselidmaatschap heeft plaatsgevonden (zie ook paragraaf 4.4). In tabel 4.12 is de normtabel van de DMT te vinden.

Tabel 4.12 Normgegevens DMT naar afnamemoment

Afname	M	SD	p10	p20	p25	p40	p50	p60	p75	p80	p90
m3	17.88	12.44	6.58	8.33	9.29	11.48	13.64	15.93	20.10	24.11	32.60
e3	26.30	15.70	9.13	12.57	13.97	18.62	23.01	26.87	35.50	39.32	46.57
m4	45.04	18.67	21.68	26.89	30.06	39.18	44.02	49.23	57.77	61.23	69.75
e4	49.68	20.40	23.06	31.00	34.23	43.54	48.43	54.93	63.21	67.14	77.72
m5	62.35	20.03	35.97	44.58	47.24	57.74	63.28	68.48	76.90	79.95	87.85
e5	67.33	19.49	40.40	50.76	54.62	63.42	68.03	73.08	80.57	83.81	92.26
m6	75.85	20.40	47.02	59.35	64.16	73.26	77.20	81.78	90.12	92.76	99.50
e6	80.20	19.30	54.67	65.10	68.49	76.34	82.04	86.39	93.34	95.99	104.49
m7	87.41	19.26	63.09	73.30	77.12	84.80	89.45	94.39	100.85	103.01	109.34
e7	90.59	19.33	64.73	76.06	80.71	88.02	93.27	97.60	104.10	106.52	112.89
m8	96.56	18.95	71.57	81.66	84.09	91.74	98.09	101.84	110.16	112.48	119.70

In figuur 4.8 is de informatie van tabel 4.12 visueel weergegeven in (cumulatieve) scoreverdelingen. We zien dat de gemiddelde vaardigheid van de leerlingen steeds toeneemt tussen de opeenvolgende metingen, dat de variantie in de fase van het aanvankelijk lezen kleiner is dan in latere fasen van de leesontwikkeling, en dat de afstanden tussen de grenswaarden voor niveau-indelingen A-E en I-V behoorlijk groot zijn. Dit betekent dat de DMT gedurende de gehele basisschoolperiode goed in staat is om groepen leerlingen met een verschillend leesvaardigheidsniveau van elkaar te onderscheiden.

Figuur 4.8 Cumulatieve vaardigheidsverdelingen van medio groep 3 tot medio groep 8



Na het vaststellen van de normering is (als controle) nagegaan hoe deze normering uitpakt als we deze toepassen op de steekproef van leerlingen die heeft deelgenomen aan het normeringsonderzoek. Bij de controle hebben we alle mogelijkheden die op basis van de afnameregels zijn toegestaan, meegenomen. We zijn uitgegaan van de volgende grenswaarden:

- I score  $\geq$  p80
- II p60  $\leq$  score < p80
- III p40  $\leq$  score < p60
- IV p20  $\leq$  score < p40
- V score < p20

Op grond van de hierboven gekozen indeling verwachten we dat 20 procent van de leerlingen niveau I scoort, 20 procent niveau II, 20 procent niveau III, 20 procent niveau IV en 20 procent niveau V. Tabel 4.13 laat zien dat deze verdeling in voldoende mate teruggevonden wordt in onze steekproef van leerlingen. Daarbij maakt het weinig uit of gekozen wordt voor een volledige afname of voor een selectie van leeskaarten.

Tabel 4.13 Aantal en percentage leerlingen in niveaus I-V

Afname	Aantal leerlingen in niveaus I-V					Percentage leerlingen in niveaus I-V				
	I	II	III	IV	V	I	II	III	IV	V
<i>Kaart 1, 2 en 3</i>										
e3	145	143	121	109	137	22.1	21.8	18.5	16.6	20.9
m4	124	101	97	110	100	23.3	19.0	18.2	20.7	18.8
e4	183	154	162	157	164	22.3	18.8	19.8	19.1	20.0
m5	128	119	98	121	119	21.9	20.3	16.8	20.7	20.3
e5	136	146	147	142	149	18.9	20.3	20.4	19.7	20.7
m6	109	122	113	117	115	18.9	21.2	19.6	20.3	20.0
e6	142	137	165	135	157	19.3	18.6	22.4	18.3	21.3
m7	110	125	109	126	120	18.6	21.2	18.5	21.4	20.3
e7	137	122	143	135	134	20.4	18.2	21.3	20.1	20.0
m8	124	114	124	117	125	20.5	18.9	20.5	19.4	20.7
<i>Kaart 1 en 2</i>										
m3	126	111	99	90	138	22.3	19.7	17.6	16.0	24.5
<i>Kaart 2 en 3</i>										
m5	122	110	107	112	134	20.9	18.8	18.3	19.1	22.9
e5	146	144	127	146	157	20.3	20.0	17.6	20.3	21.8
m6	125	103	108	120	120	21.7	17.9	18.8	20.8	20.8
e6	170	132	139	140	155	23.1	17.9	18.9	19.0	21.1
m7	147	101	108	114	120	24.9	17.1	18.3	19.3	20.3
e7	165	120	121	127	138	24.6	17.9	18.0	18.9	20.6
m8	155	113	113	95	128	25.7	18.7	18.7	15.7	21.2

## 5 Betrouwbaarheid en meetnauwkeurigheid

### 5.1 Betrouwbaarheid

Omdat we beschikken over een zuivere schatter van de leessnelheid, het aantal correct gelezen woorden per tijdseenheid (minuut) op een standaard leeskaart, 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

$$(1) \quad \tilde{\theta}_v = \frac{S_v}{\sum_j d_{vj} \tau_j \sigma_j}$$

waarin  $S_v$  het totaal aantal woorden voorstelt dat leerling  $v$  heeft gelezen, dat wil zeggen: het totaal over alle leeskaarten die de leerling heeft gelezen. De designvariabele  $d_{vj}$  geeft aan of kaart  $j$  gelezen is door leerling  $v$  ( $d_{vj} = 1$ ) of niet ( $d_{vj} = 0$ );  $\sigma_j$  is de moeilijkheidsparameter van de leeskaart: hoe groter deze parameter hoe makkelijker de taak. De grootte  $\tau_j$  is geen parameter maar een bekende constante die de omvang van de taak uitdrukt. Bij de DMT is dit de toegestane tijd (1 minuut per leeskaart).

