

Leerlingvolgsysteem

**Wetenschappelijke verantwoording van de toetsserie Leestechiek & Leestempo**

**R.S.H. Krom en F.H. Kamphuis**

Leerlingvolgsysteem

## **Wetenschappelijke verantwoording van de toetsserie Leestechiek & Leestempo**

**R.S.H. Krom en F.H. Kamphuis**

augustus 2001

# Inhoud

Woord vooraf	5
Inleiding	6

## Deel 1 Wetenschappelijke verantwoording van de toetsen Leestechniek

Rationale	7	
1	Beschrijving van de toetsen Leestechniek	7
1.1	Construct	7
1.2	Doelen	7
1.3	Doelgroep	8
1.4	Samenstelling	8
1.5	Inhoud en operationalisatie	8
1.5.1	Itemvorm	8
1.5.2	Grondwoorden	9
1.5.3	Afleidings	10
1.5.4	Foutanalyse	10
1.6	Kalibratie en opgavenselectie	10
1.6.1	Meetmodel	10
1.6.2	Afnamedesign	13
1.6.3	Modeltoetsing	14
1.6.4	Itemselectie	14
2	Meetnauwkeurigheid van de toetsen Leestechniek	14
2.1	Globale meetnauwkeurigheid	14
2.2	Lokale meetnauwkeurigheid in tabelvorm	15
2.3	Lokale meetnauwkeurigheid in grafiekvorm	15
2.4	Betrouwbaarheidsintervallen	16
3	Validiteit van de toetsen Leestechniek	16
3.1	Inhoudsvaliditeit	16
3.1.1	Representativiteit van de grondwoorden	17
3.1.2	Bekendheid van de grondwoorden	17
3.2	Begripsvaliditeit	18
3.2.1	Correlaties met 'soortgenoten'	18
3.2.2	Vaardigheidstoename	19
3.3	Criteriumvaliditeit	20
4	Normering van de toetsen Leestechniek	20
4.1	Referentiegroep	20
4.1.1	Schoolgewicht	20
4.1.2	Regionale spreiding	21
4.2	Referentiegegevens	22

## Deel 2 Wetenschappelijke verantwoording van de toetsen Leestempo

Rationale	23	
5	Beschrijving van de toetsen Leestempo	23
5.1	Construct	23
5.2	Doelen	23
5.3	Doelgroep	24
5.4	Samenstelling	24

5.5	Inhoud en operationalisatie	25	
5.5.1	De lengte van de toetsen	25	
5.5.2	De moeilijkheid van de toetsen	25	
5.5.3	Toetsvorm	26	
5.5.4	Toetsscores	27	
5.5.5	Typering van leesproblemen	28	
5.6	Kalibratie en toetsselectie	28	
5.6.1	Meetmodel	28	
5.6.2	Parameterschattingen	29	
5.6.3	Afnamedesign	30	
5.6.4	Vaardigheidsscores	30	
5.6.5	Toetsselectie	31	
6	Meetnauwkeurigheid van de toetsen	Leestempo	31
6.1	Globale meetnauwkeurigheid	31	
6.2	Lokale meetnauwkeurigheid	32	
7	Validiteit van de toetsen	Leestempo	36
7.1	Inhoudsvaliditeit	36	
7.1.1	Representativiteit van het tekstmateriaal		36
7.1.2	Conceptuele moeilijkheid van het tekstmateriaal		37
7.2	Begripsvaliditeit	38	
7.2.1	Correlaties met een 'soortgenoot'	38	
7.2.2	Vaardigheidstoename	38	
7.2.3	Het effect van een instructievariant	39	
7.2.4	Het effect van raden	39	
7.3	Criteriumvaliditeit	40	
8	Normering van de toetsen	Leestempo	41
8.1	Referentiegroep	42	
8.2	Referentiegegevens	43	
<b>Deel 3 Wetenschappelijke verantwoording van de koppeling van de Leestechiek- en Leestemposchaal</b>			
9	Leestechiek en Leestempo: één doorlopende schaal		45
9.1	De 'ankering' Leestechiek – Leestempo	45	
9.2	Regressie-analyse	45	
9.3	Metten van de technische leesvaardigheid (z)onder tijdsdruk		46
9.4	Stabiliteit van de positie van leerlingen in de vaardigheidsverdelingen		46
	Literatuur	49	

## Woord vooraf

Deze verantwoording levert tezamen met de inhoud van het toetspakket Leestechneik & Leestempo (Krom, 2001) alle informatie die nodig is voor een snelle en efficiënte beoordeling van de kwaliteit van de betreffende meetinstrumenten. Het genoemde materiaal maakt een beoordeling van de toetsen in Leestechneik & Leestempo mogelijk op de volgende aspecten:

- Uitgangspunten bij de toetsconstructie
- Kwaliteit van het toetsmateriaal
- Kwaliteit van de handleiding
- Wijze van normering
- Meetnauwkeurigheid
- Validiteit

Het voorliggend document heeft met name betrekking op de uitgangspunten bij de constructie (de hoofdstukken 1, 5 en 9), de wijze van normering (de hoofdstukken 4 en 8), de meetnauwkeurigheid (de hoofdstukken 2 en 6) en de validiteit (de hoofdstukken 3 en 7 ) van de toetsen in Leestechneik & Leestempo. De kwaliteit van het toetsmateriaal en de handleiding is te bepalen door kennis te nemen van de inhoud van het toetspakket.

## Inleiding

Het meetinstrument dat in voorliggende publicatie verantwoord wordt heet Leestechniek & Leestempo. De 'dubbele' naam van het instrument geeft al aan dat het hier om een combinatie van twee toetsseries gaat: toetsen Leestechniek en toetsen Leestempo. Hoe verhouden de beide series zich tot elkaar?

De toetsserie Leestechniek is te beschouwen als een 'voorloper' (in didactische zin) op de toetsen in de toetsserie Leestempo. Dit komt tot uitdrukking in de verschillende toetsinhouden: in Leestechniek, voor leerlingen in groep 3, gaat het om het lezen van woorden; in Leestempo, voor de groepen 4 tot en met 8, om het lezen van teksten. Het komt eveneens tot uitdrukking in de onderscheiden toetsprocedures: de toetsen Leestechniek laten de leerlingen in alle rust ontsleutelen; de toetsen Leestempo introduceren een zekere tijdsdruk.

Ondanks deze verschillen liggen de scores die leerlingen op de toetsen in beide series behalen op een en dezelfde vaardigheidsschaal. Met het pakket Leestechniek & Leestempo kunnen leerkrachten de ontwikkeling van hun leerlingen dan ook volgen van groep 3 tot en met groep 8.

Voor de overzichtelijkheid worden beide toetsseries hieronder desalniettemin afzonderlijk verantwoord. Leestechniek in deel 1 en Leestempo in deel 2. De koppeling van beide vaardigheidsschalen tot één doorlopende schaal wordt verantwoord in deel 3.

# Deel 1 Wetenschappelijke verantwoording van de toetsen Leestechniek

## Rationale

Binnen het leesonderwijs op de basisschool wordt een onderscheid gemaakt tussen technisch lezen en begrijpend lezen. Het technisch lezen is geen doel op zich maar kan worden gezien als een voorwaardelijke activiteit voor het leren begrijpen van teksten. Om een tekst te kunnen begrijpen, dat wil zeggen de betekenis of betekenissen van de tekst te kunnen achterhalen, is het van belang dat de woorden in die tekst nauwkeurig en vlot ontsleuteld worden. Het efficiënt leren ontsleutelen van woorden geldt als een van de belangrijkste doelstellingen van het aanvankelijk leesonderwijs. In vrijwel alle methoden voor aanvankelijk lezen besteedt men reeds in een zeer vroeg stadium aandacht aan dit (technische) aspect van het lezen. Leerkrachten zullen dan ook van tijd tot tijd willen nagaan hoe hun leerlingen er op dit punt voor staan. Daarom zijn binnen het Cito-Leerlingvolgsysteem technisch-leestoetsen ontwikkeld, die aan het onderwijs ter beschikking gesteld worden onder de naam Leestechniek.

## 1 Beschrijving van de toetsen Leestechniek

### 1.1 Construct

De toetsen Leestechniek meten de technische leesvaardigheid van leerlingen tijdens het aanvankelijk leesonderwijs. Het gaat hier om de vaardigheid in het ontsleutelen van geschreven woorden en het herkennen van deze woorden als dragers van betekenis. Technisch lezen, op deze manier omschreven, omvat dus meer dan louter het verklanken van woorden. Het is een omschrijving die direct aansluit bij het uitgangspunt dat technisch lezen geen doel op zich is, maar voorwaardelijk voor het begrijpend lezen. Het ontsleutelen of decoderen zet het geschreven woord om in een verklanke vorm ('verklankt' betekent hier niet per se hardop!), zodat aan het woord zijn betekenis toegekend kan worden.

### 1.2 Doelen

De toetsen Leestechniek hebben twee hoofddoelen en twee nevendoelelen. De hoofddoelen zijn: niveau- en progressiebepaling; de nevendoelelen hebben betrekking op het signaleren van zwakke lezers met specifieke decodeerproblemen en het selecteren van passend leesmateriaal.

#### Niveaubepaling

De beide toetsafnamen in het kader van Leestechniek geven de leerkracht informatie over het leesvaardigheidsniveau van zijn leerlingen, individueel of als groep. Iedere behaalde leesvaardigheidsscore kan daartoe normgericht geïnterpreteerd worden op basis van de vaardigheidsverdeling in een adequate referentiegroep (zie deze publicatie: paragraaf 4.1). De referentiegroep is op basis van de scores van de leerlingen in deze groep in vijf niveaugroepen verdeeld. Niveaugroep A: de leerlingen in het bovenste kwartiel. Niveaugroepen B plus C: de leerlingen in de twee volgende kwartielen (B: de leerlingen boven het gemiddelde; C: de leerlingen onder het gemiddelde). Niveaugroepen D en E: de leerlingen in het onderste kwartiel (E: de 10% leerlingen met de laagste scores; D: de 15% leerlingen daarboven).

#### Progressiebepaling

De beide toetsen van Leestechniek geven de leerkracht informatie over de ontwikkeling van de leesvaardigheid van zijn leerlingen, individueel of als groep, gedurende de fase van het aanvankelijk lezen. Het gehanteerde meetmodel (zie deze publicatie: paragraaf 1.6.1) maakt het namelijk mogelijk om de scores van een leerling op verschillende toetsen, op verschillende momenten afgenomen, onderling te vergelijken. De ruwe scores op de beide toetsen – de aantallen opgaven goed – zijn daartoe te transformeren in scores op één vaardigheidsschaal. Deze unidimensionele vaardigheidsschaal die aan de twee toetsen in Leestechniek ten grondslag ligt is ontwikkeld met behulp van het *One Parameter Logistic Model* (Verhelst, 1993; Verhelst & Glas, 1995).

## Signalering

Het eerste neven doel van Leestechniek betreft de signalering van leerlingen met specifieke leesproblemen. Deze signaalfunctie van de toetsen wordt mogelijk gemaakt doordat bij de keuze van de afleiders (de onjuiste antwoordalternatieven) in de items van Leestechniek steeds uitgegaan is van decodeerfouten die kenmerkend zijn voor zwakke lezers in het beginstadium van het leren lezen. Omdat deze fouten onder te brengen zijn in afzonderlijke fouttypen kan een analyse van de door een leerling gemaakte fouten 'diagnostische' informatie over de onderliggende problematiek met betrekking tot het lezen van de leerling opleveren. (Zie verder in deze publicatie: paragraaf 1.5.4)

## Selectie van leesstof

Een tweede belangrijk neven doel van Leestechniek is het afstemmen van de leesstof op de leesvaardigheid. Deze afstemming kan plaatsvinden omdat de scores die leerlingen op Leestechniek behalen omgezet kunnen worden in een CILT-waarde.

CILT staat voor Cito Index voor de LeesTechniek, een index voor zowel de technische *leesvaardigheid* van lezers als de technische *leesbaarheid* of moeilijkheid van teksten (Staphorsius & Verhelst, 1997). Zoals de vaardigheidsscores van leerlingen op Leestechniek omgezet kunnen worden in een CILT, kan ook de moeilijkheid die een tekst leestechnisch gesproken heeft, uitgedrukt worden in een CILT. Met behulp van de CILT worden leesvaardigheid en leesbaarheid (= vereiste leesvaardigheid) op één schaal gebracht. Dit maakt het mogelijk om voor een leerling leestekstjes te selecteren met een moeilijkheid die afgestemd is op de leestechniek waarover hij of zij kan beschikken. (Zie verder Krom, 2001, pagina 35)

Jeugdt teksten uit boeken, tijdschriften, kranten, brochures, enzovoort worden door het Cito van een CILT voorzien en voor een deel opgenomen in een computerprogramma op cd-rom dat onder de naam Boekenwijzer aan het onderwijs ter beschikking wordt gesteld en waarmee leerkrachten hun leesonderwijs kunnen individualiseren. Het programma stelt de gebruiker in staat om uit een lijst van – op dit moment – ruim 2000 jeugdboeken een selectie te maken die past bij de technische leesvaardigheid van een leerling, zoals bepaald met behulp van de toetsen Leestechniek. Deze cd-rom maakt deel uit van het toetspakket Leestechniek & Leestempo.

## 1.3 Doelgroep

De toetsen Leestechniek zijn bedoeld voor leerlingen in groep 3 van het basisonderwijs. De toetsen zijn ook geschikt voor leerlingen in het speciaal basisonderwijs. De toetshandleiding bevat een afzonderlijk hoofdstuk met extra instructies voor het gebruik van de toetsen in het speciaal basisonderwijs (zie Krom, 2001, pagina 95 e.v.).

## 1.4 Samenstelling

Er zijn twee toetsen Leestechniek geconstrueerd: M3 en E3, respectievelijk voor leerlingen halverwege (medio) en eind groep 3. Vanwege de omvang van de toetsen (zie voor de aantallen opgaven: tabel 1.1) vindt de afname van elk van beide toetsen in tweeën plaats. De toetsen bestaan dan ook uit twee delen. Per toets is steeds één opgavenboekje beschikbaar; dit bevat de items uit beide delen van de toets, alsmede enkele voorbeeldopgaven. Naast de handleiding voor de leerkracht – met onder andere de afname-instructie en de score- en normtabellen – bevat het toetspakket een kopieerbaar groepsoverzicht en leerlingrapport, en analyseformulieren. Ook bevinden zich in de handleiding de toetssleutels ten behoeve van de handmatige scorebepaling. Behalve met de hand kunnen de toetsen in Leestechniek nagekeken worden met behulp van het Computerprogramma Leerlingvolgsysteem (Rosier, 2001). Met dit programma kunnen leerkrachten de prestaties van hun leerlingen tevens nader analyseren

Tabel 1.1 Aantallen opgaven per toets in Leestechniek

Toets	Aantal opgaven
M3	46
E3	54

## 1.5 Inhoud en operationalisatie

### 1.5.1 Itemvorm

Ieder van beide toetsen Leestechniek is een toets met de itemvorm: tekening-plus-grondwoord-plus-afleiders (zie figuur 1.1). Het grondwoord is het goede antwoord. De afleiders zijn de vier onjuiste antwoordalternatieven.

Figuur 1.1 Een voorbeelditem uit Leestechniek M3



- naam
- ram
- kraam
- raam
- maar

De leerling leest de woorden (stil) voor zich zelf – Leestechniek is een collectief afneembaar instrument – en zet een kruisje in het hokje voor een van de antwoordalternatieven.

In de tekening wordt de betekenis van het grondwoord weergegeven, maar op de tekening staat meer afgebeeld dan het begrip waarnaar het grondwoord verwijst. In de woorden van de geestelijke vader van deze itemvorm – in Leestechniek overigens niet met drie maar met vijf antwoordalternatieven uitgewerkt: *‘Daarbij staat op het plaatje niet één voorwerp afgebeeld, maar wordt een betekenisvolle context gegeven van meerdere voorwerpen, waarbij een van die voorwerpen – of een combinatie van die voorwerpen (toevoeging van de auteurs) – de betekenisreferent vormt van het grondwoord’* (Verhoeven, 1982, pagina 33). Dit is gedaan om te voorkomen dat een leerling bij het maken van een opgave de tekening als uitgangspunt kan nemen. De leerling zou in dat geval namelijk de betekenis van het grondwoord rechtstreeks kunnen afleiden uit de tekening en uitgaande van deze betekenis op basis van afzonderlijke letter-klankkoppelingen de onjuiste alternatieve woorden weg kunnen strepen. De auditieve synthese – het in de juiste volgorde samenvoegen van de afzonderlijke klanken – zou dan achterwege kunnen blijven bij het beantwoorden van opgaven in Leestechniek. En dat is uiteraard ongewenst.

Voor elk van de opgaven in de toetsen Leestechniek is nagegaan of leerlingen in de doelgroep de betekenisreferent van het grondwoord in de tekening herkennen. In de toetsen zijn geen opgaven opgenomen waarin tekeningen in dit opzicht moeilijkheden opleverden in proefversies van de toetsen.