Dat deze schatter zuiver is betekent dat zijn verwachte waarde gelijk is aan de 'werkelijke' parameter  $\theta_v$ . 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  $\theta_v$  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:

$$(3) \quad \text{Var}(\tilde{\theta}_v | \theta_v) = SE^2(\tilde{\theta}_v | \theta_v) = \frac{S_v}{\left( \sum_j d_{vj} \tau_j \sigma_j \right)^2}$$

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:

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

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

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

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

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

Teller en noemer in de breuk in het rechterlid van (6) kunnen we consistent schatten als we beschikken over een representatieve steekproef. De geobserveerde variantie is gelijk aan 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.

Formule (6) hebben we uitgerekend voor de toegestane combinaties van de leeskaarten bij alle afname-momenten. De resultaten staan in tabel 5.1.

Tabel 5.1 *Betrouwbaarheid van de  $\tilde{\theta}$ -schatting en test-hertest correlatie voor toegestane combinaties van DMT-leeskaarten*

Halfjaargroep	Betrouwbaarheid $\theta$ -schatting			Betrouwbaarheid test-hertest		
	1 + 2	2 + 3	1 + 2 + 3	1 + 2	2 + 3	1 + 2 + 3
Medio groep 3	0,966			0,951		
Eind groep 3			0,976			0,969
Medio groep 4			0,965			0,955
Eind groep 4			0,966			0,957
Medio groep 5		0,921	0,955		0,925	0,943
Eind groep 5		0,905	0,946		0,91	0,931
Medio groep 6		0,905	0,945		0,909	0,934
Eind groep 6		0,889	0,936		0,894	0,924
Medio groep 7		0,883	0,932		0,885	0,922
Eind groep 7		0,879	0,927		0,881	0,919
Medio groep 8		0,863	0,917		0,866	0,91

De betrouwbaarheidsindices in tabel 5.1 voor de toegestane combinaties van leeskaarten hebben hoge waarden (.86 of hoger). Voor toetsen van het type waar geen zware consequenties voor leerlingen aan verbonden zijn, zoals de DMT, geeft de COTAN (Commissie Testaangelegenheden van het Nederlands Instituut voor Psychologen) 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 DMT goed te noemen voor alle onderscheiden afname-momenten.

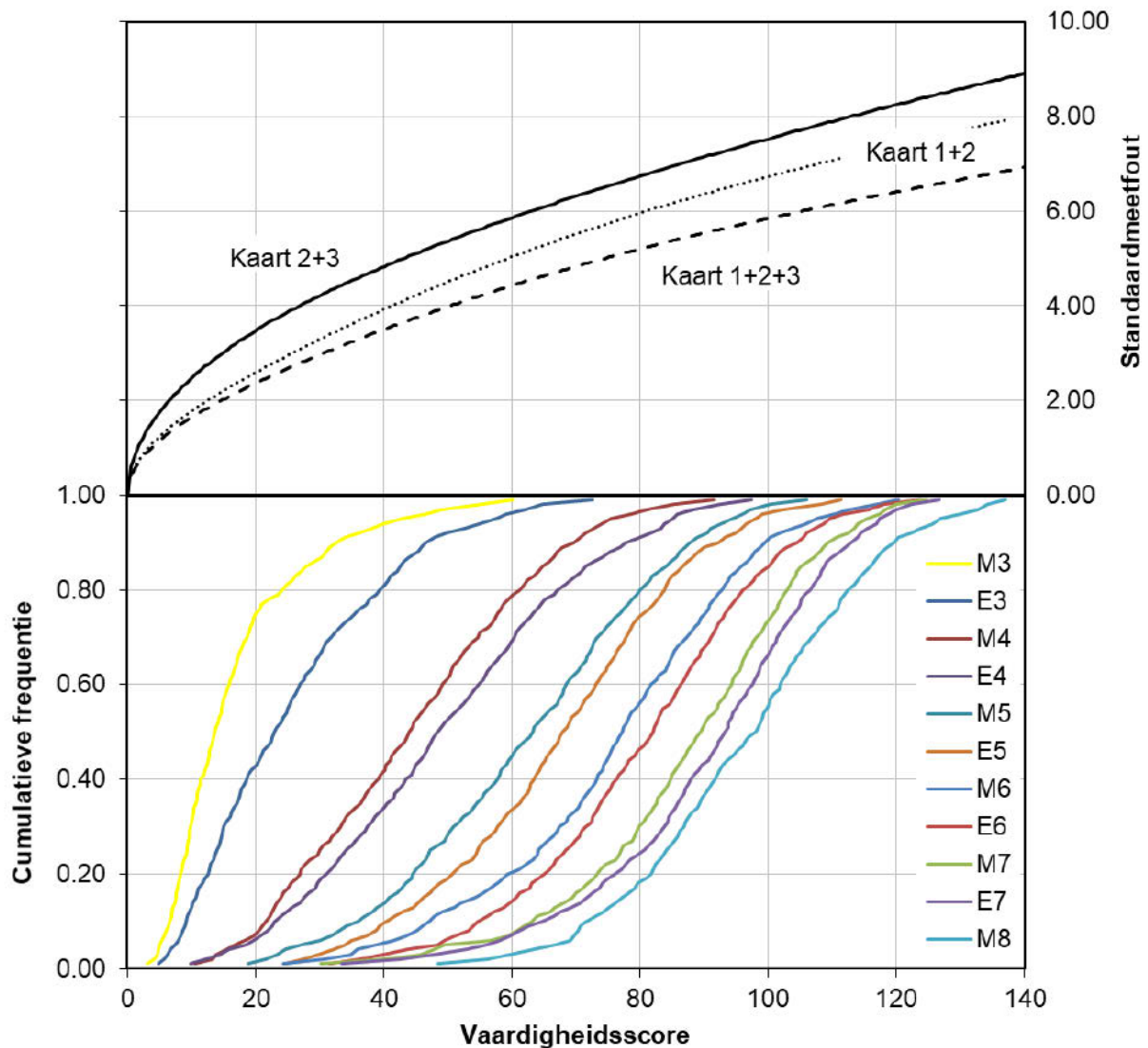
Er heeft geen test-hertest onderzoek plaatsgevonden. De afnamecontext van de DMT-kaarten leent zich daar niet goed voor. Het feit dat alle leeskaarten echter gekalibreerd zijn, maakt het mogelijk een hertest te simuleren. We hebben een dubbele afname gesimuleerd van 100.000 leerlingen. Daarbij hebben we enerzijds de vaardigheidsverdeling van alle leerlingen, 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 geen 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 toegestane combinaties van DMT-leeskaarten. De uitkomsten komen goed overeen met eerder berekende coëfficiënten en leiden dan ook tot dezelfde conclusies met betrekking tot de betrouwbaarheid van de DMT.