### 1.5.2 Grondwoorden

De eenlettergrepige grondwoorden in Leestechniek M3 zijn van het type km, mk en mkm (m: medeklinker, k: klinker). De grondwoorden in Leestechniek E3 kunnen in vier categorieën worden ondergebracht:

- eenlettergrepige woorden met aan het begin of aan het eind twee medeklinkers achter elkaar (mmkm en mkmm);
- eenlettergrepige woorden met aan het begin en aan het eind twee medeklinkers achter elkaar (mmkmm);
- tweelettergrepige woorden met daarin een gesloten lettergreep en een stomme e in de andere lettergreep;
- twee- en drielettergrepige woorden met daarin een open lettergreep en een stomme e in de andere lettegre(e)p(en).

Tabel 1.2 geeft het aantal items per woordcategorie in Leestechniek E3.

Tabel 1.2 Aantal items per woordcategorie in leestechniek E3

Woordcategorie	Aantal items
eenlettergrepige woorden met mmkm of mkmm	22
eenlettergrepige woorden met mmkmm	7
tweelettergrepige woorden met een gesloten lettergreep	10
twee- of drielettergrepige woorden met een open lettergreep	15

De grondwoorden in Leestechniek zijn woorden waarvan aangenomen kan worden dat ze tot de woordenschat van het merendeel van de leerlingen in groep 3 behoren. Deze aanname steunt op gegevens over de frequentie van voorkomen van woorden in jeugdlektuur en op gegevens over de verwervingsperiode van woorden, dat wil zeggen de leeftijd waarop woorden, over het algemeen, geleerd worden. (Zie verder deze publicatie: paragraaf 3.1.2)

### 1.5.3 Afleiders

De afleiders of onjuiste antwoordalternatieven – in alle gevallen: bestaande woorden – zijn afgeleid van het grondwoord; de afwijkingen ten opzichte van het grondwoord staan voor door zwakke lezers min of meer frequent gemaakte technisch-leesfouten:

- twee van de vier alternatieven komen in principe tot stand door het verwisselen van een letter(combinatie); bijvoorbeeld: *raam* wordt *naam* door het verwisselen van 'r' en 'n'; *raam* wordt *ram* door het verwisselen van 'aa' en 'a';

In sommige opgaven beginnen zowel het grondwoord als de afleiders met een hoofdletter. Dit omdat het optreden van leesfouten ten gevolge van letterverwisseling niet beperkt hoeft te blijven tot de 'kleine' letters; ook bij het ontsleutelen van woorden die beginnen met een hoofdletter kan dit type leesfout zich manifesteren;

- twee van de vier alternatieven komen in principe tot stand door:
  - het weglaten van een letter; of door:
  - het toevoegen van een letter (*raam* wordt *kraam*); of door:
  - het dooreengooien van letters (*raam* wordt *maar*).

### 1.5.4 Foutenanalyse

Binnen de toetsen Leestechneik worden aldus twee foutsoorten onderscheiden. Dit onderscheid kan '(...) dienen als uitgangspunt voor een diagnostische interpretatieprocedure van de toetsresultaten van zwakke leerlingen'

(Verhoeven, 1982, p. 34). Technisch lezen rust op twee pijlers: kennis van de letter-klank koppelingen en de vaardigheid in het (auditief) synthetiseren van de klanken tot woorden. Op deze twee punten onderscheiden zwakke lezers zich van goede en normale lezers. Een zwakke lezer die nog niet alle letter-klank koppelingen kent, zal vaak letters verwarren. Een dergelijke lezer zal dan ook nog regelmatig letters verwisselen, dat wil zeggen in de toetsen Leestechneik veel fouten van de eerste soort maken. Een zwakke lezer die de auditieve synthetisatie nog onvoldoende beheerst, zal bij het samenvoegen van wel degelijk correct verklankte letters in veel gevallen: een letter weglaten, een letter toevoegen of letters dooreengooien, dat wil zeggen in Leestechneik veel fouten van de tweede soort maken.

Iedere opgave in Leestechneik bevat, naast het (juiste) grondwoord, vier onjuiste antwoordalternatieven of afleiders. De afwijkingen van deze afleiders ten opzichte van het grondwoord staan voor de leesfouten die hierboven beschreven zijn. Er is steeds naar gestreefd om per opgave twee afleiders op te nemen die op een fout van de eerste soort wijzen én twee afleiders die wijzen op een fout van de tweede soort). Opgaven waarbij dit niet gerealiseerd kon worden – iedere afleider moet een bestaand woord zijn; om leerlingen niet met 'verkeerde' woordbeelden te confronteren – bevatten maximaal één afleider waarin beide soorten fouten 'doorwerken'. Een dergelijke afleider kan niet bijdragen aan de signalering van het ene, dan wel het andere type zwakke lezer. Leestechneik M3 bevat zeven van deze items; Leestechneik E3 veertien.

Als een leerling opmerkelijk veel fouten maakt van dezelfde soort – vuistregel: 70 à 80% van het totaal aantal fouten – dan is dat een signaal. In de handleiding bij de toetsen wordt aangeraden om voor leerlingen met vaardigheidsscores in het onderste kwartiel een foutenanalyse uit te voeren. Leerkrachten kunnen deze analyse met behulp van het Computerprogramma Leerlingvolgsysteem (Rosier, 2001) door hun pc laten uitvoeren, voor het analyseren van de fouten 'met de hand' kunnen zij gebruik maken van speciale formulieren – voor elke toets één – die in kopieerbare vorm in het toetspakket zijn opgenomen.

## 1.6 Kalibratie en opgavenselectie

De opgaven in de toetsen Leestechneik zijn afkomstig uit de opgavenbank Leestechneik. Deze opgavenbank bevat 169 items die op een en dezelfde vaardigheidsschaal liggen; 46 van deze items werden opgenomen in Leestechneik M3 en 54 in Leestechneik E3.

Hieronder wordt nu eerst de kalibratieprocedure beschreven. Daarna verantwoorden we de definitieve selectie van opgaven uit de gekalibreerde itembank.

### 1.6.1 Meetmodel\*

Er is gekalibreerd met behulp van het *One Parameter Logistic Model* (OPLM; zie Verhelst, 1993 en Verhelst, Glas & Verstralen, 1994), een op de itemresponstheorie (IRT) gebaseerd meetmodel. Dergelijke modellen verschillen in een aantal opzichten nogal sterk van de klassieke testtheorie (Verhelst, 1993; Glas & Verhelst, 1993). Bij de klassieke testtheorie staan de toets en de toetsscore centraal. Het theoretisch belangrijkste begrip in deze theorie is de zogeheten *ware score*, de gemiddelde score die de persoon zou behalen indien de test een oneindig aantal keren onder dezelfde condities

---

\* De inhoud van paragraaf 1.6.1 is grotendeels ontleend aan Staphorsius, Verhelst & Kleintjes (2000)

zou worden afgenomen. Die notie geeft een van de belangrijkste (praktische) obstakels van deze theorie voor ons onderzoek weer: het is problematisch om toetsscores te vergelijken die verkregen zijn in een onvolledig design. De items in de opgavenbank Leestechiek zijn bestemd voor leerlingen halverwege groep 3 en voor leerlingen eind groep 3. Door het relatief grote aantal items en door de spreiding in moeilijkheid is het niet haalbaar en ook niet zinvol om alle items door alle leerlingen in ons onderzoek te laten maken. We zijn aangewezen op een onvolledig design en het is problematisch om toetsscores die via een dergelijk design verkregen zijn te vergelijken. Hoewel er methoden bestaan binnen de klassieke testtheorie om toetsscores te equivaleren (Engelen & Eggen, 1993), schiet deze benadering te kort als het gaat om de centrale vraag: hoe weten we dat de equivalering zinvol is? Op die vraag heeft de IRT een antwoord.

In de IRT staat het te meten begrip of de te meten eigenschap centraal. De IRT beschouwt het antwoord op een item als een indicator voor de mate waarin die eigenschap aanwezig is. Het verband tussen eigenschap en itemantwoord is van probabilistische aard en wordt weergegeven in de zogenaamde itemresponsfunctie. Die geeft aan hoe groot de kans is op een correct antwoord als functie van de onderliggende eigenschap of vaardigheid. Formeler: zij  $X_i$  de toevalsvariabele die het antwoord op item  $i$  voorstelt.  $X_i$  neemt de waarde 1 aan in geval van een correct antwoord en 0 in geval van een fout antwoord. Als symbool voor de vaardigheid kiezen we  $\theta$ . We wijzen erop dat  $\theta$  niet rechtstreeks observeerbaar is. Dat zijn alleen de antwoorden op de opgaven. Dat is de reden waarom  $\theta$  een 'latente' variabele wordt genoemd. (Men duidt de modellen die onder de IRT ressorteren dan ook wel aan als 'latente trek'-modellen.) De itemresponsfunctie  $f_i(\theta)$  is gedefinieerd als een conditionele kans:

$$f_i(\theta) = P(X_i = 1 | \theta) \quad (1.1)$$

Een IRT-model is een speciale toepassing van (1.1) waarbij aan de functie een meer of minder specifieke functionele vorm wordt toegekend. Een eenvoudig en zeer populair voorbeeld is het zogenaamde Rasch-model (Rasch, 1960) waarin  $f_i(\theta)$  is gegeven door:

$$f_i(\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (1.2)$$

waarin  $\beta_i$  de moeilijkheidsparameter van item  $i$  is. Dat is een onbekende grootte die geschat wordt uit de observaties. De grafiek van (1.2) is weergegeven in figuur 1.2 voor twee items,  $i$  en  $j$ , die in moeilijkheid verschillen. Deze figuur illustreert dat de itemresponsfunctie een stijgende functie is van  $\theta$ . Hoe groter de vaardigheid, des te groter de kans op een juist antwoord. Indien de latente vaardigheid precies gelijk is aan de moeilijkheidsparameter  $\beta_i$  krijgen we:

$$f_i(\beta_i) = \frac{\exp(\beta_i - \beta_i)}{1 + \exp(\beta_i - \beta_i)} = \frac{1}{1+1} = \frac{1}{2} \quad (1.3)$$

Daaruit volgt onmiddellijk een interpretatie voor de parameter  $\beta_i$ : het is de 'hoeveelheid' vaardigheid die nodig is voor de kans van precies een half om het item  $i$  juist te beantwoorden. Uit de figuur blijkt duidelijk dat voor item  $j$  en grotere vaardigheid nodig is om diezelfde kans te bereiken, maar dit is hetzelfde als te zeggen dat item  $j$  moeilijker is dan item  $i$ . We kunnen de parameter  $\beta_j$  dus terecht omschrijven als de moeilijkheidsparameter van item  $j$ . De implicatie van het bovenstaande is dat 'moeilijkheid' en 'vaardigheid' op dezelfde schaal liggen.

Formule (1.2) is geen beschrijving van de werkelijkheid, het is een hypothese over de werkelijkheid die getoetst kan worden op haar houdbaarheid. Hoe een dergelijke toetsing grofweg verloopt, maken we duidelijk aan de hand van figuur 1.2. Daaruit blijkt dat, voor welk vaardigheidsniveau dan ook, de kans om item  $j$  juist te beantwoorden steeds kleiner is dan de kans op een juist antwoord op item  $i$ . Daaruit volgt de statistisch te toetsen voorspelling dat de verwachte proportie juiste antwoorden op item  $j$  kleiner is dan op item  $i$  in een willekeurige steekproef van personen.

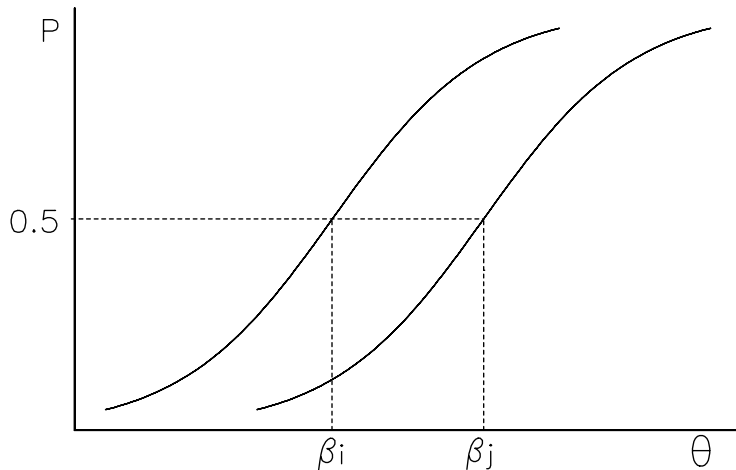
Splitst men nu een grote steekproef in twee deelsteekproeven, een 'laaggroep', met de vijftig procent laagste scores, en een 'hooggroep', met de vijftig procent hoogste scores, dan kan men nagaan of de geobserveerde p-waarden van de opgaven in beide deelsteekproeven op dezelfde wijze geordend zijn. Daarvan kan strikt genomen alleen sprake zijn als, uitgedrukt in termen van de klassieke testtheorie, alle opgaven eenzelfde discriminatie-index hebben. Dat echter blijkt lang niet altijd zo te zijn. Ook in ons geval niet. Vele van onze items blijken dan ook niet beschreven te kunnen worden met het Rasch-model. We zijn daarom op zoek gegaan naar een ander model.

Voordat we het door ons gebruikte model introduceren, moeten we eerst een opmerking maken over het schatten van de moeilijkheidsparameters in het Rasch-model.

Hoewel de literatuur verschillende schattingsmethoden beschrijft, zijn er twee die tegenwoordig veel worden gebruikt. De eerste, die wordt aangeduid als de 'marginale grootste aannemelijkheidsmethode' (in het Engels: *marginal maximum*

*likelihood*, verder afgekort als MML), veronderstelt naast (1.2) ook nog dat de vaardigheid  $\theta$  in de populatie een bepaalde verdeling heeft. De meeste computerprogramma's die IRT-analyses kunnen uitvoeren, veronderstellen een normale verdeling. Bovendien stelt deze methode de voorwaarde dat de steekproef die voor de schatting gebruikt wordt uit die verdeling een aselechte steekproef is. Is aan deze dubbele veronderstelling (normaliteit en een aselechte steekproef) niet voldaan, dan is het model verkeerd gespecificeerd. Dit heeft tot gevolg dat er systematische fouten kunnen optreden in de schattingen van de itemparameters. In ons design (zie deze publicatie: paragraaf 1.6.2) is aan de eis van aselectie zeker niet voldaan. Om die reden hebben we dan ook afgezien van de MML-schattingmethode.

Figuur 1.2 Twee itemresponscurven in het Rasch-model



De tweede vaak toegepaste schattingsmethode is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: *conditional maximum likelihood*, verder aangeduid als CML). Die maakt gebruik van het feit dat in het Rasch-model een afdoende steekproefgrootte (*sufficient statistic*) bestaat voor de latente variabele  $\theta$ , namelijk de ruwe score of het aantal correct beantwoorde items. Dat betekent grofweg dat, indien de itemparameters bekend zijn, alle informatie die het antwoordpatroon over de vaardigheid bevat, kan worden samengevat in de ruwe score; het doet er dan verder niet meer toe welke opgaven goed en welke fout zijn gemaakt. Hieruit vloeit voort dat de conditionele kans op een juist antwoord op item  $i$ , gegeven de ruwe score, een functie is die alleen afhankelijk is van de itemparameters en onafhankelijk van de waarde van  $\theta$  (zie voor een gedetailleerde uiteenzetting: Verhelst, 1992.). De CML-schattingmethode maakt van deze functie gebruik. Het belangrijkste voordeel van deze methode is dat ze geen enkele veronderstelling maakt over de verdeling van de vaardigheid in de populatie, en dat het er niet toe doet hoe de steekproef is getrokken. Zelfs als dezelfde leerling verschillende keren aan het onderzoek deelneemt vormt dit geen enkel bezwaar (Verhelst & Kleintjes, 1993). De CML-schattingmethode is echter niet bij elk meetmodel toepasbaar. Verhelst & Eggen (1989) stellen een meetmodel voor, het zogenaamde éénparameter logistisch model (*One Parameter Logistic Model*, afgekort: OPLM), waarin CML mogelijk is en dat, anders dan het Rasch-model, wel bestand is tegen 'omwisseling' van 'proporties juist' in verschillende steekproeven. De itemresponsfunctie van het OPLM is gegeven door:

$$f_i(\theta) = \frac{\exp[\alpha_i(\theta - \beta_i)]}{1 + \exp[\alpha_i(\theta - \beta_i)]} \quad (1.4)$$

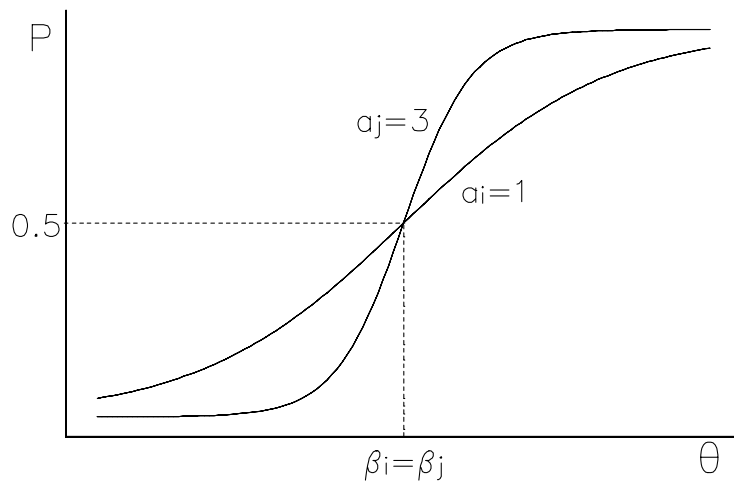
Waarin  $\alpha_i$  de zogenaamde discriminatie-index van het item is. Door deze indices te beperken tot (positieve) gehele getallen, en door ze a-priori als constanten in te voeren, is het mogelijk CML-schattingen van de itemparameters  $\beta_j$  te maken (Verhelst & Eggen, 1989). In figuur 1.3 is de itemresponscurve weergegeven van twee items  $i$  en  $j$ , die even moeilijk zijn maar verschillend discrimineren.

De schattingen worden berekend met het computerprogramma OPLM (Verhelst, Glas & Verstralen, 1994). Dit programma voert eveneens statistische toetsen uit op grond waarvan kan worden bepaald of het model de gegevens adequaat beschrijft. Omdat een aantal van deze toetsen bijzonder gevoelig is voor een verkeerde specificatie van de discriminatie-indices, zijn de uitkomsten van deze toetsen bruikbaar als modificatie-indices: ze geven een aanwijzing in welke richting deze discriminatie-indices moeten worden aangepast om een betere overeenkomst tussen model en gegevens te verkrijgen.

Kalibratie van items volgens het OPLM is dan ook een iteratief proces waarin alternerend de *model fit* van items wordt onderzocht middels statistische toetsing en de waarden van de discriminatie-indices worden aangepast op grond van de resultaten van deze toetsen. Deze aanpassingen geschieden in de praktijk op basis van een en hetzelfde

gegevensbestand. Er kan dus kanskapitalisatie optreden. Indien een steekproef een voldoende grootte heeft, is het effect van deze kanskapitalisatie echter gering (Verhelst, Verstralen & Eggen, 1991).

**Figuur 1.3** Twee itemresponscurven in het OPLM: zelfde moeilijkheid, verschillende discriminatie



Hoewel het OPLM aanzienlijk flexibeler is dan het Rasch-model, heeft het met dit model toch een nadeel gemeen, waardoor het bij het kalibreren van meerkeuze-opgaven niet zonder meer bruikbaar is. Uit de formules (1.2) en (1.4) volgt dat, indien  $\theta$  zeer klein is, de kans op een juist antwoord zeer dicht in de buurt van nul komt. Voor onze toepassing zou dit impliceren dat, indien een item voor een leerling te moeilijk is, de kans op een juist antwoord zeer dicht bij nul moet liggen. In het geval dat we met meerkeuze-opgaven te maken hebben, moeten we rekening houden met de raadkans die varieert met het aantal alternatieven. Er bestaan modellen die rekening houden met de raadkans (Lord & Novick, 1968), maar die laten geen CML-schattingmethode toe. De ongeschiktheid van het Rasch-model of OPLM voor meerkeuzevragen is echter relatief: indien de items in vergelijking met de vaardigheid van de leerling niet al te moeilijk zijn, blijkt dat het effect van het raden op de overeenkomst tussen model en gegevens niet erg groot is. Door een verstandige dataverzamelingsprocedure toe te passen kan het OPLM toch toegepast worden op meerkeuzevragen, waarbij de overeenkomst tussen model en data de uiteindelijke doorslag over die geschiktheid moet geven.

**1.6.2 Afnamedesing**

In de vorige paragraaf merkten we al op dat in het onderzoek dat aan de opgavenbank Leestechiek ten grondslag ligt, is uitgegaan van een onvolledig design: niet alle leerlingen in de steekproef maakten alle opgaven. In figuur 1.4 geven we dit design weer.

**Figuur 1.4** Weergave van het afnamedesign voor Leestechiek

		Itemverzamelingen																				
Medio 3	A	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	B	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	C	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	D	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	E	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	F	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
Eind 3	A	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	B	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	C	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	D	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	E	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█
	F	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█

Medio 3 en Eind 3: de twee afnamemomenten (halverwege en eind groep 3)  
A tot en met F: de zes substeekproeven (elk ongeveer 200 leerlingen; zie voor de exacte leerlingaantallen paragraaf 4.2)

De itemverzamelingen (de grijs gearceerde blokken) in figuur 1.4 bevatten in totaal 178 unieke opgaven. Aan elke leerling in de zes substeekproeven werden op twee afnamemomenten een of meer itemverzamelingen voorgelegd. In de figuur is te zien dat de verschillende itemverzamelingen overlap vertonen. (Op afnamemoment medio 3 bijvoorbeeld maken de leerlingen in substeekproef F de helft van de items die de leerlingen in substeekproef A maken plus de helft van de items die de leerlingen in substeekproef E maken; een gedeelte van deze items wordt op afnamemoment eind 3 ook gemaakt door respectievelijk de leerlingen in de substeekproeven B en A.) Door deze overlap wordt ervoor gezorgd dat het design verbonden is, een noodzakelijke voorwaarde om CML-schattingen van de itemparameters te kunnen bepalen.

### 1.6.3 Modeltoetsing

Ten behoeve van de toetsing van het gebruikte meetmodel wordt de totale groep van leerlingen die een bepaalde verzameling van opgaven gemaakt heeft, ingedeeld in een aantal (meestal acht) zogeheten scoregroepen. Elke groep bestaat uit leerlingen met een ongeveer even hoge score. In de toetsingsprocedure in het computerprogramma OPLM (Verhelst, Glas & Verstralen, 1994) worden de geobserveerde proporties juiste antwoorden van deze groepen afgezet tegen de proporties die we op grond van de parameterschattingen kunnen voorspellen. We besluiten tot *model fit* als de geobserveerde proporties binnen het 95%-betrouwbaarheidsinterval van de (geschatte) voorspelde proporties liggen. (Zie uitvoeriger: Verhelst, 1994).

Uit de toetsing van de gegevens verzameld met de Leestechniekitems blijkt dat 169 van de 178 opgaven in de totale itemverzameling bij het model passen. Dit is een aanduiding dat het meetinstrument dat we ontwikkeld hebben en het meetmodel dat we gebruikten toereikend zijn om het gedrag van leerlingen te verklaren. Bovendien blijkt eruit dat de hier gemeten verschillen tussen leerlingen te verklaren zijn door één unidimensionaal concept.

### 1.6.4 Itemselectie

De gekalibreerde itembank Leestechniek bevat dus 169 opgaven. Aan het begin van paragraaf 1.6 vermeldden we reeds dat 100 van deze items in de toetsen Leestechniek zijn opgenomen: 46 in M3 en 54 in E3. Bij de definitieve selectie van opgaven voor deze toetsen hanteerden we de volgende criteria:

- de mening van leerkrachten 'groep 3' over de bekendheid van de grondwoorden bij het merendeel van de leerlingen in hun groep op beide afnamemomenten (halverwege en eind groep 3);
- de mening van leerkrachten en *screeners* over de tekeningen in de opgaven: beeldt de tekening het grondwoord duidelijk en eenduidig af?;
- de mening van leerkrachten en *screeners* over het pedagogisch gehalte van de opgaven (zo is een opgave waarin een jongen veren uit de staart van een haan trekt niet geselecteerd);
- de 'diagnostische' functionaliteit van opgaven: in hoeverre maakt het item, of beter gezegd: maken de antwoordalternatieven van het item, foutenanalyse mogelijk.

De mening van de leerkrachten inventariseerden we door middel van vragenlijsten die we bij elke afname met het toetsmateriaal naar de scholen stuurden. Voor de *screeners* stelden we na de kalibratieprocedure zogeheten screeningspakketten samen.

## 2 Meetnauwkeurigheid van de toetsen Leestechniek

### 2.1 Globale meetnauwkeurigheid

Tabel 2.1 bevat informatie over de meetnauwkeurigheid van de toetsen in Leestechniek. In het vorige hoofdstuk werd onder meer aangegeven dat de leerlingen die deelgenomen hebben aan het onderzoek rond Leestechniek allen slechts een deel gemaakt hebben van de items die uiteindelijk in de toetsen opgenomen zijn. De betrouwbaarheid van de toetsen in klassieke zin is dan ook niet rechtstreeks te bepalen. Het is echter wel mogelijk om de betrouwbaarheid van iedere toets te schatten door gebruik te maken van het feit dat alle items die zijn opgenomen in de toetsen OPLM-geschaald zijn. Ook andere beschrijvende gegevens, zoals de gemiddelde score en de standaardmeetfout zijn te schatten op grond van het feit dat de toetsen volledig bestaan uit OPLM-gekalibreerde items. Om relevante beschrijvende gegevens bij de verschillende toetsen te genereren, is gebruik gemaakt van het programma OPTAL (Verstralen 1997).

Tabel 2.1 geeft allereerst aan wat de maximumscore is op iedere toets. Deze is eenvoudigweg gelijk aan het aantal opgaven dat deel uitmaakt van de toets. De derde kolom geeft de geschatte gemiddelde scores van de leerlingen op de beide toetsen. De vierde kolom bevat informatie over de geschatte standaardmeetfout van iedere toets. De laatste kolom laat de geschatte betrouwbaarheidscoëfficiënt van de beide toetsen zien. Deze coëfficiënten bedragen 0,89 en

0,92. Dit zijn waarden die acceptabel zijn voor toetsen waar, zoals in dit geval, geen zware consequenties voor leerlingen aan verbonden zijn. Zo geeft de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen) aan dat voor toetsen van dit type een betrouwbaarheidscoëfficiënt lager dan 0,70 onvoldoende is, een betrouwbaarheidscoëfficiënt tussen de 0,70 en de 0,80 voldoende, en een betrouwbaarheidscoëfficiënt hoger dan 0,80 goed (COTAN, 1999, pagina 24).

Tabel 2.1 Beschrijvende gegevens bij de toetsen Leestechiek

Toets	Maximumscore	Gemiddelde	Standaardmeetfout	Betrouwbaarheid
M3	46	40,90	1,92	0,89
E3	54	42,96	2,60	0,92

De twee betrouwbaarheidscoëfficiënten voldoen aan de COTAN-normen op het terrein van de meetnauwkeurigheid. De coëfficiënten van beide toetsen in de toetsserie verdienen de COTAN-kwalificatie 'goed'.

## 2.2 Lokale meetnauwkeurigheid in tabelvorm

De betrouwbaarheidscoëfficiënten in tabel 2.1 laten zien dat de toetsen voldoen aan de COTAN-normen wat betreft de meetnauwkeurigheid. Deze coëfficiënten hebben echter alleen betrekking op de globale meetnauwkeurigheid van de toetsen Leestechiek en geven geen beeld van de lokale meetnauwkeurigheid van de toetsen.

De zogeheten betrouwbaarheidstabellen 2.2 en 2.3 doen dat wel. Zo laat tabel 2.2 bijvoorbeeld zien dat 82,8 procent van de leerlingen die met hun geschatte vaardigheidsscore op Leestechiek M3 in scoregroep E vallen, ook met hun werkelijke vaardigheidsscore binnen deze scoregroep vallen. Anders gezegd: de kans dat een E-leerling op basis van zijn prestaties op Leestechiek M3 terecht als een E-leerling wordt bestempeld is ongeveer 83% procent.

Tabel 2.2 Betrouwbaarheidstabel bij Leestechiek M3

Scoregroep	Scoregroep waarin de ware score valt				
	E	D	C	B	A
E	82,8	24,4	0,7	0,0	0,0
D	17,0	59,3	31,8	2,4	0,1
C	0,2	15,6	51,7	33,5	7,2
B	0,0	0,7	14,4	48,0	38,0
A	0,0	0,0	1,4	16,1	54,6

Tabel 2.3 Betrouwbaarheidstabel bij Leestechiek E3

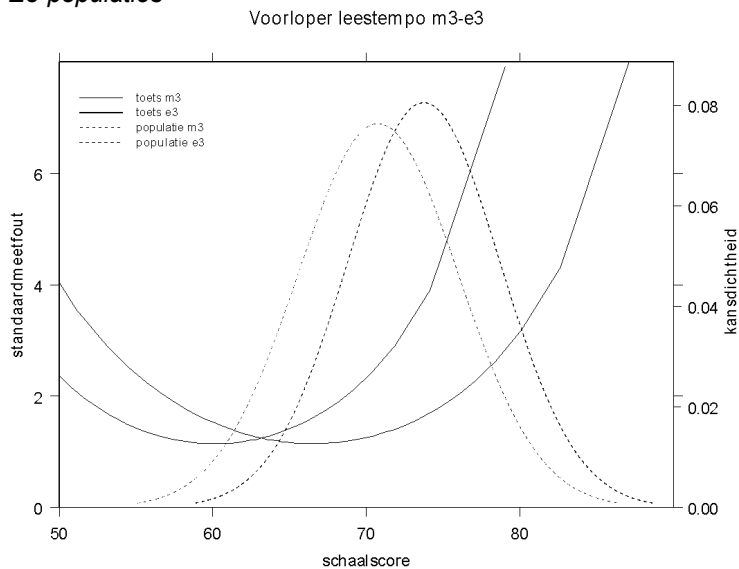
Scoregroep	Scoregroep waarin de ware score valt				
	E	D	C	B	A
E	84,4	22,3	0,3	0,0	0,0
D	15,5	64,6	29,9	1,2	0,0
C	0,1	13,0	58,9	40,3	2,3
B	0,0	0,2	10,6	49,9	36,6
A	0,0	0,0	0,3	8,5	61,1

## 2.3 Lokale meetnauwkeurigheid in grafiekvorm

Figuur 2.1 geeft nog eens grafisch weer hoe het gesteld is met de lokale meetnauwkeurigheid van de toetsen Leestechiek. In deze figuur staat voor elk van beide toetsen de grootte van de meetfout afgebeeld. De figuur maakt

duidelijk dat de meetfout het kleinst is in de lagere vaardigheidsregioenen. De toetsen concentreren zich dus op het goed in kaart brengen van de vaardigheid van de minder goede leerlingen.  
 In de figuren zijn tevens de kansdichtheidsfuncties voor de normgroepen op de twee afnamemomenten opgenomen. Deze laten zien hoe de vaardigheid van de leerlingen in de populatie verdeeld is over de vaardigheidsschaal.

**Figuur 2.1** Grootte van de meetfouten voor de toetsen Leestechneik en de kansdichtheidsfuncties voor de M3-, en E3-populaties



## 2.4 Betrouwbaarheidsintervallen

Nog gedetailleerdere informatie over de meetnauwkeurigheid van de toetsen is te vinden in de handleiding bij de toetsen. In een bijlage van de handleiding (Krom, 2001, pagina 105) bevinden zich namelijk de zogeheten score-intervaltabellen. Deze geven voor iedere ruwe score op elke van beide toetsen het 90%-betrouwbaarheidsinterval voor de bijbehorende vaardigheidsschatting.

**Tabel 2.4** Een gedeelte van de score-intervaltabellen bij de toetsen Leestechneik

Toets-score	Schaalscores	
	M3	E3
(...)	(...)	(...)
21	58 – 60	63 – 65
22	58 – 61	63 – 66
23	59 – 61	64 – 66
24	59 – 61	64 – 66
25	60 – 62	64 – 67
26	60 – 62	65 – 67
(...)	(...)	(...)

## 3 Validiteit van de toetsen Leestechneik

### 3.1 Inhoudsvaliditeit

#### 3.1.1 Representativiteit van de grondwoorden

In het kader van het onderzoek naar de inhoudsvaliditeit van de toetsen bespreken we allereerst in hoeverre de woordcategorieën in respectievelijk Leestechneik M3 en E3 zich weerspiegelen in de inhoud van het aanvankelijk leesonderwijs.

Nu zijn er meerdere methoden voor beginnende lezers in ons land, die onderling uiteraard in meer of mindere mate verschillen, qua inhoud en fasering, maar grofweg zien we toch steeds een opbouw van de lesstof langs ongeveer de volgende lijnen. In de eerste maanden leren leerlingen woorden lezen uit de categorie mkm. Na enige maanden wordt het leren lezen uitgebreid naar woorden uit de categorieën mmkm, mkmm en mmkmm (en daarmee naar de specifieke problematiek van het ontsleutelen van woorden eindigend op -d, -ng, -nk, -cht) en naar eenvoudige tweelettergrepige woorden. Naast complexere vertegenwoordigers van eerder aangeboden woordcategorieën (bijvoorbeeld de derde-persoons t na een stam op d) komen in een volgende fase van het leesonderwijs woorden uit de categorieën (m)mmm- en -mmm(m) aan bod, alsmede twee- en meerlettergrepige woorden waaronder woorden met potentiële leesmoelijkheden als open lettergrepen, verkleinvormen en voor- en achtervoegsels (ge-, be-, ver- en -ig, -lijk, enzovoort). Ter vergelijking: in Leestechiek M3, af te nemen in januari groep 3, doorgaans dus na zo'n vier maanden leesonderwijs, zijn de grondwoorden op drie na afkomstig uit de categorie mkm. Twee van de drie uitzonderingen komen uit de categorie km en één uit de categorie mk. Leestechiek E3, afnameperiode juni groep 3, introduceert naast eenlettergrepige woorden uit de categorieën mmkm, mkmm en mmkmm, tweelettergrepige woorden met gesloten en open lettergrepen (waaronder enige verkleinwoorden en woorden met een voorvoegsel) en een enkel drielettergrepig woord.

### 3.1.2 Bekendheid van de grondwoorden

In het kader van het onderzoek naar de inhoudsvaliditeit van Leestechiek M3 en E3 is ook nagegaan hoe aannemelijk het is dat de grondwoorden in de toetsen bekend zijn bij het merendeel van de leerlingen in de doelgroep. Een woordbekendheidslijst stond ons daarbij niet ter beschikking, daarom moesten we de variabele 'woordbekendheid' benaderen. We hebben twee benaderingen gehanteerd: via de gebruiksfrequentie van de woorden en via de verwervingsleeftijd van de woorden.