## 5.2 Nauwkeurigheid

### Lokale meetnauwkeurigheid (grafisch)

De hiervoor vermelde betrouwbaarheidscoëfficiënten hebben betrekking op de *globale* meetnauwkeurigheid van de toegestane combinaties van leeskaarten, maar geven geen beeld van de *lokale* meetnauwkeurigheid. Dat doet figuur 5.1 wel. In deze figuur staat voor combinaties van DMT-leeskaarten de grootte van de meetfout afgebeeld. In de onderste helft van figuur 5.1 staan de cumulatieve vaardigheidsverdelingen voor de afnamemomenten.

Figuur 5.1 Grootte van de meetfout voor combinaties van DMT-leeskaarten



Figuur 5.1 maakt duidelijk dat voor combinaties van leeskaarten de meetfout toeneemt met toenemende vaardigheid, maar alleszins binnen de perken blijft; zeker voor de aanbevolen combinatie van leeskaart 1+2+3 (m.u.v. afnamemoment M3).

### Lokale meetnauwkeurigheid (tabellarisch)

De betekenis van de meetnauwkeurigheid voor de beslissingen die met de toetsen genomen worden, is af te leiden uit classificatie-/misclassificatietabellen. Tabel 5.2 laat het effect van de lokale meetnauwkeurigheid zien voor de afnamecombinatie van drie kaarten voor afnamemoment E4. Uit de tabel komt naar voren dat

84% van de leerlingen die eind jaargroep 4 met hun geschatte vaardigheidsscore aan niveaugroep A (zie paragraaf 2.3 voor de interpretatie van de vaardigheidsniveaus) worden toegekend, ook met hun werkelijke vaardigheidsscore in deze groep vallen. Anders gezegd, de kans dat een A-leerling (terecht) als een A-leerling wordt bestempeld is 84%. Verder laat de tabel zien dat ongeveer 16% van de leerlingen in niveaugroep A een werkelijke vaardigheidsscore heeft die in scoregroep B valt. Het percentage leerlingen in niveaugroep A dat in de werkelijke scoregroep C, D of E valt, is kleiner dan 1%. Bijlage 1 bevat classificatie/misclassificatietabellen voor de toegestane afnamecombinaties voor alle afnamemomenten. Uit de bijlage komt naar voren dat de lokale meetnauwkeurigheid groter is bij de afname van drie kaarten dan bij de afname van twee kaarten. Daarom wordt in de handleiding bij de toetsen ook aanbevolen om drie leeskaarten af te nemen, met uitzondering van het afnamemoment medio groep 3, wanneer de woorden van leeskaart 3 nog te hoog gegrepen zijn voor de meeste leerlingen.

Tabel 5.2 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind jaargroep 4 (E4)

Score-groepen V t/m I	Scoregroep waarin de ware score valt					Score-groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	85,1	14,9	0	0	0	E	82,8	17,2	0	0	0
IV	11,7	69,2	18,8	0,3	0	D	10,4	66,6	22,9	0,1	0
III	0,1	18	65,3	16,4	0,2	C	0	10,6	75,2	14,1	0
II	0	0,7	23,5	64,1	11,6	B	0	0,1	17,2	71,8	11
I	0	0	0,5	19,1	80,5	A	0	0	0,2	16,3	83,6

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, óf een scoregroep daarboven óf een scoregroep daaronder. In tabel 5.2 zijn dit de gearceerde cellen. Dit ambitieniveau is gebaseerd op de veronderstelling dat geen enkele toets perfect meet en dat er dus altijd sprake is van foutieve classificaties. In dat licht is de maximale accuraatheid die op het individuele niveau bereikt kan worden plus of minus één scoregroep. De tweede maat wordt op verschillende plekken in de literatuur beschreven. De maat laat zien hoe vaak de classificatie op basis van de geschatte vaardigheidsscore gemiddeld gezien overeenstemt met de classificatie op basis van de (gesimuleerde) werkelijke vaardigheidsscore. Bij een ideale, maar in de praktijk niet te realiseren, toetsafname lijkt de *Marginal Classification Accuracy* rond 0,75 - 0,80 uit te komen. In de praktijk liggen de waarden vaak tussen 0,60 en 0,70.

De samenvattende indices voor afnamemoment E4 voor de afnamecombinatie van leeskaart 1+2+3 zijn te vinden in tabel 5.3. Deze tabel laat zien dat de uitkomsten in lijn liggen met het ambitieniveau zoals dat geformuleerd is door Pilliner (1969). Gemiddeld gezien scoort 100 procent van de leerlingen in een niveaugroep ook in werkelijkheid in die niveaugroep, óf een niveaugroep daarboven óf daaronder. De *Marginal Classification Accuracy* ligt boven de 74 procent Dit betekent dat de classificatie op basis van de geschatte vaardigheidsscore bij afnamemoment eind groep 4 op basis van drie leeskaarten gemiddeld gezien in ongeveer driekwart van de gevallen overeenkomt met de classificatie op basis van de (gesimuleerde) werkelijke vaardigheidsscore. De samenvattende indices voor de andere afnamemomenten op basis van alle toegestane afnamecombinaties zijn te vinden in bijlage 2. Deze bijlage toont aan dat de samenvattende indices bij afname van drie leeskaarten min of meer hetzelfde beeld laten zien als bij



onderstaande tabel. Alleen de Marginal Classification Accuracy daalt in de hogere groepen naar ongeveer 70%.

*Tabel 5.3 Samenvattende indices voor de afnamecombinatie van leeskaarten 1+2+3 eind jaargroep 4*

	marginal classification accuracy	accuracy plus/minus 1 niveau
	leeskaart 1+2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	76	99,9
vaardigheidsniveaus I t/m V	73,5	99,7



## 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 DMT echter niet bedoeld is voor voorspellend gebruik, is criteriumvaliditeit niet van toepassing.

### 6.1 Inhoudsvaliditeit

Beoogd is de DMT onafhankelijk samen te stellen van de verschillende onderwijsmethoden voor technisch lezen. De getoetste woorden zijn aselekt gekozen uit het BasiLex-corpus op basis van criteria die geheel los staan van het al dan niet voorkomen van deze woorden in de op basisscholen gebruikte lesboekjes. Vervolgens is een procedure gebruikt om op basis van 8 verschillende stappen de leeskaarten samen te kunnen stellen (zie paragraaf 3.2.2). Daarna zijn de kaarten aangepast na opmerkingen van Vlaamse en Nederlandse vakexperts, leerkrachten en na analyse van de proeftoetsresultaten. Voor de definitieve leeskaarten geldt dat ze, ook na de herzieningen, voldeden aan de vooraf gewenste verdeling over woordtypen en leesmoeilijkheden. Dit alles vormt een degelijke basis voor de inhoudsvaliditeit van de DMT.

### 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 DMT onder een zekere tijdsdruk af te nemen. Ook wordt in deze paragraaf de keuze voor het opnemen van bestaande woorden (in tegenstelling tot pseudowoorden) verantwoord. In paragraaf 3.2.1 worden de overige aspecten die kenmerkend zijn voor de opzet van de DMT verantwoord. In de eerste plaats gaat het daarbij om de keuze om leerlingen hardop te laten lezen en in de tweede plaats om het hanteren van het verdelen van woordtypen over drie verschillende kaarten.

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 DMT is geoperationaliseerd, vormt dit een degelijke basis voor de begripsvaliditeit van de DMT.

### 6.3 Convergente en divergente validiteit

De mate waarin de DMT gerelateerd is aan soortgenootinstrumenten als de AVI-toets (Cito, 2017) en de EMT (Een-Minut-Test), (Brus en Voeten, 1979) en niet-soortgenootinstrumenten is een belangrijke aanwijzing voor de begripsvaliditeit van de toets. In de tabellen 6.1 tot en met 6.6 worden correlaties gerapporteerd tussen scores op de DMT, de AVI-toets (uitgesplitst naar tijd en fouten), de EMT en de toetsen Begrijpend lezen en Woordenschat die in het kader van het Cito-Leerlingvolgsysteem primair en speciaal onderwijs zijn ontwikkeld. Met uitzondering van de recent gepubliceerde AVI-toets (die tegelijkertijd

met de nieuwe uitgave van de DMT wordt aangeboden ter verantwoording) zijn alle bovengenoemde toetsen positief beoordeeld door de Cotan of de Expertgroep PO. Zoals in hoofdstuk 4 is beschreven, maakten de EMT en de AVI-toets deel uit van het normeringsonderzoek. De scores van de deelnemende leerlingen op de toetsen Begrijpend Lezen en Woordenschat zijn in een later stadium, middels datareturn via het computerprogramma LOVS, opgevraagd. In de tabellen zijn de cellen leeg gelaten waarbij de correlaties betrekking hadden op minder dan 50 leerlingen.

Tabel 6.1 tot en met 6.6 laten zien dat de scores op de DMT, de EMT en AVI-leestijd onderling sterk samenhangen, wat naar verwachting is. 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. Ten slotte is de samenhang tussen de DMT en de toetsen Begrijpend lezen en Woordenschat matig tot laag. Dit komt omdat bij deze toetsen andere vaardigheden dan alleen technisch lezen een rol spelen. Samengevat kunnen we de hoge correlaties van de DMT met de soortgenotentoetsen en de matige tot lage correlaties met de niet-soortgenotentoetsen beschouwen als een aanwijzing voor de begripsvaliditeit van de DMT.