#### Woordbekendheid benaderd via gebruiksfrequentie

Voor deze benadering gingen we uit van de positie van de grondwoorden in Leestechiek op de frequentielijst van woordvormen naar dalende orde in Staphorsius, Krom & De Geus (1988). Dit is een van de frequentielijsten die tot stand kwam in een onderzoek naar de frequentie van voorkomen van woord(vorm)en en letter(combinatie)s in een corpus van 480 fictie- en non-fictieteksten voor de jeugd. Het totale corpus bevat ruim 200000 woorden (*types*); het aantal *tokens*, dat wil zeggen het aantal verschillende woorden, in het corpus bedraagt 18270. In tabel 3.1 staan per toets de aantallen grondwoorden per woordfrequentiepositieklasse vermeld: hoeveel procent van de grondwoorden valt in de klasse van de eerste duizend meest frequente woorden, hoeveel procent in de klasse van de tweede duizend, enzovoort.

Tabel 3.1 Aantal grondwoorden in Leestechiek per woordfrequentiepositieklasse

Woordfrequentie- positieklasse	Aantallen grondwoorden			
	in M3		in E3	
tot 1000	24%	(24%)	17%	(17%)
1000-2000	17%	(41%)	17%	(34%)
2000-3000	20%	(61%)	17%	(51%)
3000-4000	9%	(70%)	17%	(68%)
4000-5000	13%	(83%)	6%	(74%)
5000-6000	2%	(85%)	7%	(81%)

Uit tabel 3.1 blijkt uit de cumulatieve percentages tussen haakjes dat 85% (M3), respectievelijk 81% (E3) van de grondwoorden in Leestechiek tot de 6000 meest frequente woorden in bovengenoemd corpus behoren. Dat corpus bestaat, zoals gezegd, uit teksten specifiek geschreven voor de jeugd. Van de 6000 frequentste woorden in dat corpus – dat in totaal dus meer dan 18000 *tokens* omvat – mogen we vermoedelijk wel aannemen dat ze bekend zijn bij het merendeel van de leerlingen in onze doelgroep.

#### Woordbekendheid benaderd via verwervingsleeftijd

Voor deze benadering gingen we uit van de verwervingsperiode van de grondwoorden in Leestechiek, zoals weergegeven in Krom (1990). Deze publicatie maakt gebruik van gegevens van de Nieuwe Streeflijst Woordenschat (Kohnstamm, Schaerlaekens, De Vries, Akkerhuis & Frooninckx (1981) en van een serie onderzoeksrapporten van Van Loon-Vervoorn (1985a, 1985b, 1985c) en Van Loon-Vervoorn, Ridderbos & Rietveld (1985).

In het onderzoek van Kohnstamm, et al. is nagegaan hoe wenselijk leraren basisonderwijs het vinden dat bepaalde woorden door zesjarigen (passief) gekend worden. Voor ieder van de aan hen voorgelegde woorden kwam een maat voor de wenselijkheid tot stand, uitgedrukt in het percentage leraren dat begrip van het betreffende woord gewenst achtte.

Het onderzoek van Van Loon-Vervoorn maakt aannemelijk dat deze wenselijkheidsmaat, dit percentage, gehanteerd kan worden als numerieke maat voor de verwervingsleeftijd van woorden. De correspondentie tussen de percentages en de onderscheiden verwervingsperiodes zet Krom (1990) in een tabel, die we in tabel 3.2 gedeeltelijk reproduceren.

*Tabel 3.2   Wenselijkheid en verwerving*

Wenselijkheids- percentage- klasse	Verwervingsperiode in het basisonderwijs
88-100	begin groep 1
68-87	eind groep 2
49-67	groep 3
30-48	groep 4 of 5
11-29	groep 6 of 7
0-10	groep 8

Voor ieder van de grondwoorden in Leestechneik M3 en E3 is nagegaan in welke percentageklasse het woord valt. Tabel 3.3 vat het resultaat van deze inventarisatie samen.

*Tabel 3.3   Aantal grondwoorden in Leestechneik per wenselijkheidspercentageklasse*

Wenselijkheids- percentage- klasse	Aantallen grondwoorden	
	in M3	in E3
88-100	96% (96%)	78% (78%)
68-87	2% (98%)*	20% (98%)
49-67		2% (100%)

Uit tabel 3.3 blijkt uit de cumulatieve percentages tussen haakjes dat met een hoge mate van waarschijnlijkheid aangenomen mag worden dat de grondwoorden in Leestechneik bij de leerlingen in onze doelgroep bekend zullen zijn. Bijna alle grondwoorden (98% in beide toetsen) zijn immers al verworven in groep 1 of 2.

Een compleet overzicht van de grondwoorden in Leestechneik M3 en E3 met bijbehorende frequentie- en verwervingsgegevens geeft Krom (2000).

## **3.2   Begripsvaliditeit**

Met de toetsen Leestechneik M3 en E3 een (één) onderliggende latente vaardigheid die we 'de technische leesvaardigheid' mogen noemen? Een eerste indicatie in de richting van een bevestigend antwoord op deze vraag levert de toetsing van het hier gehanteerde meetmodel (zie deze publicatie: paragraaf 1.6.1). Het positieve resultaat van deze toetsing duidt erop dat aan de in Leestechneik opgenomen opgaven een unidimensionaal begrip ten grondslag ligt.

### **3.2.1   Correlaties met 'soortgenoten'**

De mate waarin de opgaven van Leestechneik vervolgens gerelateerd zijn aan 'soortgenoot'-opgaven en/of -oordelen is een tweede aanwijzing voor de begripsvaliditeit van de toetsen.

\* Omdat één van de grondwoorden van M3 (*kok*) niet aan de leraren werd voorgelegd, bereikt het cumulatieve percentage de 100% niet

Nu is Leestechiek geruime tijd na het gereedkomen van de eerste generatie toetsen in het Cito-Leerlingvolgsysteem (LVS) ontwikkeld. De leerlingen in het normerings- en kalibratie-onderzoek in het kader van de ontwikkeling van de toetsen Leestechiek maken dus geen deel uit van de oorspronkelijke LVS-steekproef; hun scores op de opgaven Leestechiek kunnen daardoor niet gerelateerd worden aan scores op andere onderdelen van het Cito-Leerlingvolgsysteem.

Wel beschikken we over gegevens uit onderzoek waarin de samenhang werd nagegaan tussen enige externe criteria en de scores op de opgaven uit de voorgangers van Leestechiek: Technisch lezen 1, 2 en 3 (Verhoeven, 1980a en 1980b; Krom, 1988). Deze opgaven verschillen weliswaar van de opgaven in Leestechiek wat het aantal antwoordalternatieven betreft – om een ‘robuustere’ signalering te bewerkstelligen werd het aantal alternatieven uitgebreid van drie naar vijf (twee afleiders per foutensoort in plaats van één) – maar dat is in dit kader geen fundamentele wijziging; het gaat in de opgaven van Leestechiek om dezelfde grondwoorden en dezelfde illustraties. In tabel 3.4 worden de correlaties gerapporteerd tussen de scores op deze opgaven en een drietal ‘soortgenoten’ plus het oordeel van de groepsleerkrachten over de leesvaardigheid van hun leerlingen.

Tabel 3.4 Correlaties met externe criteria

	M3		E3
	Versie 1	Versie 2	
Leerkrachtoordeel over de leesvorderingen	.87	.87	.78
Eén-minuut-test	.43	.56	.48
Grafemendictee	.66	.61	-
Fonemische synthesesoets	.61	.63	-

Verhoeven (1982, pagina 38) verklaart de vrij lage correlaties met de (Caesar) Eén-minuut-testscores ‘vanuit het feit, dat laatstgenoemde test voor een deel uit globaalwoorden van één onderwijsleermethode bestaat, terwijl deze woorden in de geconstrueerde toetsen juist zijn vermeden.’

Dat de samenhang met een snelheidsinstrument – de Eén-minuut-test vraagt leerlingen om binnen één minuut zoveel mogelijk woorden van een woordenlijst voor te lezen – inderdaad hoger kan zijn, laat recent eigen onderzoek zien. Halverwege groep 4 (januari 2000) namen we bij de leerlingen die betrokken waren bij de normering en de kalibratie van de Leestechiekopgaven de toets Leestempo M4 (of M4-Extra) én een ankertoets Leestechiek af. (Zie ook deze publicatie: deel 3.) De correlatie tussen de scores op beide instrumenten bedraagt: .67.

### 3.2.2 Vaardigheidstoename

Een derde indicatie voor de begripsvaliditeit van de toetsen Leestechiek is te vinden in de in het vorige hoofdstuk (figuur 2.1; paragraaf 2.3) afgebeelde kansdichtheidsfuncties van de populaties op de verschillende afnamemomenten. Tabel 3.5 geeft de daar gegeven grafische informatie over de gemiddelde vaardigheid en spreiding van de vaardigheid in de verschillende normgroepen nog eens kwantitatief weer.

Tabel 3.5 Gemiddelde vaardigheid en bijbehorende standaardafwijking voor de beide populaties

Populatie op	Gemiddelde vaardigheid	Standaardafwijking
Medio 3	70,76	5,23
Eind 3	73,73	4,95

Medio en Eind 3: de afnamemomenten (halverwege en eind groep 3)

Deze tabel bevat informatie over de gemiddelde technische leesvaardigheid, dat wil zeggen het gemiddelde van de geschatte vaardigheden, en de spreiding van de vaardigheid op de beide afnamemomenten, i.e. voor de beide populaties.

De gemiddelde leesvaardigheid van de leerlingen blijkt tussen meetmoment ‘medio groep 3’ en ‘eind groep 3’ te zijn toegenomen.

### 3.3 Criteriumvaliditeit

Het antwoord op de vraag naar de criteriumvaliditeit van de toetsen Leestechneik valt geheel samen met het antwoord op deze vraag voor de toetsen Leestempo, het evaluatief-didactisch 'vervolg' van Leestechneik. De criteriumvaliditeit van zowel Leestechneik als Leestempo wordt dan ook besproken in deel 2 van deze publicatie (in paragraaf 7.3).

## 4 Normering van de toetsen Leestechneik

De toetsen Leestechneik zijn ontwikkeld buiten het onderzoek dat is uitgevoerd in het kader van de ontwikkeling van de eerste generatie LVS-toetsen. De Leestechneikitems zijn dus afgenomen aan leerlingen in een andere steekproef dan die waarin bijvoorbeeld de Leestempotoetsen werden afgenomen.

De (gestratificeerd getrokken) steekproef van scholen en leerlingen in het onderzoek van het oorspronkelijke LVS wordt beschreven en verantwoord in deel 2 van deze publicatie. Daar wordt de representativiteit van die steekproef ten opzichte van de populatie op twee variabelen onderzocht: de samenstelling van de schoolbevolking zoals die blijkt uit de schoolscore voor de formatiebepaling (het zogeheten schoolgewicht) en de regionale spreiding.

In de beide onmiddellijk volgende paragrafen vergelijken we – op dezelfde variabelen – de steekproef van leerlingen die in 1999 en 2000 deelnamen aan het landelijk longitudinale onderzoek rond Leestechneik met de populatie ten tijde van de start van het LVS-onderzoek, de zogeheten 'startpopulatie LVS'. In de laatste paragraaf van dit hoofdstuk geven we de norm- of referentiegegevens voor de beide afnamemomenten van Leestechneik.

### 4.1 Referentiegroep

#### 4.1.1 Schoolgewicht

Ten behoeve van de samenstelling van het scholencohort dat aan het oorspronkelijke LVS-onderzoek deelnam, werd het bestand van Nederlandse basisscholen op basis van hun schoolgewicht opgesplitst in drie strata; een procedure vergelijkbaar met die in het onderzoek in het kader van de Periodieke Peiling van het Onderwijsniveau (Wijnstra, 1988, pagina 7 e.v.):

- Stratum 1: scholen met een schoolgewicht  $\leq 1,05$ ; in dit stratum zijn vooral scholen vertegenwoordigd met relatief weinig arbeiderskinderen en relatief veel kinderen uit gezinnen met een hogere sociaal-economische status.
- Stratum 2: scholen met een schoolgewicht tussen 1.06 en 1.15; de scholen in dit stratum nemen een middenpositie in tussen de beide andere strata.
- Stratum 3: scholen met een schoolgewicht  $> 1.15$ ; scholen in stratum 3 kenmerken zich door een relatief groot aantal allochtone leerlingen.

Uit deze drie strata werd een aselechte steekproef van scholen getrokken. (Voor de berekening van landelijke normen is later een wegingsprocedure toegepast.) In hoofdstuk 8 wordt de representativiteit van deze (oorspronkelijke) LVS-steekproef onderzocht. In tabel 4.1 hieronder staan per stratum het aantal en de proportie scholen in de steekproef waarin Leestechneik is afgenomen, en de proportie in de populatie ten tijde van de start van het longitudinale LVS-onderzoek, de populatie dus waaruit de steekproef is getrokken die betrokken was bij de totstandkoming van de eerste generatie LVS-toetsen, waaronder Leestempo.

Tabel 4.1 Proportie van scholen in de strata 1, 2 en 3 in de Leestechneiksteekproef en in de 'startpopulatie LVS'

Stratum	Steekproef Leestechneik		Startpopulatie LVS
	Aantal scholen	Proportie	Proportie
1	21	0,4884	0,4852
2	12	0,2791	0,3181
3	10	0,2326	0,1966
	43		

Toetsen van de nulhypothese 'de proporties in de Leestechnieksteekproef zijn gelijk aan die in de startpopulatie LVS' geeft een overschrijdingskans van .884. We verwerpen de nulhypothese niet; de verdeling over de strata verschilt in steekproef en populatie niet wezenlijk.

#### 4.1.2 Regionale spreiding

Voor de representativiteitscontrole op het gebied van de regionale spreiding is op basis van de eerste twee cijfers van de postcode een indeling in veertien aaneengesloten gebieden gemaakt. (Zie uitvoeriger: Wijnstra, 1990, pagina 69 e.v.) Deze gebiedsindeling wordt weergegeven in tabel 4.2. Bij de omvang van de gebieden is rekening gehouden met de scholendichtheid. Ieder gebied telt in totaal 6% tot 8% van de basisscholen in Nederland.

Tabel 4.2 Regionale indeling

Gebieds- code	Postcodes beginnend met	Regio- omschrijving
A	10-15	Amsterdam e.o.
B	16-20	Kop Noord-Holland
C	21-24, 27	Leiden e.o.
D	25, 26, 30, 31	Rotterdam/Den Haag
E	32, 43-49	Zeeland/W-Brabant
F	28, 29, 33, 34, 41, 42	Gorinchem e.o.
G	12, 35-39	Utrecht
H	40, 65-70, 73	Arnhem/Nijmegen
I	50-56	Midden/Oost N-Brabant
J	57-64	Limburg
K	71, 72, 74-76	Twente/Achterhoek
L	77, 79-83	NW Overijssel/ZW Drente
M	84-92	Friesland
N	78, 93-99	Groningen

In hoofdstuk 8 wordt de regionale spreiding in de LVS-steekproef onderzocht door deze te vergelijken met de spreiding in de populatie. In tabel 4.3 hieronder staan de gegevens over de regionale spreiding in de steekproef van de scholen (en leerlingen) die betrokken waren bij de ontwikkeling van de toetsen Leestechniek.

Tabel 4.3 Regionale spreiding van de scholen in de Leestechniek-steekproef

Gebieds- code	Regio- omschrijving	Aantal scholen
A	Amsterdam e.o.	1
B	Kop Noord-Holland	3
C	Leiden e.o.	4
D	Rotterdam/Den Haag	6
E	Zeeland/W-Brabant	3
F	Gorinchem e.o.	1
G	Utrecht	2
H	Arnhem/Nijmegen	3
I	Midden/Oost N-Brabant	5
J	Limburg	3
K	Twente/Achterhoek	5
L	NW Overijssel/ZW Drente	5
M	Friesland	1
N	Groningen	1

Uit iedere regio is minstens één school vertegenwoordigd. De regio's Amsterdam e.o., Gorinchem e.o., Friesland en Groningen zijn ondervertegenwoordigd; de regio Rotterdam/Den Haag is oververtegenwoordigd.

## 4.2 Referentiegegevens

Op grond van het onderzoek met de Leestechniekitems in bovenomschreven referentiegroep beschikken we over de gegevens (gemiddelden en P10, P25 en P75-waarden) waarmee we de verdeling van de vaardigheid *technisch lezen* op twee meetmomenten kunnen typeren. Deze gegevens staan in tabel 4.4.

Tabel 4.4 *Typering van de vaardigheidsverdeling*

Op afnamemoment	P10	P25	GEM	P75	Aantal leerlingen
Medio 3	64.1	67.4	70.8	74.5	1205
Eind 3	67.6	70.6	73.7	77.0	1168

Medio 3 en Eind 3: de beide afnamemomenten (halverwege en eind groep 3)  
P10, P25 en P75: het tiende, vijftiende en vijftiende percentiel  
GEM: het gemiddelde  
Aantal leerlingen: het aantal leerlingen per afnamemoment

In de handleiding bij de toetsen zijn deze waarden gekoppeld aan de in het Cito-Leerlingvolgsysteem gebruikelijke vaardigheidsniveaus A tot en met E (Krom, 2001, pagina 25). In tabel 4.5 geven we deze koppeling weer.