Tabel 6.1 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 3 (n = 843)

	DMT medio	DMT eind	AVI tijd medio	AVI tijd eind	AVI fouten medio	AVI fouten eind	EMT medio	Begrijpend lezen einde	Woordenschat medio	Woordenschat einde
DMT medio	1,00									
DMT eind	0,85	1,00								
AVI tijd medio	0,85	0,90	1,00							
AVI tijd eind	0,75	0,93	0,92	1,00						
AVI fouten medio	-0,48	-0,54	-0,56	-0,53	1,00					
AVI fouten eind	-0,51	-0,66	-0,59	-0,63	0,87	1,00				
EMT medio	0,79	0,71	0,74	0,67	-0,34	-0,46	1,00			
Begrijpend lezen einde	0,39	0,43	0,41	0,46	-0,46	-0,64		1,00		
Woordenschat medio	0,02	-0,05	-0,06	-0,06	-0,09	-0,21		0,31	1,00	
Woordenschat einde	0,16	0,09	0,02	0,10	-0,17	-0,23		0,40	0,93	1,00

Tabel 6.2 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 4 (n = 1003)

	DMT medio	DMT eind	AVI tijd medio	AVI tijd eind	AVI fouten medio	AVI fouten eind	EMT medio	Begrijpend lezen einde	Woordenschat medio	Woordenschat einde	DMT medio
DMT medio	1,00										
DMT eind	0,97	1,00									
AVI tijd medio	0,87	0,87	1,00								
AVI tijd eind	0,87	0,92	0,96	1,00							
AVI fouten medio	-0,67	-0,65	-0,51	-0,59	1,00						
AVI fouten eind	-0,66	-0,68	-0,51	-0,59	0,96	1,00					
EMT medio	0,95	0,98	0,90	0,93	-0,64	-0,66	1,00				
Begrijpend lezen medio	0,51	0,52	0,54	0,59	-0,49	-0,45		1,00			
Begrijpend lezen eind	0,44	0,51	0,50	0,53	-0,34	-0,40		0,79	1,00		
Woordenschat medio	0,27	0,35	0,43	0,54	-0,30	-0,19		0,60	0,58	1,00	
Woordenschat eind	0,38	0,49	0,45	0,56	-0,37	-0,36		0,67	0,64	0,81	1,00

Tabel 6.3 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 5 (n = 919)

	DMT medio	DMT eind	AVI tijd medio	AVI tijd eind	AVI fouten medio	AVI fouten eind	EMT medio	Begrijpend lezen medio	Begrijpend lezen eind	Woordenschat medio
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			
Begrijpend lezen medio	0,497	0,489	0,658	0,556	-0,491	-0,450		1,000		
Begrijpend 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 6.4 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 6 (n = 935)

	DMT medio	DMT eind	AVI tijd medio	AVI tijd eind	AVI fouten medio	AVI fouten eind	EMT medio	Begrijpend lezen medio	Woordenschat medio	Woordenschat eind
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			
Begrijpend 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 6.5 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 7 (n = 881)

	DMT medio	DMT eind	AVI tijd medio	AVI tijd eind	AVI fouten medio	AVI fouten eind	EMT medio	Begrijpend lezen medio	Begrijpend lezen eind	Woordenschat medio	Woordenschat eind
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				
Begrijpend lezen medio	0,275	0,372	0,610	0,579	-0,276	-0,379		1,000			
Begrijpend 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 6.6 Correlaties tussen de DMT, de AVI-toets en andere toetsen in groep 8 (n = 604)

	DMT medio	AVI tijd medio	AVI fouten medio	EMT medio	Begrijpend lezen medio	Woordenschat medio
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		
Begrijpend 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. Van ongeveer 2000 leerlingen in de steekproef was aangegeven of er sprake was van dyslexie of niet. 10% van deze groep was aangemerkt als dyslectisch. Bij het maken van effectschattingen is Cohens D gebruikt. De interpretatie van de effectgroottes is volgens Cohen (1988). De resultaten van de effect-schattingen zijn opgenomen in tabel 6.7. De tabel laat zien dat voor de variabelen regio en verstedelijking geen betekenisvolle verschillen worden gevonden. Voor de variabele wel of geen dyslexie worden grote 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 6.7 Resultaten effectschattingen voor sekse, leerlinggewicht en wel of geen dyslexie

	DMT medio	DMT eind
sekse	0,03	0,06
leerlinggewicht	0,01	-0,6
wel of geen dyslexie	0,32	0,48





## 7 Samenvatting

In dit samenvattende hoofdstuk geven we kort weer wat in de voorafgaande hoofdstukken is besproken. In hoofdstuk 1 wordt aangegeven dat De DMT uit het Cito Volgsysteem primair en speciaal onderwijs bedoeld is 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 uitgangpunten van de toetsconstructie. Na elkaar komen aan de orde: de meetpretentie van de DMT, 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 DMT 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 welke procedure is gevolgd bij het samenstellen van de leeskaarten. Ten slotte wordt aangetoond dat de definitieve leeskaarten voldeden aan de beoogde verdeling van woorden over woordtypen en leesmoelijkheden.

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 uitgebalancheerde sets leeskaarten aan leerlingen konden worden voorgelegd. Vervolgens wordt de representativiteit van de steekproef beschreven. Aangetoond wordt dat de scoreverdelingen van leerlingen in representatieve *subsamples* van de steekproef volledig overlappen met die van de totale steekproef. Dit betekent dat het voor de normering weinig uitmaakt of deze gebaseerd is op de volledige steekproeven of de *subsamples*. Daarom is ervoor gekozen om de normering te baseren op de volledige steekproeven. Daarnaast wordt het meetmodel beschreven en wordt gerapporteerd hoe op basis van dit meetmodel de kalibratie heeft plaatsgevonden. Ten slotte is weergegeven hoe de normering tot stand is gekomen en wordt aangetoond dat de DMT 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 het aantal correct gelezen woorden per tijdseenheid op een standaardkaart te berekenen. Vervolgens worden betrouwbaarheidsindices gepresenteerd. Deze zijn met 0,86 en hoger 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 lokale meetnauwkeurigheid groter is bij de afname van drie kaarten dan bij de afname van twee kaarten. Daarom wordt in de handleiding bij de toetsen ook aanbevolen om drie leeskaarten af te nemen, met uitzondering van het afnamemoment medio groep 3, wanneer de woorden van leeskaart 3 nog te hoog gegrepen zijn voor de meeste leerlingen. Ten slotte wordt de betekenis van de meetnauwkeurigheid voor de beslissingen die met de toetsen genomen worden, gepresenteerd in classificatie-/misclassificatietabellen. De samenvattende indices geven aan dat de uitkomsten in lijn liggen met het ambitieniveau zoals geformuleerd door Pilliner (1969).

In hoofdstuk 6 wordt ingegaan op de validiteit van de DMT. In de eerste plaats wordt aangetoond dat de gevolgde procedure bij het samenstellen van de DMT garant staat voor de inhoudsvaliditeit van de toets. Vervolgens wordt betoogd dat de manier waarop het construct Technisch Lezen bij de DMT is geoperationaliseerd, een degelijke basis vormt voor de begripsvaliditeit van de DMT. Daarnaast wordt uitgelegd dat de hoge correlatie van de DMT met de AVI-toets (tijd) en de EMT als bewijs van hoge

convergente validiteit geldt. De lage tot matige correlaties met de toetsen Begrijpend lezen en Woordenschat kunnen als bewijs van divergente validiteit worden opgevat. Ten slotte wordt aangetoond dat de prestaties van relevante subgroepen op de DMT aansluiten bij theoretische verwachtingen ten aanzien van de prestaties van deze 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-Minuut-Test vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout Testmateriaal BV.
- Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, 12, 65-85.
- 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.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00054.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J (1992). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- 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.
- Huntsman, L.A., & Lima, S.D. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research, 31*, 289–306.
- 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.
- Koning, L. (2009). *Nieuwe AVI-toetsen en de DMT; wat is er mis mee?* Lekkerkerk: Pravoo.
- 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.
- New, B., Ferrand, L., Pallier, C. & Brysbaert, M. (2006). *Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project*. *Psychonomic Bulletin and Review, 13*, 1, 45-52.
- 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-Minuut-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. (<http://dx.doi.org/10.1037/a0026703>)
- 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.
- 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.

- Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1140-1154. (doi:10.1037/0278-7393.21.5.1140)
- Taylor, T.E., & Lupker, S. . (2001). Sequential effects in naming: A time-criterion account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27, 117-138.
- Tellings, A., Hulsbosch, M., Vermeer, A. & van den Bosch, A. (2014). BasiLex: an 11.5 million word corpus of Dutch texts written for children. *Computational Linguistics in the Netherlands*, 4, 191-208.
- 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.
- Verhoeven, L. en Aarnoutse, C. (2000). *Tussendoelen beginnende geletterdheid*. Nijmegen: Expertisecentrum Nederlands.
- 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-/misclassificatietabel voor de combinatie van leeskaarten 1+2 medio groep 3*

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	76,1	18,3	5,2	0,4	0,0	E	67,7	25,8	6,3	0,2	0,0
IV	17,8	47,7	26,6	7,9	0,0	D	11,6	51,2	30,6	6,6	0,0
III	1,0	22,7	51,7	23,9	0,7	C	0,3	18,3	56,1	23,8	1,6
II	0,0	1,1	23,5	64,1	11,2	B	0,0	0,5	22,6	62,2	14,7
I	0,0	0,0	0,1	28,3	71,6	A	0,0	0,0	0,1	21,1	78,8

Tabel 2 *Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind groep 3*

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	84,1	15,8	0,2	0,0	0,0	E	78,2	21,0	0,8	0,0	0,0
IV	11,8	71,1	17,0	0,1	0,0	D	11,0	62,2	26,7	0,0	0,0
III	0,0	18,2	69,6	12,2	0,0	C	0,1	8,1	82,3	9,6	0,0
II	0,0	0,1	23,8	69,7	6,4	B	0,0	0,0	20,5	72,5	7,0
I	0,0	0,0	0,0	20,4	79,6	A	0,0	0,0	0,0	19,1	80,9

Tabel 3 *Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 medio groep 4*

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	85,3	14,6	0,0	0,0	0,0	E	83,0	16,7	0,3	0,0	0,0
IV	12,6	70,7	16,4	0,3	0,0	D	13,4	64,3	22,3	0,0	0,0
III	0,0	20,0	63,8	16,1	0,1	C	0,0	13,8	73,1	13,0	0,1
II	0,0	1,3	23,7	66,4	8,6	B	0,0	0,0	19,9	71,1	9,0
I	0,0	0,0	0,6	21,1	78,3	A	0,0	0,0	0,3	17,5	82,2

Tabel 4 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind groep 4

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	85,1	14,9	0,0	0,0	0,0	E	82,8	17,2	0,0	0,0	0,0
IV	11,7	69,2	18,8	0,3	0,0	D	10,4	66,6	22,9	0,1	0,0
III	0,1	18,0	65,3	16,4	0,2	C	0,0	10,6	75,2	14,1	0,0
II	0,0	0,7	23,5	64,1	11,6	B	0,0	0,1	17,2	71,8	11,0
I	0,0	0,0	0,5	19,1	80,5	A	0,0	0,0	0,2	16,3	83,6

Tabel 5 Classificatie-/misclassificatietabel voor de combinatie van kaarten 2+3 medio groep 5