Tabel 4.5 *Referentiegegevens Leestechniek*

Vaardigheids- niveau	Medio 3	Eind 3
A	> 74	> 77
B	71 – 74	74 – 77
C	68 – 70	71 – 73
D	65 – 67	68 – 70
E	< 65	< 68

A: de leerlingen in het bovenste kwartiel (Q1)  
B en C: de leerlingen in Q2 en Q3 (B: de leerlingen boven het gemiddelde; C: de leerlingen onder het gemiddelde)  
D en E: de leerlingen in Q4 (E: de 10% laagstscorende leerlingen; D: de 15% leerlingen daarboven)

We verstrekken geen afzonderlijke referentiegegevens voor allochtone leerlingen omdat hun aantal in het normerings- en kalibratie-onderzoek in het kader van de ontwikkeling van de toetsen Leestechniek te gering was; het kwam de 150 niet te boven. Eenzelfde beslissing hebben we moeten nemen ten aanzien van het verstrekken van referentiegegevens voor allochtone leerlingen bij de toetsen Leestempo (zie deze publicatie: paragraaf 8.2).

## Deel 2 Wetenschappelijke verantwoording van de toetsen Leestempo

### Rationale

Binnen het leesonderwijs op de basisschool wordt een onderscheid gemaakt tussen technisch lezen en begrijpend lezen. Het technisch lezen is geen doel op zich maar kan worden gezien als een voorwaardelijke activiteit voor het leren begrijpen van teksten. Om een tekst te kunnen begrijpen, dat wil zeggen de betekenis of betekenissen van de tekst te kunnen achterhalen, is het van belang dat de woorden in die tekst nauwkeurig en vlot ontsleuteld worden. Nauwkeurig ontsleutelen is dé doelstelling in het aanvankelijk leesonderwijs; daarop richten zich dan ook de toetsen Leestechneik voor groep 3. (Zie deze publicatie: deel 1.) Daarna wordt *vlot lezen* een even belangrijke doelstelling. Vanaf het moment dat leerlingen de letter-klankkoppelingen kennen en de klanken tot woorden weten te synthetiseren, vanaf het moment met andere woorden dat van accuraat ontsleutelen sprake is, dient dit gekoppeld te zijn aan een zeker leestempo dat gestaag toeneemt over de tijd. Leerkrachten zullen daarom van tijd tot tijd willen nagaan hoe snel hun leerlingen – zonder dat zij de nauwkeurigheid uit het oog verliezen – kunnen lezen. Binnen het Cito-Leerlingvolgsysteem zijn dan ook toetsen ontwikkeld die dit aspect van de leesvaardigheid evalueren; de toetsen worden in serieverband aan het onderwijs ter beschikking gesteld onder de naam Leestempo.

## 5 Beschrijving van de toetsen Leestempo

### 5.1 Construct

De toetsserie Leestempo meet *leessnelheid* bij leerlingen in het primair onderwijs. Het gaat in Leestempo echter niet om een pure snelheidsmeting. Leestempo stelt niet vast hoeveel woorden een leerling in een bepaalde tijd kan lezen, maar stelt vast of een leerling dat wat hij in een bepaalde tijd gelezen heeft, ook *goed* gelezen heeft; althans zó goed dat hij uit kan maken dat bepaalde woorden wél in het gelezene thuishoren en andere niet. In Leestempo, met andere woorden, kan een leerling laten zien hoe vlot en nauwkeurig hij leest. Nog weer anders gezegd, is Leestempo een meetinstrument waarin de variabele *leessnelheid* gebruikt wordt om gegevens over de technische leesvaardigheid van leerlingen te verzamelen: Leestempo meet de technische leesvaardigheid onder tijdsdruk.

### 5.2 Doelen

De toetsen Leestempo hebben twee hoofd- en twee neven-doelen. De hoofd-doelen zijn: niveau- en progressiebepaling; neven-doelen hebben betrekking op het signaleren van bepaalde typen leesproblemen bij zwakke lezers en het selecteren van passende leesstof voor leerlingen.

#### Niveaubepaling

Iedere afzonderlijke toetsafname in het kader van Leestempo geeft de leerkracht informatie over het leesvaardigheidsniveau van zijn leerlingen, individueel of als groep. Iedere behaalde leesvaardigheidsscore kan daartoe normgericht geïnterpreteerd worden op basis van de vaardigheidsverdeling in een adequate referentiegroep (zie deze publicatie: paragraaf 8.1). De referentiegroep is op basis van de scores van de leerlingen in deze groep in vijf niveaugroepen verdeeld. Niveaugroep A: de leerlingen in het bovenste kwartiel. Niveaugroepen B plus C: de leerlingen in de twee volgende kwartielen (B: de leerlingen boven het gemiddelde; C: de leerlingen onder het gemiddelde). Niveaugroepen D en E: de leerlingen in het onderste kwartiel (E: de 10% leerlingen met de laagste scores; D: de 15% leerlingen daarboven).

#### Progressiebepaling

Het geheel van toetsen in de toetsserie geeft de leerkracht informatie over de ontwikkeling van de leesvaardigheid van zijn leerlingen, individueel of als groep, gedurende (bijna) de gehele basisschoolperiode. Het gehanteerde meetmodel (zie deze publicatie: paragraaf 5.6.1) maakt het namelijk mogelijk om de scores van een leerling op verschillende toetsen, op verschillende momenten afgenomen, onderling te vergelijken. De (ruwe) scores 'aantallen opgaven goed' op de verschillende toetsen zijn daartoe te transformeren in scores op één vaardigheidsschaal. Deze unidimensionele vaardigheidsschaal die aan de twaalf toetsen in Leestempo ten grondslag ligt is ontwikkeld met behulp van het *Classical Latent Additive Test Model* (Moosbrugger & Müller, 1982; Kramers, 1990).

## Signalering

Het eerste nevendoel van Leestempo betreft het signaleren van specifieke leesproblemen bij zwak lezende leerlingen. Deze signalering is mogelijk doordat iedere afname van Leestempo een dubbele indicatie oplevert; een indicatie van de vlotheid van het lezen en een indicatie van de nauwkeurigheid. Met behulp van Leestempo kunnen we drie probleemttypen op het spoor komen. In de 2x2-tabel hieronder (figuur 5.1) worden – in de cellen b, c en d – deze probleemttypen schematisch weergegeven. (Zie verder deze publicatie: paragraaf 5.5.5)

Figuur 5.1 Signalering met Leestempo

		Vlotheid	
		+	-
Nauwkeurigheid	+	a	b
	-	c	d

## Selectie van leesstof

Een belangrijk nevendoel van Leestempo is ook het afstemmen van de leesstof op de leesvaardigheid. Deze afstemming kan plaatsvinden omdat de scores die leerlingen op Leestempo behalen omgezet kunnen worden in een CILT-waarde. CILT staat voor Cito Index voor de LeesTechniek, een index voor zowel de technische *leesvaardigheid* van lezers als de technische *leesbaarheid* of moeilijkheid van teksten (Staphorsius & Verhelst, 1997). Zoals de vaardigheidsscores van leerlingen op Leestempo omgezet kunnen worden in een CILT, kan ook de moeilijkheid die een tekst leestechisch gesproken heeft, uitgedrukt worden in een CILT. Met behulp van de CILT worden leesvaardigheid en leesbaarheid (= vereiste leesvaardigheid) op één schaal gebracht. Dit maakt het mogelijk om voor een leerling leesteksten te selecteren met een moeilijkheid die afgestemd is op de leestechiek waarover hij of zij kan beschikken. (Zie verder Krom, 2001, pagina 35.)

Jeugdtteksten uit boeken, tijdschriften, kranten, brochures, enzovoort worden door het Cito van een CILT voorzien en voor een deel opgenomen in een computerprogramma op cd-rom dat onder de naam Boekenwijzer aan het onderwijs ter beschikking wordt gesteld en waarmee leerkrachten hun leesonderwijs kunnen individualiseren. Het programma stelt de gebruiker in staat om uit een lijst van – op dit moment – ruim 2000 jeugdboeken een selectie te maken die past bij de technische leesvaardigheid van een leerling, zoals bepaald met behulp van de toetsen Leestempo. Deze cd-rom maakt deel uit van het toetspakket Leestechiek & Leestempo.

## 5.3 Doelgroep

De toetsserie Leestempo is bedoeld voor leerlingen in de groepen 4 tot en met 8 van het basisonderwijs; niet voor alle achtstegroepers overigens, alleen voor de zwakke lezers onder hen. Uit het onderzoek waarbinnen de ontwikkeling van de toetsen Leestempo plaatsvond, bleek namelijk dat de leesvaardigheid van de meeste leerlingen – zoals gemeten met deze toetsen – vanaf groep 7 nog maar weinig toeneemt. Een meting in groep 8 is voor het merendeel van de leerlingen dan ook overbodig. Omdat wij er echter vermoedden dat leerkrachten de ontwikkeling van juist de zwakste lezers zo lang mogelijk, dus tot in de achtste groep, zullen willen volgen, is besloten om voor deze leerlingen toch een afname in groep 8 mogelijk te maken.

De toetsen zijn ook geschikt voor leerlingen in het speciaal basisonderwijs. Leestempo kan worden afgenomen zodra een leerling de letter-klankkoppelingen kent en de klanken tot woorden kan synthetiseren. De toetshandleiding bevat een afzonderlijk hoofdstuk met extra instructies voor het gebruik van de toetsen in het speciaal basisonderwijs (zie Krom, 2001, pagina 95 e.v.). In aanvulling daarop zijn uitgebreide informatie en handreikingen met betrekking tot het afnemen van de toetsen Leestempo bij speciale leerlingen te vinden in De Groot, Huizing & Noteboom (1999, pagina 99 e.v.)

## 5.4 Samenstelling

Leestempo bevat zes standaard toetsen: M4, E4, M5/E5, M6/E6, M7/E7 en M8. M staat voor 'medio', E voor 'eind'. De toetsen zijn achtereenvolgens bedoeld voor leerlingen halverwege groep 4, voor leerlingen eind groep 4, voor leerlingen in de groepen 5, 6 en 7, en voor zwak lezende leerlingen halverwege groep 8. De aanduiding M5/E5 bijvoorbeeld betekent dat deze toets zowel halverwege als einde leerjaar kan worden afgenomen; voor beide

meetmomenten zijn normen beschikbaar. Dit is een uitbreiding van de gebruiksmogelijkheden van de toetsen ten opzichte van de vorige versie van Leestempo (Krom, 1996).

Naast de zes standaardtoetsen bevat het pakket Leestempo zes zogeheten Extra toetsen: M4-Extra, E4-Extra, M5/E5-Extra, M6/E6-Extra, M7/E7-Extra en M8-Extra. In paragraaf 5.5.4 van deze publicatie wordt duidelijk gemaakt wanneer deze toetsen gebruikt moeten worden.

Iedere toets is in een eigen opgavenboekje opgenomen. Aan de toets gaat steeds een geïllustreerde voorbeeldtoets vooraf. Enkele voorbeelden uit de vorige versie werden in de huidige versie vervangen. Het aantal opgaven in iedere toets bedraagt 100. Het aantal opgaven in de voorbeeldtoetsen loopt op, van vijf in het voorbeeld bij M4 tot en met tien in het voorbeeld bij M8.

Naast de handleiding voor de leerkracht – met onder andere de afname-instructie en de score- en normtabellen – bevat het toetspakket een kopieerbaar groepsoverzicht en leerlingrapport. Ook bevinden zich in de handleiding de toetssleutels ten behoeve van de handmatige scorebepaling. Behalve met de hand kunnen de toetsen in Leestempo nagekeken worden met behulp van het Computerprogramma Leerlingvolgsysteem (Rosier, 2001).

## **5.5 Inhoud en operationalisatie**

Leestempo bevat groepsgewijs afneembare toetsen die de leerlingen stil lezen en niet, zoals gebruikelijk bij de meeste 'soortgenoot'-instrumenten, hardop. Dat is een essentieel verschil, aangezien stil lezen en hardop lezen tamelijk onvergelijkbare processen zijn. (Allington, 1984; Armbruster & Wilkinson, 1991; Burge, 1983; Leu Jr., 1982; Mosenthal, 1976-1977; Mosenthal, 1978).

Daarnaast verschilt Leestempo van het merendeel van de 'soortgenoot'-instrumenten waar het de aard van het te lezen materiaal betreft. Tijdens de afname van een Leestempotoets krijgen leerlingen geen rijtjes woorden, pseudo-woorden of geïsoleerde zinnen voorgelegd, maar moeten zij een 'lopende' tekst lezen; woorden in een samenhangende context. Een belangrijk verschil omdat de contextuele informatie in een tekst een rol kan spelen bij de woordherkenning: door gebruik te maken van de informatie uit de context – het deel van de tekst dat reeds verwerkt is – zouden zwakke technisch-lezers hun decodeerproblemen tot op zekere hoogte kunnen compenseren (Curtis & Glaser, 1983; Goldsmith-Phillips, 1989; Kim & Goetz, 1994; Leu Jr., DeGroff & Simons, 1986; Nicholson, 1991; Nicholson, Lilas & Rzoska, 1988; Stanovich, 1980; Stanovich, West & Freeman, 1981).

De taak waar leerlingen in de toetsen Leestempo voor staan, is dus het stil lezen van teksten. Een werkvorm die gekozen werd om zo dicht mogelijk bij het 'alledaagse lezen' te blijven. Lezen, in ieder geval buiten school, is immers nagenoeg altijd: stil lezen van teksten. Dát is het lezen waar het leesonderwijs uiteindelijk op gericht is. En in die zin draagt de in Leestempo gehanteerde toetsvorm dan ook bij aan de ecologische validiteit van het instrument.

### **5.5.1 De lengte van de toetsen**

De toetsen zijn elk gebaseerd op een tekst van ongeveer 1000 woorden. Een tekstlengte die zeker voor leerlingen in de lagere groepen van de basisschool ongebruikelijk is. Er wordt dan ook niet verwacht dat de gemiddelde leerling een tekst in de gegeven leestijd helemaal zal kunnen lezen. Voor ons doel is het zelfs gewenst dat het merendeel van de leerlingen na de gegeven leestijd niet tot het einde van de tekst gevorderd is. Wij willen namelijk leerlingen, wat betreft leesvaardigheid, onderscheiden; iets dat onmogelijk wordt als alle leerlingen ongeveer een zelfde toetsscore behalen. En dit zou in de hand gewerkt worden wanneer een voorgelegde tekst zo kort zou zijn dat ook de allertraagste lezer deze in de gegeven tijd helemaal kan lezen. Goede lezers kunnen zich dan niet onderscheiden van zwakke lezers zodat een dergelijk toetsing geen informatie over vaardigheidsverschillen op zou leveren.

Hier kan een vergelijking met alternatieve meetinstrumenten gemaakt worden: zoals er bijvoorbeeld maar weinig leerlingen in groep 4 zullen zijn die in acht minuten (de leestijd van Leestempo M4; zie deze publicatie: tabel 5.2) 1000 woorden lezen, zullen er maar weinig leerlingen in groep 4 zijn die in één minuut de 116 woorden op een leeskaart van de Eén-Minuu-Test (Brus & Voeten, 1973) of de 150 woorden op een leeskaart van de Drie-Minuten-Toets (Verhoeven, 1995) lezen. Instrumenten als de EMT, de DMT en Leestempo die leessnelheid gebruiken om gegevens te verzamelen over de leesvaardigheid van leerlingen, hebben een dergelijke ruime 'uitloop' nu eenmaal nodig. In de afname-instructie van Leestempo is overigens een zinsnede opgenomen om ontmoediging van de leerling bij het zien van deze grote hoeveelheid tekst te voorkomen.

### **5.5.2 De moeilijkheid van de toetsen**

*Moeilijkheid* staat hier voor de moeilijkheid die het begrijpend lezen van deze teksten voor leerlingen met zich meebrengt. Van M4 naar M7/E7 neemt de moeilijkheid van de in de standaard- en extra toetsen gebruikte teksten toe. De teksten die aan de basis stonden van M8 en M8-Extra, de toetsen voor zwakke lezers, hebben een moeilijkheid die ligt tussen die van de toetsen voor groep 5 en die van de toetsen voor groep 6.

Deze ordening is – nadat aan de auteurs van de op verzoek geschreven teksten gevraagd is zoveel mogelijk ‘op niveau’ te schrijven – tot stand gebracht met behulp van een door het Cito ontwikkelde leesbaarheidsformule. (Zie voor achtergrondinformatie: Staphorsius, 1992; 1994) De hier bedoelde moeilijkheid van een tekst kan worden uitgedrukt in een zogeheten CLIB, een waarde tussen 0 en 100. De afkorting CLIB staat voor: Cito LeesIndex voor het Basis- en speciaal onderwijs. De CLIB van een tekst zegt iets over de moeilijkheid of leesbaarheid van die tekst of - vanuit een ander perspectief bekeken - over de hoeveelheid leesvaardigheid die nodig is om die tekst met begrip te kunnen lezen.

In tabel 5.1 staan de CLIB-waarden die behoren bij de teksten die het uitgangspunt waren voor de constructie van de toetsen in Leestempo.