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	84,0	15,9	0,1	0,0	0,0	E	80,9	18,9	0,1	0,0	0,0
IV	11,5	68,9	19,1	0,6	0,0	D	7,7	67,5	24,6	0,1	0,0
III	0,0	18,6	60,4	20,1	0,9	C	0,0	9,8	74,2	15,7	0,3
II	0,0	0,6	22,7	59,0	17,8	B	0,0	0,0	18,6	65,3	16,1
I	0,0	0,0	0,7	15,9	83,4	A	0,0	0,0	0,2	15,7	84,1

Tabel 6 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 medio groep 5

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	82,2	17,5	0,2	0,0	0,0	E	77,7	22,1	0,2	0,0	0,0
IV	9,3	68,7	21,3	0,7	0,0	D	7,8	66,6	25,5	0,1	0,0
III	0,1	20,0	58,5	20,9	0,6	C	0,1	9,3	73,8	16,8	0,1
II	0,0	0,7	25,7	57,2	16,3	B	0,0	0,0	21,5	63,2	15,2
I	0,0	0,0	1,1	17,2	81,7	A	0,0	0,0	0,4	16,1	83,4

Tabel 7 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 eind groep 5

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	86,1	13,7	0,2	0,0	0,0	E	79,2	20,8	0,1	0,0	0,0
IV	11,3	65,0	22,3	1,5	0,0	D	9,6	70,5	19,7	0,2	0,0
III	0,1	18,0	56,5	23,8	1,6	C	0,0	10,8	68,2	20,4	0,6
II	0,0	1,1	24,3	55,4	19,2	B	0,0	0,1	18,9	63,3	17,7
I	0,0	0,0	1,5	19,7	78,8	A	0,0	0,0	0,5	18,6	80,9

Tabel 8 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind groep 5

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
	83,8	15,9	0,3	0,0	0,0	E	78,9	21,0	0,1	0,0	0,0
IV	10,8	61,7	26,0	1,5	0,0	D	6,9	67,3	25,6	0,3	0,0
III	0,1	18,1	55,6	25,2	0,9	C	0,0	10,5	65,7	23,3	0,4
II	0,0	1,4	22,0	58,7	17,9	B	0,0	0,1	17,9	64,3	17,7
I	0,0	0,0	1,4	17,4	81,2	A	0,0	0,0	0,5	15,8	83,7

Tabel 9 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 medio groep 6

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	84,9	14,9	0,2	0,0	0,0	E	82,7	17,2	0,1	0,0	0,0
IV	9,6	66,4	20,5	3,4	0,0	D	6,1	69,3	24,0	0,6	0,0
III	0,1	23,8	46,8	27,8	1,5	C	0,0	11,7	64,7	22,9	0,7
II	0,0	3,0	22,8	56,7	17,5	B	0,0	0,2	20,2	63,9	15,7
I	0,0	0,0	1,3	19,9	78,8	A	0,0	0,0	0,5	19,8	79,8

Tabel 10 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 medio groep 6

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	79,3	20,4	0,2	0,0	0,0	E	80,3	19,6	0,1	0,0	0,0
IV	8,4	65,7	23,3	2,5	0,0	D	5,4	65,6	28,6	0,4	0,0
III	0,2	19,7	51,2	28,3	0,7	C	0,0	10,2	65,7	23,9	0,3
II	0,0	2,4	19,0	63,2	15,4	B	0,0	0,3	15,7	69,4	14,6
I	0,0	0,0	0,8	16,7	82,4	A	0,0	0,0	0,4	16,4	83,2

Tabel 11 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 eind groep 6

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	81,7	17,4	0,9	0,0	0,0	E	78,8	20,9	0,3	0,0	0,0
IV	13,6	58,4	25,7	2,2	0,1	D	7,5	64,8	27,2	0,5	0,0
III	0,6	21,6	54,1	21,3	2,4	C	0,0	11,9	67,3	19,8	1,0
II	0,0	2,1	30,1	47,5	20,2	B	0,0	0,2	25,8	56,2	17,8
I	0,0	0,0	3,1	19,5	77,4	A	0,0	0,0	1,4	20,1	78,4

Tabel 12 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind groep 6

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
	77,4	21,8	0,8	0,0	0,0	E	71,2	28,3	0,5	0,0	0,0
IV	11,0	58,9	28,7	1,3	0,0	D	6,5	61,3	32,0	0,2	0,0
III	0,5	16,9	59,8	21,6	1,2	C	0,0	9,7	70,3	19,6	0,4
II	0,0	1,2	24,8	54,7	19,3	B	0,0	0,1	20,2	63,0	16,7
I	0,0	0,0	1,6	16,1	82,3	A	0,0	0,0	0,7	15,9	83,4

Tabel 13 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 medio groep 7

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	77,1	21,3	1,6	0,0	0,0	E	75,9	23,5	0,6	0,0	0,0
IV	14,2	57,1	26,3	2,3	0,1	D	6,5	63,1	29,0	1,5	0,0
III	0,8	23,9	52,6	20,3	2,5	C	0,1	15,8	60,9	21,8	1,4
II	0,0	3,3	32,6	44,5	19,6	B	0,0	0,8	26,4	54,5	18,3
I	0,0	0,1	3,9	21,8	74,2	A	0,0	0,0	1,7	20,6	77,7

Tabel 14 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 medio groep 7

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	77,0	22,2	0,8	0,0	0,0	E	71,3	28,3	0,4	0,0	0,0
IV	9,8	62,0	27,0	1,2	0,0	D	4,7	64,0	30,5	0,8	0,0
III	0,3	17,9	60,3	20,2	1,3	C	0,0	11,0	66,7	21,8	0,5
II	0,0	1,2	27,0	53,0	18,7	B	0,0	0,3	20,5	62,4	16,8
I	0,0	0,0	1,7	18,4	80,0	A	0,0	0,0	0,7	17,1	82,2

Tabel 15 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 eind groep 7

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	78,2	20,7	1,1	0,0	0,0	E	74,0	25,6	0,4	0,0	0,0
IV	13,2	59,7	23,6	3,4	0,1	D	5,1	67,6	26,3	0,9	0,0
III	0,6	25,0	48,0	22,9	3,4	C	0,0	16,0	62,8	19,8	1,3
II	0,0	4,5	31,3	45,1	19,1	B	0,0	0,7	28,0	54,0	17,3
I	0,0	0,1	4,3	21,2	74,4	A	0,0	0,0	1,9	21,1	77,0

Tabel 16 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 eind groep 7

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
	78,1	21,5	0,4	0,0	0,0	E	68,6	31,3	0,2	0,0	0,0
IV	9,7	65,7	23,0	1,6	0,0	D	4,4	68,6	26,7	0,3	0,0
III	0,3	20,8	56,2	21,3	1,4	C	0,1	11,0	69,1	19,4	0,5
II	0,0	1,6	27,4	53,3	17,6	B	0,0	0,2	22,2	62,2	15,4
I	0,0	0,0	1,6	17,8	80,5	A	0,0	0,0	0,7	17,7	81,6

Tabel 17 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 2+3 medio groep 8

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	75,4	21,8	2,7	0,1	0,0	E	69,1	29,4	1,5	0,0	0,0
IV	16,4	52,1	28,3	3,2	0,0	D	8,2	60,4	30,5	0,9	0,0
III	1,4	22,3	51,3	23,3	1,7	C	0,1	14,4	64,9	19,5	1,0
II	0,0	2,9	28,4	51,7	17,0	B	0,0	0,4	26,9	56,1	16,6
I	0,0	0,1	2,5	22,7	74,7	A	0,0	0,0	1,5	21,4	77,1

Tabel 18 Classificatie-/misclassificatietabel voor de combinatie van leeskaarten 1+2+3 medio groep 8

Score- groepen V t/m I	Scoregroep waarin de ware score valt					Score- groepen E t/m A	Scoregroep waarin de ware score valt				
	V	IV	III	II	I		E	D	C	B	A
V	77,9	21,0	1,2	0,0	0,0	E	66,8	32,5	0,7	0,0	0,0
IV	12,4	59,3	26,9	1,3	0,0	D	5,4	63,3	30,9	0,3	0,0
III	0,5	18,5	59,5	20,9	0,6	C	0,1	10,2	71,6	17,8	0,3
II	0,0	1,0	23,6	60,3	15,0	B	0,0	0,1	21,4	64,2	14,3
I	0,0	0,0	0,8	18,3	80,9	A	0,0	0,0	0,4	17,7	81,9

## Bijlage 2

Tabel 1 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 3*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	65,3	n.v.t.	n.v.t.	97,6	n.v.t.	n.v.t.
vaardigheidsniveaus I t/m V	64,1	n.v.t.	n.v.t.	96,9	n.v.t.	n.v.t.

Tabel 2 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment eind groep 3*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	n.v.t.	76,8	n.v.t.	n.v.t.	99,9
vaardigheidsniveaus I t/m V	n.v.t.	n.v.t.	74,4	n.v.t.	n.v.t.	99,9

Tabel 3 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 4*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	n.v.t.	74,7	n.v.t.	n.v.t.	99,9
vaardigheidsniveaus I t/m V	n.v.t.	n.v.t.	72,9	n.v.t.	n.v.t.	99,5

Tabel 4 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment eind groep 4*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	n.v.t.	76	n.v.t.	n.v.t.	99,9
vaardigheidsniveaus I t/m V	n.v.t.	n.v.t.	73,5	n.v.t.	n.v.t.	99,7

Tabel 5 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 5*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	74,4	73,3	n.v.t.	99,8	99,8
vaardigheidsniveaus I t/m V	n.v.t.	71,6	70,6	n.v.t.	99,4	99,3

Tabel 6 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment eind groep 5*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	72	71,7	n.v.t.	99,7	99,7
vaardigheidsniveaus I t/m V	n.v.t.	69	68,9	n.v.t.	98,8	98,9

Tabel 7 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 6*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	70,9	72,3	n.v.t.	99,6	99,7
vaardigheidsniveaus I t/m V	n.v.t.	67,9	69,7	n.v.t.	98,1	98,6

Tabel 8 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment eind groep 6*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	68,6	70,6	n.v.t.	99,2	99,6
vaardigheidsniveaus I t/m V	n.v.t.	65,1	67,8	n.v.t.	97,7	98,7

Tabel 9 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 7*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	65,7	69,6	n.v.t.	98,7	99,4
vaardigheidsniveaus I t/m V	n.v.t.	62,2	67,4	n.v.t.	97,1	98,7



Tabel 10 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment eind groep 7*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	66,4	70,5	n.v.t.	98,8	99,6
vaardigheidsniveaus I t/m V	n.v.t.	62,8	68,1	n.v.t.	96,6	98,6

Tabel 11 *Samenvattende indices voor verschillende afnamecombinaties van leeskaarten voor afname-moment medio groep 8*

	marginal classification accuracy			accuracy plus/minus 1 niveau		
	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3	leeskaart 1+2	leeskaart 2+3	leeskaart 1+2+3
vaardigheidsniveaus E t/m A	n.v.t.	65,4	70,4	n.v.t.	98,9	99,7
vaardigheidsniveaus I t/m V	n.v.t.	61,6	67,9	n.v.t.	97,1	98,9





**Cito**

Amsterdamseweg 13

6814 CM Arnhem

Postbus 1034

6801 MG Arnhem

T (026) 352 11 11

Fotografie: Ron Steemers