Tabel 5.1 De conceptuele moeilijkheid van de teksten in Leestempo uitgedrukt in CLIB-waarden

Toets	Tekst	CLIB
M4	<i>Wie is raar?</i>	11
E4	<i>Wilde dieren in de stad</i>	17
M5/E5	<i>De teleurstelling</i>	30
M6/E6	<i>De kat van de buurvrouw</i>	49
M7/E7	<i>Het liefste lievelingsdier</i>	63
M8	<i>Op Schiphol</i>	33
M4-Extra	<i>Het verhaal van de schrijver</i>	19
E4-Extra	<i>De fijnste en de naaste kamer</i>	21
M5/E5-Extra	<i>Charlie</i>	31
M6/E6-Extra	<i>De zes pleegkinderen</i>	42
M7/E7-Extra	<i>De Nijmeegse Vierdaagse</i>	65
M8-Extra	<i>Dertien</i>	38

In paragraaf 7.1.2 onderzoeken we in hoeverre de teksten wat hun moeilijkheid betreft passen bij het gemiddelde (begrijpend) leesvaardigheidsniveau van de leerlingen in de verschillende groepen van het basisonderwijs.

### 5.5.3 Toetsvorm

De toetsvorm van Leestempo kan in het kort omschreven worden als een tekst waarin leerlingen 'al lezende' om gemiddeld het tiende woord, maar niet vaker dan eenmaal per zeven woorden, met een keuzeprobleem worden geconfronteerd. De leerlingen krijgen, afhankelijk van de groep waarin ze zitten, een bepaalde tijd om deze tekst te lezen (zie voor de leestijd per toets: tabel 5.2). Algemeen gesproken zullen snelle lezers na de gegeven leestijd verder gevorderd zijn in de tekst dan langzame lezers.

Tabel 5.2 De leestijden in Leestempo

Toets	Leestijd
M4 en M4-Extra	8 minuten
E4 en E4-Extra	8 minuten
M5/E5 en M5/E5-Extra	7 minuten
M6/E6 en M6/E6-Extra	6 minuten
M7/E7 en M7/E7-Extra	5 minuten
M8 en M8-Extra	7 minuten

Om achteraf te kunnen vaststellen tot wáár in de tekst een leerling gevorderd is, laten we hem woorden in de tekst onderstrepen. Alle geplaatste streepjes samen geven aan hoe ver hij gekomen is. Ieder streepje geeft een keuze aan die de leerling maakte tussen drie orthografisch minimaal verschillende alternatieven: het woord dat in de tekst 'hoort' (het goede antwoord) en twee daarvan afgeleide (foute) woorden. Beide afleiders staan – vergelijk Leestechneik – voor door zwakke lezers min of meer frequent gemaakte (technisch) leesfouten:

- het verwisselen van letters, bijvoorbeeld 'r' en 'n': *starter* wordt dan gelezen als *starten*;
- het weglaten van een letter, bijvoorbeeld: *starten* wordt *staten*;

- het toevoegen van een letter, bijvoorbeeld: *staten* wordt *straten*
- het dooreengooien van letters, bijvoorbeeld: *straten* wordt *starten*.

Leerlingen zouden, om snelle lezers te lijken, lukraak woorden aan kunnen strepen. Deze leerlingen moeten achteraf onderscheiden kunnen worden van werkelijk snelle lezers. Iedere keuze voor een van de drie antwoordalternatieven moet dan ook ondubbelzinnig als een juiste of als een onjuiste keuze bestempeld kunnen worden. Het criterium daarvoor ligt in de tekst. De onjuiste alternatieven passen niet, het te kiezen woord past wél in de tekst. Het passen of niet passen van een alternatief kan spelen op syntactisch niveau (het alternatief verstoort al of niet de zinsbouw), op semantisch niveau (het alternatief past qua betekenis al of niet in de context) en op beide niveaus.

Op deze manier is in de toets een vorm van controle ingebouwd. Een dergelijke controle verstoort onvermijdelijk het normale leesproces. Van de leerling wordt, naast het lezen van een tekst, immers nóg een taak gevraagd: het maken van keuzes tussen steeds drie woorden en het 'noteren' van het resultaat van iedere keuze, door middel van het onderstrepen van een van de woorden. De uitvoering van deze extra taak echter is op zich weer een functie van de leesvaardigheid: vlotte lezers 'zien' de verschillen tussen de woorden sneller dan langzame lezers.

In figuur 5.2 ter illustratie van het bovenstaande een gedeelte van een van de toetsen. De opgaven van de toets zijn te herkennen aan de onderlijningen.

*Figuur 5.2 Een deel van de Leestempotoets voor medio groep 4 (verkleind)*

Wie is raar?

Stel je eens voor: je graaft een kuil luik kuil. Een diepe kuil.

Eén meter, twee meter meten mester, tien, honderd, duizend

meter. Je graaft steeds maar verte verder veder. Dit wordt

geen gewone kuil. Dit wordt de diepste kuil de die drie er

bestaat. Het wordt een kuil helemaal tot tol tolt de andere

kant van de wereld.

Net als in Leestechneik zijn de onjuiste antwoordalternatieven ook hier steeds 'bestaande' woorden. Dit voorkomt dat leerlingen van een Leestempotoets een soort *visual scan* taak kunnen maken. Het opnemen van non-woorden in de opgaven van Leestempo zou leerlingen namelijk de mogelijkheid bieden om goede scores te behalen zonder daarvoor de tekst te hoeven lezen: uitsluitend door per opgave op zoek te gaan naar het enige antwoordalternatief dat een woord is.

#### **5.5.4 Toetsscores**

Iedere afname van een Leestempotoets levert per leerling twee ruwe scores op. De score *aantal gelezen* en de score *aantal goed*. De score *aantal gelezen* is het aantal opgaven dat door de leerling gemaakt wordt, en is gelijk aan het volgnummer van de laatste opgave waarin onder één (of meer) van de drie antwoordalternatieven een streepje is gezet. De score *aantal goed* komt tot stand door van de score *aantal gelezen* het aantal gemaakte fouten af te trekken. Fout gerekend worden, behalve verkeerd geplaatste streepjes, ook opgaven waarin meer dan één woord onderstreept is, alsmede opgaven waarin geen van de drie woorden is onderstreept maar niet de opgaven waaraan de leerling niet is toegekomen.

De score *aantal goed* wordt omgezet in een vaardigheidsscore, welke overigens in eerste instantie alleen op de rapportageformulieren vermeld wordt als het aantal opgaven dat een leerling fout gemaakt heeft niet te groot is in vergelijking met het aantal opgaven dat hij in totaal gemaakt heeft. De handleiding bij de toets spreekt in dat geval van een acceptabele discrepantie tussen de beide ruwe scores van een leerling. De afname-instructie 'zegt' tegen de leerlingen: *Lees zo vlug als je kunt, maar niet zo vlug dat je fouten maakt*. Eigenlijk moet de leerling zijn leessnelheid dus zo kiezen dat hij de toets nog juist foutloos maakt, maar dat zou toch een wat al te zware eis zijn. Momenten van geringe concentratie of vergissingen kunnen immers ook goede lezers parten spelen. Daarom is een zekere marge ingebouwd: in de toetsen voor groep 4 moeten de leerlingen 80% of meer van de gemaakte opgaven ook goed gemaakt hebben; in de toetsen voor de hogere leerjaren: 85%. Lagere percentages duiden op niet-acceptabele discrepanties. De 85%-grens komt uit de traditie van de zogeheten criteriumtoetsen; de empirie leerde echter dat deze grens voor leerlingen halverwege groep 4 nog te streng is, zodat de grens voor toets M4 werd verlaagd tot 80%. Op basis van beide criteria bleek in het onderzoek waarin de toetsen ontwikkeld werden het aantal leerlingen met niet-acceptabele discrepanties tussen de 5% en 14% te liggen.

Leerlingen die te veel fouten in een toets gemaakt hebben – leerkrachten kunnen dit in één oogopslag aflezen in een tabel in de bijlage van de handleiding, en scoren dergelijke leerlingprestaties verder niet – krijgen op het betreffende afnamemoment nogmaals een Leestempotoets voorgelegd. Daartoe zijn in het toetspakket de Extra toetsen opgenomen. Als een leerling ook in een Extra toets te veel fouten maakt, wordt de dan behaalde vaardigheidsscore wél genoteerd op de rapportageformulieren; echter voorzien van een speciale aantekening.

### 5.5.5 Typering van leesproblemen

Op basis van de vaardigheidsscores en de discrepantiewaarden kunnen met Leestempo vervolgens drie typen probleemlezers onderscheiden worden (vergelijk deze publicatie: figuur 5.1):

- leerlingen die vlot maar met te veel fouten lezen; dit zijn leerlingen die vaardigheidsscores hoger dan die van het onderste kwartiel in de referentiegroep koppelen aan een niet-acceptabele discrepantie tussen de score *aantal gelezen* en de score *aantal goed*;
- leerlingen die nauwkeurig maar langzaam lezen; dit zijn leerlingen met vaardigheidsscores in het onderste kwartiel en een acceptabele verhouding tussen hun scores *aantal gelezen* en *aantal goed*;
- leerlingen die langzaam lezen en daarbij te veel leesfouten maken; deze leerlingen koppelen lage vaardigheidsscores (onderste kwartiel) aan niet-acceptabele discrepantiewaarden.

Hoofdstuk 9 van de handleiding bij de toetsen bevat suggesties voor hulp aan deze leerlingen. Aan een Hulpboek Leestechneek & Leestempo wordt gewerkt.

## 5.6 Kalibratie en toetsselectie

De toetsen in de toetsserie Leestempo zijn afkomstig uit de opgavenbank Leestempo. Deze opgavenbank bevat negentien Leestempotoetsen die op een en dezelfde vaardigheidsschaal liggen; twaalf van deze toetsen werden in de onderhavige toetsserie opgenomen.

Hieronder wordt nu eerst de kalibratieprocedure beschreven. Daarna verantwoorden we de uiteindelijke selectie van toetsen uit de gekalibreerde opgavenbank.

### 5.6.1 Meetmodel\*

Er is gekalibreerd met behulp van het *Classical Latent Additive Test Model* (CLAT; zie Moosbrugger & Müller, 1982). Een model met de klassieke testtheorie als basis, maar uitgebreid met de aanname dat de moeilijkheid van de toets en de vaardigheid van de leerling op één schaal zijn weer te geven, een aanname die we kennen uit de eerder (paragraaf 1.6) besproken latente trek- of IRT-modellen. CLAT is bij uitstek geschikt om te worden toegepast op toetsen als Leestempo die in de literatuur wel omschreven worden als *authentieke* toetsen (Ratz, 1985). Authentieke toetsen bestaan, algemeen gesproken, uit een of meer 'superitems' en hebben – naast uiteraard een hoog *real life* gehalte – de volgende eigenschappen:

- de items in superitems zijn onderling niet verwisselbaar;
- er kunnen geen items uit superitems worden weggelaten;
- er kunnen geen items aan superitems worden toegevoegd.

Ook in Leestempo vormen de items een onverbreekelijk geheel. Het is onmogelijk iets aan een van de items te veranderen zonder dat dit gevolgen heeft voor de toets als geheel. Op de toets is een *range* van scores mogelijk, omdat de toetsscore is samengesteld uit de scores van de afzonderlijke items, welke dichotoom gescoord zijn. Het is duidelijk dat een toets als Leestempo in zijn geheel beoordeeld moet worden. Het heeft geen zin om een analyse op het niveau van de individuele items uit te voeren. In Leestempo worden dan ook uitsluitend totaalscores van leerlingen bepaald: de score *aantal gelezen* en de score *aantal goed*.

De score *aantal goed* van een leerling  $v$  op toets  $i$  wordt gerepresenteerd door het symbool  $X_{vi}$ . Het model veronderstelt een continue reactievariabele  $Y_{vi}$ , die is gedefinieerd als:

$$Y_{vi} = f(X_{vi}) \tag{5.1}$$

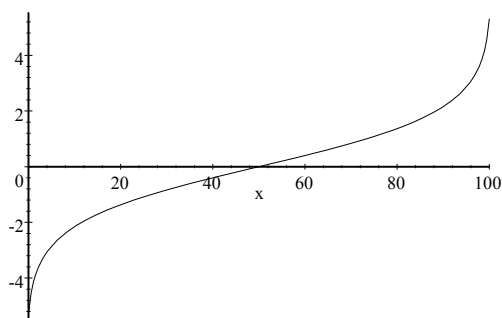
waarbij  $f$  een monotone functie voorstelt, in deze toepassing gedefinieerd als:

$$f(X_{vi}) = \ln \frac{x_{vi} + \frac{1}{2}}{100 - x_{vi} + \frac{1}{2}} \tag{5.2}$$

\* De inhoud van paragraaf 5.6.1 is grotendeels ontleend aan Bechger, Van der Lubbe & Krom (2001)

In figuur 5.3 wordt de grafiek van deze functie weergegeven.

Figuur 5.3 De grafiek van  $f(X_{vi})$



De grafiek laat zien dat de functie 0 wordt als een leerling 50 van de 100 opgaven goed heeft gemaakt. Verder is de curve symmetrisch rond  $x = 50$ , i.e.  $f(50 - a) = -f(50 + a)$ , als  $a$  een willekeurige integer is. Wanneer we aannemen dat de toets bestaat uit 100 Rasch-items die alle even moeilijk zijn, dan is  $f(X_{vi})$  een goede schatter van de latente trek (de zogeheten *Warm-schatter*).

De definitie van een *error*-variabele  $\varepsilon_{vi} \equiv Y_{vi} - E[Y_{vi}]$  geeft aanleiding tot de decompositie:

$$Y_{vi} = \tau_{vi} + \varepsilon_{vi} \tag{5.3}$$

waarbij  $E(Y_{vi}) \equiv \tau_{vi}$  de ware score van leerling  $v$  voorstelt ( $E(\varepsilon_{vi}) = 0$ ). Dit is het model van de klassieke testtheorie. Voor de ware score geldt onder het gebruikte model dat:

$$\tau_{vi} = \alpha_i (\xi_v - \beta_i) \tag{5.4}$$

De parameter  $\beta_i$  is de relatieve moeilijkheid van toets  $i$ ,  $\xi_v$  is de vaardigheid van leerling  $v$  en  $\alpha_i$  is het discriminerend vermogen van toets  $i$ . Bij het schatten wordt rekening gehouden met onvolledige data. Modeltoetsing en procedures voor het schatten van de parameters verlopen in grote lijnen volgens dezelfde principes als in OPLM (zie deze publicatie: paragraaf 1.6) en worden uiteengezet in Kramers (1990).

### 5.6.2 Parameterschattingen

De parameters  $\alpha_i$  en  $\beta_i$  zijn geschat met behulp van de gegevens van meerdere afnamen van de toetsen uit de opgavenbank Leestempo. De waarden van de parameters worden beschouwd als constant. Zie voor het afnamedesign paragraaf 5.6.3 hieronder.

In tabel 5.3 staan de discriminatie- ( $\alpha_i$ ) en moeilijkheidsparameters ( $\beta_i$ ) van de negentien toetsen in de opgavenbank.

Tabel 5.3 Parameterschattingen opgavenbank Leestempo

Toets	$\alpha_i$	$\beta_i$	Toets	$\alpha_i$	$\beta_i$
LTTO01	1,54075	- 0,40764	LTTO11	1,02581	0,08036
LTTO02	1,77765	- 0,50675	LTTO12	0,80673	-0,15063
LTTO03	1,62442	- 0,61919	LTTO13	0,67318	0,07425
LTTO04	1,60609	- 0,44930	LTTO14	0,58284	0,49437
LTTO05	1,71362	- 0,63310	LTTO15	0,52421	0,19853
LTTO06	1,14636	- 0,25456	LTTO16	0,85661	0,57745
LTTO07	1,29462	- 0,39371	LTTO17	0,83694	0,12786
LTTO08	1,40564	- 0,36646	LTTO18	0,59461	1,53073
LTTO09	1,10728	- 0,30240	LTTO19	0,51921	1,18327
LTTO10	0,93846	- 0,18309			

De schatting van de meetfout, de standaarddeviatie van de *error*-variabele ( $\varepsilon_{vi}$ ) bedraagt voor alle toetsen: 0,29921.

LTTO 1 – LTTO19: codenummers van de Leestempotoetsen in de bank  
 $\alpha_i$ : discriminatie-index  
 $\beta_i$ : relatieve moeilijkheid

### 5.6.3 Afnamedesign

De negentien toetsen in de opgavenbank Leestempo werden in het kader van de ontwikkeling van het Cito-Leerlingvolgsysteem in een landelijk longitudinaal onderzoek op negen afnamemomenten voorgelegd aan in totaal vijf substeekproeven van leerlingen.

Door het wegvallen van een deel van de subsidie moest vanaf afnamemoment 'halverwege groep 5' de omvang van het oorspronkelijke cohort gehalveerd worden. (Zie ook deze publicatie: paragraaf 8.1). Figuur 5.4 geeft het afnamedesign voor en na deze ingreep weer.

Figuur 5.4 Weergave van het afnamedesign voor Leestempo

		Toetsen																		
Medio 4	A	■	■	■																
	B				■	■	■													
	C	■	■	■																
Eind 4	A				■	■	■													
	B							■	■	■										
	C	■	■	■																
Medio 5	B*							■	■	■										
	D										■	■	■							
	E													■	■	■				
Eind 5	B*																			
	D																			
	E																			
Medio 6	B*																			
	D																			
	E																			
Eind 6	B*																			
	D																			
	E																			
Medio 7	B*																			
	D																			
	E																			
Eind 7	B*																			
	D																			
	E																			
Medio 8	B*																			
	D																			
	E																			

Medio 4 tot en met Medio 8: de negen afnamemomenten (steeds halverwege en einde groep)  
 A tot en met E: de vijf substeekproeven (A en C: elk ongeveer 750 leerlingen, B: 600 leerlingen; B\*, D en E: elk ongeveer 300 leerlingen (B\* bevat ongeveer de helft van de leerlingen uit B); zie voor de exacte leerlingaantallen paragraaf 8.2)

Per afnamemoment maakte iedere leerling steeds twee Leestempotoetsen; een leerling krijgt een toets slechts één keer voorgelegd. Het design maakt zowel horizontale linking (tussen toetsen op één afnamemoment) als verticale linking (tussen toetsen op opeenvolgende afnamemomenten) mogelijk.

### 5.6.4 Vaardigheidsscores

Per toetsafname wordt de vaardigheid van de leerling berekend uit de score *aantal goed*, als volgt:

- Bereken  $y_{vi} = f(x_{vi})$ . Dit geeft een schatting van  $E(Y_{vi}) = \alpha_i (\xi_v - \beta_i)$
- Hieruit krijgen we een schatting van  $\hat{\xi}_v = \frac{y_{vi} + \alpha_i \beta_i}{\alpha_i}$
- Vervolgens kan de vaardigheid naar iedere willekeurige schaal overgebracht worden.

De uiteindelijke schaal- of vaardigheidsscore wordt als volgt berekend:

$$SS = 15(\hat{\xi}_v + 6,21)$$

Deze transformatie is zo gekozen dat de scores op het soort 'handzame' schaal komt te liggen dat binnen het Cito-Leerlingvolgsysteem gebruikelijk is.

### 5.6.5 Toetsselectie

De gekalibreerde opgavenbank Leestempo bevat zoals gezegd negentien toetsen. Aan het begin van paragraaf 5.6 vermeldden we reeds dat twaalf van deze negentien toetsen in de uiteindelijke toetsserie werden opgenomen. Bij de definitieve selectie van toetsen hanteerden we de volgende criteria:

- de betrouwbaarheidscoëfficiënten van de toetsen; we besloten om alleen toetsen met een betrouwbaarheidscoëfficiënt hoger dan 0,80 te selecteren (zie verder hoofdstuk 6 van deze publicatie);
- de mening van leerkrachten en *screeners* over de geschiktheid van de teksten op basis waarvan de toetsen ontwikkeld werden; 'geschiktheid' betreft hier niet zozeer toetstechnische kwaliteiten, als wel tekstinhoudelijke zaken als: moeilijkheid voor de betreffende doelgroep, 'amusementswaarde' en pedagogische en/of didactische geschiktheid qua onderwerp, taalgebruik en dergelijke.

De mening van de leerkrachten inventariseerden we door middel van vragenlijsten die we bij elke afname met het toetsmateriaal naar de scholen stuurden. Voor de *screeners* stelden we na de kalibratieprocedure zogeheten screeningspakketten samen.

## 6 Meetnauwkeurigheid van de toetsen Leestempo

### 6.1 Globale meetnauwkeurigheid

Tabel 6.1 bevat informatie over de meetnauwkeurigheid van de toetsen in Leestempo. De laatste kolom laat zien wat de geschatte betrouwbaarheidscoëfficiënt van de verschillende toetsen is. Deze varieert van 0,83 tot 0,93. Dit zijn waarden die acceptabel zijn voor toetsen waar, zoals in dit geval, geen zware consequenties voor leerlingen aan verbonden zijn. Zo geeft de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen) aan dat voor toetsen van dit type een betrouwbaarheidscoëfficiënt lager dan 0,70 onvoldoende is, een betrouwbaarheidscoëfficiënt tussen de 0,70 en de 0,80 voldoende, en een betrouwbaarheidscoëfficiënt hoger dan 0,80 goed (COTAN, 1999, pagina 24).

Tabel 6.1     *Standaardmeetfout en betrouwbaarheidscoëfficiënten in Leestempo*

Toets	Standaardmeetfout	Betrouwbaarheid
M4	2,8	0,869
E4	2,6	0,886
M5/E5; meetmoment medio groep 5	2,3	0,906
M5/E5; meetmoment eind groep 5	2,3	0,927
M6/E6; meetmoment medio groep 6	4,4	0,850
M6/E6; meetmoment eind groep 6	4,4	0,877
M7/E7; meetmoment medio groep 7	5,2	0,914
M7/E7; meetmoment eind groep 7	5,2	0,858
M8	3,9	0,879
M4-Extra	2,5	0,888
E4-Extra	2,8	0,872
M5/E5-Extra; meetmoment medio groep 5	3,2	0,834
M5/E5-Extra meetmoment eind groep 5	3,2	0,869
M6/E6-Extra meetmoment medio groep 6	4,8	0,826
M6/E6-Extra meetmoment eind groep 6	4,8	0,857
M7/E7-Extra meetmoment medio groep 7	5,4	0,911
M7/E7-Extra meetmoment eind groep 7	5,4	0,853
M8-Extra	4,1	0,871

Alle betrouwbaarheidscoëfficiënten voldoen aan de COTAN-normen op het terrein van de meetnauwkeurigheid. De coëfficiënten van alle toetsen in de toetsserie verdienen de COTAN-kwalificatie 'goed'.

## 6.2 Lokale meetnauwkeurigheid

De betrouwbaarheidscoëfficiënten laten zien dat de toetsen voldoen aan de COTAN-normen op het terrein van de meetnauwkeurigheid. Ze hebben echter alleen betrekking op de globale meetnauwkeurigheid van de toetsen Leestempo en geven geen beeld van de lokale meetnauwkeurigheid van de toetsen. De zogeheten betrouwbaarheidstabellen 6.2 tot en met 6.19 doen dat wel. Zo laat tabel 6.2 bijvoorbeeld zien dat 83,0 procent van de leerlingen die met hun geschatte vaardigheidsscore op Leestempo M4 in scoregroep E vallen, ook met hun werkelijke vaardigheidsscore binnen deze scoregroep vallen. Anders gezegd: de kans dat een E-leerling terecht als een E-leerling wordt bestempeld is 83% procent.

Tabel 6.2 *Betrouwbaarheidstabel bij Leestempo M4*

Scoregroep	Scoregroep waarin de ware score valt				
	E	D	C	B	A
E	83,0	16,6	,8	0,0	0,0
D	15,9	55,3	18,1	1,1	0,0
C	1,1	26,5	57,7	22,5	,6
B	0,0	1,6	22,2	56,0	11,3
A	0,0	0,0	1,1	20,4	88,0

Tabel 6.3 *Betrouwbaarheidstabel bij Leestempo E4*

Scoregroep	Scoregroep waarin de ware score valt				
	E	D	C	B	A
E	84,5	16,8	,7	0,0	0,0
D	14,6	55,9	17,8	,9	0,0
C	,9	25,8	58,8	22,5	,6
B	0,0	1,4	21,6	56,1	10,5
A	0,0	0,0	1,1	20,4	89,0

Tabel 6.4 *Betrouwbaarheidstabel bij Leestempo M5/E5 op meetmoment medio groep 5*

Scoregroep	Scoregroep waarin de ware score valt				
	E	D	C	B	A
E	86,3	15,9	,5	0,0	0,0
D	13,3	60,1	17,5	,6	0,0
C	,5	23,1	60,7	22,4	,4
B	0,0	,9	20,4	56,5	9,1
A	0,0	0,0	,9	20,6	90,5

Tabel 6.5 *Betrouwbaarheidstabel bij Leestempo M5/E5 op meetmoment eind groep 5*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	87,9	20,8	1,1	0,0	0,0
D	11,0	49,6	13,9	,2	0,0
C	1,1	29,1	66,8	23,2	,4
B	0,0	,6	17,4	55,2	8,2
A	0,0	0,0	,8	21,3	91,4

---

Tabel 6.6 *Betrouwbaarheidstabel bij Leestempo M6/E6 op meetmoment medio groep 6*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	77,7	21,6	3,0	,1	0,0
D	17,4	39,9	16,3	1,8	0,0
C	4,8	35,2	54,9	26,9	2,0
B	,2	3,2	22,6	46,4	13,4
A	0,0	,1	3,2	24,8	84,5

---

Tabel 6.7 *Betrouwbaarheidstabel bij Leestempo M6/E6 op meetmoment eind groep 6*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	77,3	17,4	1,3	0,0	0,0
D	19,5	47,2	15,2	,7	0,0
C	3,1	33,6	60,6	23,0	,8
B	0,0	1,8	22,0	57,5	15,4
A	0,0	0,0	,9	18,7	83,8

---

Tabel 6.8 *Betrouwbaarheidstabel bij Leestempo M7/E7 op meetmoment medio groep 7*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	76,0	15,3	,8	0,0	0,0
D	21,2	49,6	13,9	,4	0,0
C	2,8	33,9	64,1	23,5	,9
B	0,0	1,2	20,2	56,9	15,0
A	0,0	0,0	,9	19,1	84,1

---

Tabel 6.9 Betrouwbaarheidstabel bij Leestempo M7/E7 op meetmoment eind groep 7

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	74,7	19,8	2,5	,1	0,0
D	20,1	41,5	16,7	2,0	0,0
C	5,0	35,1	54,4	28,0	2,5
B	,2	3,4	22,5	44,3	14,1
A	0,0	,1	3,8	25,5	83,3

---

Tabel 6.10 Betrouwbaarheidstabel bij Leestempo M8

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	77,6	13,6	,5	0,0	0,0
D	21,0	56,2	16,0	,7	0,0
C	1,4	28,9	61,1	22,6	,7
B	0,0	1,3	21,6	58,0	14,3
A	0,0	0,0	,8	18,7	85,0

---

Tabel 6.11 Betrouwbaarheidstabel bij Leestempo M4-Extra

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	84,6	15,7	,5	0,0	0,0
D	14,7	58,4	17,2	,7	0,0
C	,7	24,8	60,8	21,3	,4
B	0,0	1,1	20,9	59,1	10,5
A	0,0	0,0	,7	18,9	89,1

---

Tabel 6.12 Betrouwbaarheidstabel bij Leestempo E4-Extra

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	83,4	17,5	1,0	0,0	0,0
D	15,3	53,6	18,5	1,2	0,0
C	1,2	27,0	56,6	23,4	,7
B	0,0	1,9	22,4	54,0	11,0
A	0,0	0,0	1,5	21,4	88,3

---

Tabel 6.13 *Betrouwbaarheidstabel bij Leestempo M5/E5-Extra op meet moment medio groep 5*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	81,0	19,1	1,9	,1	0,0
D	16,7	49,4	20,8	2,6	,1
C	2,2	27,9	49,7	26,2	1,7
B	,1	3,5	23,9	45,6	11,3
A	0,0	,1	3,6	25,5	86,9

---

Tabel 6.14 *Betrouwbaarheidstabel bij Leestempo M5/E5-Extra op meet moment eind groep 5*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	83,2	24,3	3,3	,1	0,0
D	13,1	39,3	16,1	1,4	0,0
C	3,5	33,8	56,5	27,8	1,6
B	,1	2,5	20,9	44,2	10,2
A	0,0	,1	3,2	26,5	88,1

---

Tabel 6.15 *Betrouwbaarheidstabel bij Leestempo M6/E6-Extra op meet moment medio groep 6*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	75,7	22,1	3,8	,2	0,0
D	17,9	37,3	16,7	2,5	,1
C	6,0	36,1	51,9	27,9	2,7
B	,4	4,4	23,2	43,6	14,2
A	0,0	,2	4,3	25,8	83,1

---

Tabel 6.16 *Betrouwbaarheidstabel bij Leestempo M6/E6-Extra op meet moment eind groep 6*

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	75,2	18,0	1,8	0,0	0,0
D	20,4	44,4	15,7	1,2	0,0
C	4,3	35,0	58,0	24,2	1,3
B	,1	2,6	23,1	54,8	16,5
A	0,0	0,0	1,4	19,8	82,2

---

Tabel 6.17 Betrouwbaarheidstabel bij Leestempo M7/E7-Extra op meet moment medio groep 7

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	75,4	15,5	,9	0,0	0,0
D	21,5	48,8	14,1	,5	0,0
C	3,0	34,4	63,4	23,9	,9
B	0,0	1,3	20,6	56,1	15,3
A	0,0	0,0	1,0	19,5	83,7

---

Tabel 6.18 Betrouwbaarheidstabel bij Leestempo M7/E7-Extra op meet moment eind groep 7

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	74,2	19,9	2,7	,1	0,0
D	20,3	40,9	16,8	2,2	,1
C	5,3	35,4	53,6	28,5	2,7
B	,3	3,7	22,7	43,5	14,3
A	0,0	,1	4,1	25,8	82,9

---

Tabel 6.19 Betrouwbaarheidstabel bij Leestempo M8-Extra

---

Scoregroep waarin de ware score valt					
Scoregroep	E	D	C	B	A
E	76,8	13,8	,6	0,0	0,0
D	21,5	55,0	16,3	,8	0,0
C	1,6	29,6	60,0	23,1	,8
B	0,0	1,6	22,1	56,9	14,8
A	0,0	0,0	1,0	19,2	84,4

---

## 7 Validiteit van de toetsen Leestempo

### 7.1 Inhoudsvaliditeit

#### 7.1.1 Representativiteit van het tekstmateriaal

In het kader van het onderzoek naar de inhoudsvaliditeit van de toetsen zijn we eerst nagegaan in hoeverre de in de toetsen gebruikte teksten, welke – op één na – op ons verzoek geschreven zijn, beschouwd mogen worden als representanten van de populatie van teksten die aangeduid kan worden met de term ‘jeugdlectuur’. We vergelijken daartoe in eerste instantie een drietal oppervlaktekenmerken van de teksten in de toetsen met die van de teksten in het zogeheten P335-corpus in Staphorsius (1994). Dit corpus kan beschouwd worden als een representatieve steekproef uit bedoelde populatie en Staphorsius (1994, pagina 242) geeft voor deze steekproef enige beschrijvende

maten; onder andere: de gemiddelde woordlengte in letters, de gemiddelde zinslengte in woorden en het percentage frequente woorden, waarin 'frequent' gedefinieerd is als een frequentie van voorkomen van 20 of vaker in het corpus van Staphorsius, Krom & De Geus (1988). In tabel 7.1 staat een samenvatting.

Tabel 7.1 Range en gemiddelde van drie oppervlaktekenmerken voor de teksten in het P335-corpus

	Range	Gemiddelde
Gemiddelde woordlengte in letters	3.84 - 5.49	4.66
Gemiddelde zinslengte in woorden	5.85 - 26.88	12.66
Percentage frequente woorden	58.69 - 91.75	77.22

Over deze maten beschikken we ook voor de teksten in de toetsen Leestempo; zij staan in tabel 7.2.

Tabel 7.2 Waarden van de drie oppervlaktekenmerken voor de teksten in Leestempo

Toets	Tekst	Gemiddelde woordlengte	Gemiddelde zinslengte	Percentage frequente woorden
M4	<i>Wie is raar?</i>	3.80	7.21	87.78
E4	<i>Wilde dieren in de stad</i>	4.01	6.99	84.43
M5/E5	<i>De teleurstelling</i>	4.29	7.74	81.49
M6/E6	<i>De kat van de buurvrouw</i>	4.05	13.68	82.31
M7/E7	<i>Het liefste lievelingsdier</i>	4.76	15.76	74.89
M8	<i>Op Schiphol</i>	4.36	8.01	82.12
M4-Extra	<i>Het verhaal van de schrijver</i>	4.12	6.83	85.56
E4-Extra	<i>De fijnste en de naaste kamer</i>	3.97	7.08	82.12
M5/E5-Extra	<i>Charlie</i>	4.07	7.68	82.77
M6/E6-Extra	<i>De zes pleegkinderen</i>	4.15	11.28	83.82
M7/E7-Extra	<i>De Nijmeegse Vierdaagse</i>	4.73	16.68	75.44
M8-Extra	<i>Dertien</i>	4.33	9.31	81.37

De gemiddelde woordlengte in letters van een van de teksten in Leestempo (*Wie is raar?*) is iets lager dan de minimale waarde in het P335-corpus. Alle andere waarden blijven binnen de in dat corpus gevonden ranges. Wat de oppervlaktekenmerken betreft zijn de teksten in Leestempo goed vergelijkbaar met teksten in de populatie 'jeugdlectuur'.

### 7.1.2 Conceptuele moeilijkheid van het tekstmateriaal

In het kader van het onderzoek naar de inhoudsvaliditeit van de toetsen hebben we vervolgens bekeken in hoeverre de teksten in Leestempo qua conceptuele of begripsmoeilijkheid passen bij het gemiddelde (begrijpend) leesvaardigheidsniveau van de leerlingen in de verschillende groepen van het basisonderwijs. Deze controle voeren we uit met behulp van de CLIB, de Cito LeesIndex voor het Basis- en speciaal onderwijs (Staphorsius, 1992; 1994). In de CLIB kunnen we de moeilijkheid van de gebruikte teksten én de leesvaardigheid van leerlingen uitdrukken. In tabel 5.1 (deze publicatie, paragraaf 5.5.2 namen we reeds de CLIB-waarden van de teksten in Leestempo op. De CLIB kan ook omgezet worden in een Clib-niveau; een aanduiding van de gemiddelde leesvaardigheid aan het einde van een bepaald leerjaar. In tabel 7.3 staat een gedeelte van een in Staphorsius & Krom, 1998, pagina 32) gepresenteerde tabel.

Vergelijking van de CLIB-waarden in tabel 5.1 met de CLIB-categorieën in tabel 7.3 leert dat de vier toetsen voor de leerlingen van groep 4 redelijk tot goed op niveau zijn, dat de twee toetsen voor groep 5 beter passen bij het gemiddelde leesvaardigheidsniveau halverwege het leerjaar dan bij het eindejaarsniveau, dat M6/E6, M7/E7 en M7/E7-Extra beter passen bij het eindejaars- dan bij het medianiveau, dat M6/E6-Extra beter past bij het medio- dan bij het eindejaarsniveau en dat M8 en M8-Extra – voor zwakke lezers – een leesvaardigheid vragen die gemiddeld aangetroffen wordt bij leerlingen eind groep 5, begin groep 6.

Tabel 7.3 Van CLIB naar Clib-niveau

CLIB	Clib-niveau	Aanduiding Clib-niveau
8-20	4	direct om de gemiddelde CLIB eind groep 4
21-35	5	direct om de gemiddelde CLIB eind groep 5
36-48	6	direct om de gemiddelde CLIB eind groep 6
49-61	7	direct om de gemiddelde CLIB eind groep 7
62-74	8	direct om de gemiddelde CLIB eind groep 8

## 7.2 Begripsvaliditeit

Meten de toetsen Leestempo een (één) onderliggende vaardigheid die we 'de technische leesvaardigheid' (onder tijdsdruk bepaald) mogen noemen? Een eerste indicatie in de richting van een bevestigend antwoord op deze vraag levert de toetsing van het hier gehanteerde meetmodel (zie deze publicatie: paragraaf 5.6.1). Het positieve resultaat van deze toetsing duidt erop dat aan de in de toetserie Leestempo opgenomen toetsen een unidimensionaal begrip ten grondslag ligt.

### 7.2.1 Correlaties met een 'soortgenoot'

De mate waarin deze toetsen vervolgens gerelateerd zijn aan een 'soortgenoot'-instrument is een tweede aanwijzing voor de begripsvaliditeit van de toetsen. In tabel 7.4 worden correlaties gerapporteerd tussen de scores op Leestempo en de Drie-Minuten-Toets (Verhoeven, 1995) op de gemeenschappelijke meetmomenten.

Tabel 7.4 Correlaties tussen Leestempo en de Drie-Minuten-Toets

Afname-moment	Correlatie-coëfficiënt
Medio 4	.77
Eind 4	.78
Medio 5	.74
Medio 6	.73
Medio 7	.69

Hoewel Leestempo en de Drie-Minuten-Toets (DMT) beide 'leessnelheidsinstrumenten' zijn – beide gebruiken leessnelheid om leerkrachten een beeld te geven van de technische leesvaardigheid van hun leerlingen – verschillen zij op twee belangrijke punten: DMT laat hardop en Leestempo laat stil lezen. En DMT laat leerlingen woorden lezen, terwijl Leestempo teksten laat lezen. De gevonden samenhang is dan ook volgens verwachting niet perfect, maar toch redelijk hoog.

### 7.2.2 Vaardigheidstoename

Een derde indicatie voor de begripsvaliditeit van de toetsen in Leestempo is te vinden in tabel 7.5. Deze tabel bevat informatie over de gemiddelde technische leesvaardigheid, dat wil zeggen het gemiddelde van de geschatte vaardigheden, en de spreiding van de vaardigheid op de verschillende afnamemomenten, i.e. voor de verschillende populaties.

De gemiddelde leesvaardigheid van de leerlingen blijkt van afname tot afname gestaag toe te nemen. Wel maken bovenstaande gegevens duidelijk dat de groei vanaf meetmoment Medio 7 aanzienlijk kleiner is dan vóór dat afnamemoment. Daarom ook staat in de handleiding bij de toetsen (Krom, 2001, pagina 41): *'Een meting in groep 8 is voor het merendeel van de leerlingen dan ook overbodig. Het is echter goed voorstelbaar dat u en uw collega's de ontwikkeling van juist de langzaamste lezers zo lang mogelijk, dus tot in de achtste groep, willen volgen. vandaar dat besloten is om voor deze leerlingen een afname in groep 8 mogelijk te maken.'* De geleverde toetsen voor afnamemoment 'medio 8' zijn met andere woorden uitsluitend bedoeld voor de zwakke lezers in groep 8. We

gebruiken voor deze afname overigens Leestempotoetsen gebaseerd op teksten die wat hun begripsmoeilijkheid betreft het begrijpend-leesvaardigheidsniveau van de gemiddelde eind vijfde-, begin zesde-groeper vraagt.

Tabel 7.5 Gemiddelde vaardigheid en bijbehorende standaardafwijking voor de verschillende populaties

Populatie op	Gemiddelde vaardigheid	Standaardafwijking
Medio 4	79,7	7,1
Eind 4	83,2	7,3
Medio 5	85,6	7,2
Eind 5	88,6	8,2
Medio 6	92,1	10,4
Eind 6	95,2	11,7
Medio 7	99,6	17,1
Eind 7	100,8	12,9
Medio 8	101,3	10,5

Medio 4 tot en met Medio 8: de negen afnamemomenten (steeds halverwege en einde groep)

### 7.2.3 Het effect van een instructie-variant

De toetsserie Leestempo probeert leerkrachten inzicht te geven in het technisch-leesvaardigheidsniveau van hun leerlingen door gebruik te maken van 'leessnelheid'. Individuele lezers houden er echter verschillende leessnelheden op na. Welke leessnelheid een lezer op enig moment kiest hangt af van zijn *mental set*, de instelling-op-de-taak die hij tijdens het lezen 'opbrengt'. Toetsontwikkelaars willen uiteraard dat de leerlingen die hun toets maken onderling niet al te zeer verschillen wat betreft die instelling-op-de-taak. Onder andere daarom gaan toetsen vergezeld van een eenduidige afname-instructie; een noodzakelijkheid voor alle toetsen, maar zeker voor toetsen die leessnelheid gebruiken om de leestechiek van leerlingen te meten, zoals Leestempo.

Maar hoe eenduidig de afname-instructie ook is, de wijze waarop bij het afnemen van toetsen uit Leestempo de aanwijzingen gegeven worden, zal van leerkracht tot leerkracht verschillen. De een zal zich bijvoorbeeld strikt houden aan de letterlijke tekst van de afname-instructie in de toetshandleiding, de ander zal daar vrijelijk op improviseren. Dit kan van invloed zijn op de instelling-op-de-taak van hun leerlingen. Toetsen moeten daar – binnen de grenzen van het redelijke, natuurlijk – tegen kunnen. De opzet van Leestempo moet zodanig robuust zijn, dat de metingen die met het instrument verricht worden bestand zijn tegen invloeden van varianten van de afname-instructie. Als daaraan niet voldaan is, zou dat een aantasting betekenen van de validiteit van het instrument. Deze potentiële gevoeligheid moet dus worden nagegaan. Dit is gebeurd in een onderzoek waarin het variëren op 'het thema van de standaard afname-instructie' gesimuleerd werd door een tweetal Leestempotoetsen aan basisscholieren voor te leggen onder gebruikmaking van verschillende afname-instructies (Krom, Verhelst & Kamphuis, 2000). Leerlingen in de experimentele conditie van dat onderzoek werden aangezet tot sneller lezen dan normaal. Het effect van deze instructie bleek zich slechts uit te strekken tot de nauwkeurigheidsmaat (de verhouding *aantal goed / aantal gelezen*): leerlingen gingen niet vlugger lezen, wel nam het aantal leesfouten toe. In de praktijk zou dit dus als enige gevolg hebben dat enkele leerlingen ten onrechte in aanmerking zouden komen voor een extra toetsafname (zie deze publicatie: paragraaf 5.5.4). Een inefficiëntie, uiteraard, maar het betrof in het hierboven beschreven onderzoek dan ook een nogal brute schending van de standaard afnameconditie. Welke schending, voor de goede orde, dus niet leidde tot kunstmatig totstandgekomen 'te hoge' leessnelheden.

### 7.2.4 Het effect van raden

Een belangrijke vraag in het kader van de begripsvalidering van de toetsen Leestempo is ook de volgende. Kunnen leerlingen zonder de tekst echt te lezen, als het ware springend van item naar item, min of meer door raden dus, toch een redelijke score behalen? Om daarachter te komen zetten we onderzoek op, waarin de scores op onvolledige versies van Leestempo vergeleken werden met de scores op corresponderende reguliere versies (Krom, Verhelst & Kamphuis, in voorbereiding). De onvolledige versies bevatten uitsluitend de opgaven en hun onmiddellijke context, dat wil zeggen het woord dat in hun reguliere pendant vóór de opgave staat en het woord dat erna komt. We gingen er vanuit dat de onvolledige versie aldus de informatie bevat waar potentiële raders onder de leerlingen gebruik van maken om een Leestempotoets te maken.

In tabel 7.6 staan de gemiddelde ruwe scores *aantal gelezen* en *aantal goed* voor de onvolledige en de reguliere versies op een viertal afnamemomenten: eind groep 4, 5, 6 en 7.

Tabel 7.6 Ruwe scores op onvolledige en reguliere versies van Leestempo

Afname- moment	Gemiddelde scores			
	Onvolledige versie		Reguliere versie	
	Aantal gelezen	Aantal goed	Aantal gelezen	Aantal goed
Eind 4	48	32	52	45
Eind 5	51	38	57	53
Eind 6	55	43	61	57
Eind 7	54	47	59	54

In de eerste plaats valt op dat gemiddeld gesproken – en voor bijvoorbeeld P25- en P75-leerlingen is het niet anders – leerlingen op alle tijdstippen in de reguliere versie hogere scores *aantal gelezen* behalen dan in de onvolledige versie. Dit terwijl voor beide versies dezelfde leestijd kon worden gebruikt en de reguliere versie uiteraard veel meer leeswerk vraagt. Kennelijk kost het minder moeite en dus tijd om een Leestempotoets te maken zoals bedoeld, namelijk door te lezen, dan om bij iedere opgave een ‘gestuurde’ vorm van raden te moeten toepassen.

Daarnaast wordt duidelijk dat leerlingen in de onvolledige versies meer fouten maken dan in de reguliere. Gemiddeld gesproken zijn de scores *aantal goed* in de reguliere versies immers hoger dan die in de onvolledige versies. Raden, zelfs deze ‘gestuurde’ vorm, laat staan ‘blind’ raden, zal leerlingen niet makkelijk bevredigende scores op Leestempo opleveren.

### 7.3 Criteriumvaliditeit

Bij de vraag naar de criteriumvaliditeit van de toetsen Leestempo (en Leestechneik), de criteriumvaliditeit van de toetsen Leestechneik & Leestempo dus, gaat het om de mate waarin de scores van leerlingen op dit meetinstrument voorspellen welke kans op succes deze leerlingen hebben om in ‘het lezen van alledag’ bepaalde teksten (technisch) ‘aan te kunnen’. De toetsen zouden antwoord moeten geven op een vraag als: heeft een leerling met die-en-die vaardigheidsscore voldoende leestechneik om die-en-die tekst technisch te kunnen lezen?

In feite hebben we de vraag naar de ‘voorspellende waarde’ van de op Leestechneik & Leestempo behaalde toetsscores reeds beantwoord. In paragraaf 1.2 en 5.2 maakten we melding van de CILT, de Cito Index voor de LeesTechniek, een index voor zowel de technische *leesvaardigheid* van lezers als de technische *leesbaarheid* of moeilijkheid van teksten (Staphorsius & Verhelst, 1997). Zoals een vaardigheidsscore van een leerling op iedere willekeurige toets uit Leestechneik & Leestempo omgezet kan worden in een CILT-waarde, kan ook de moeilijkheid die een tekst leestechneik gesproken heeft, uitgedrukt worden in een CILT-waarde. Met behulp van de CILT, met andere woorden, worden leesvaardigheid en leesbaarheid (= *vereiste* leesvaardigheid) op één schaal gebracht. Dit maakt het mogelijk om voor een leerling teksten te selecteren met een moeilijkheid afgestemd op zijn leestechneik.

De CILT is een zogeheten ‘leesbaarheidsformule’, en in aanleg – maar alleen in aanleg; zie onder – een leesbaarheidsformule als alle andere (zie voor een overzicht van dit soort formules en het onderzoek eromheen: Staphorsius, 1994, pagina 66 e.v.). Dergelijke leesbaarheidsformules komen in grote lijnen als volgt tot stand. Van teksten in een steekproef uit een gegeven populatie wordt de leesbaarheid bepaald. Daarna onderzoekt men door welke kenmerken van die teksten de vastgestelde verschillen in leesbaarheid tussen de teksten verklaard kunnen worden. Hierbij wordt gebruik gemaakt van multi-pele regressie-analyse. De regressievergelijking die hier het resultaat van is, wordt leesbaarheidsformule of -index genoemd. De formule voor de CILT bijvoorbeeld ziet er als volgt uit.

$$\text{CILT} = 114,49 + 0,28\text{FREQ} - 12,23\text{GWL} \quad (R = 0,90; R^2 = 0,82) \quad (7.1)$$

Waarin FREQ een maat is voor *woordfrequentie*, gebaseerd op het in deze publicatie (paragraaf 3.1.2) reeds vermelde frequentie-onderzoek van Staphorsius, Krom & De Geus (1988). En waarin GWL staat voor *gemiddelde woordlengte in letters*. (Zie voor de details van het leesbaarheids-onderzoek dat de CILT tot stand bracht: Staphorsius & Verhelst, 1997).

Met behulp van een formule als deze kan vervolgens de leesbaarheid worden voorspeld van alle teksten die behoren tot de populatie waaruit de onderzochte teksten werden getrokken (zie voor een ander voorbeeld: Staphorsius & Krom, 1985). Leesbaarheidsformules als die van Staphorsius & Krom (1985) maken het weliswaar mogelijk om teksten, gelet op hun leesbaarheid, onderling te vergelijken en bijvoorbeeld naar moeilijkheid te rangordenen, maar het zijn slechts *relatieve* indexen.

Staphorsius, Verhelst & Kleintjes (1996) stellen daar de met grotere praktische waarde omgeven *domeingerichte* indexen tegenover. Dat zijn indexen op grond waarvan leerkrachten aan hun leerlingen teksten kunnen toewijzen die passen bij hun leesvaardigheid. De CILT – we zagen het al in de paragrafen 1.2 en 5.2 – is zo'n domeingerichte index.

Het onderzoek waarin de CILT van een domeingerichte betekenis werd voorzien, wordt gerapporteerd in Staphorsius & Verhelst (1997). Hieronder vatten wij dit onderzoek samen; uitgangspunt ervan wordt gevormd door de volgende vragen (ibid, pagina 160):

- [Kunnen we] *'vaststellen hoe groot de technische leesvaardigheid van een lezer moet zijn om met succes een tekst met een gegeven CILT te kunnen decoderen'* [?]
- [Kunnen we] *'behalve de door teksten veronderstelde leestechneik ook de leestechneik van lezers in de CILT'* [uitdrukken?]

Om antwoord op deze vragen te krijgen, laten Staphorsius & Verhelst om te beginnen leraren basisonderwijs (n = 87) het technisch leesvaardigheidsniveau van hun leerlingen (n = 1844) bepalen met behulp van een zogeheten tekstenschaal. Op deze schaal, bestaande uit twaalf teksten, moet iedere leerkracht van elke individuele leerling in zijn of haar groep aangeven wat de moeilijkste tekst is die de betreffende leerling *'technisch voldoende goed'* kan lezen (ibid, pagina 161). Van alle leerlingen wordt aldus bekend wat volgens hun leerkracht de moeilijkste tekst is die zij (technisch) kunnen lezen. Van die tekst, zoals van alle teksten op de tekstenschaal, is ook de CILT-waarde bekend. Daarmee beschikken de onderzoekers over een maat voor de leestechneik van de leerlingen.

Alle betrokken leerlingen maken bovendien een begrijpend-leestoets. De onderzoekers hebben daarmee dus tevens de beschikking over een maat voor de *begrijpend* leesvaardigheid van elke leerling. Met behulp van de vaardigheidsverdelingen die deze meting oplevert wordt vervolgens nagegaan welke CILT-waarde voor welke gemiddelde leerling in deze vaardigheidsverdelingen geldt: welke CILT-waarde voor de gemiddelde leerling medio groep 4, welke voor de gemiddelde leerling eind groep 4, welke voor de gemiddelde leerling medio groep 5, enzovoort. Uit het LVS-onderzoek rond de toetsserie Leestempo ten slotte is ook bekend welke vaardigheidsscores deze 'gemiddelde leerlingen' behalen op de Leestempotoetsen.

Alle gegevens combinerend valt dan te berekenen wat de gemiddelde Leestempotoetscore is van leerlingen die bijvoorbeeld de derde tekst op de tekstenschaal wel, maar de volgende nog niet voldoende goed kunnen lezen. Omdat – zoals gezegd – die tekst, net als alle teksten op de schaal, een CILT-waarde heeft, kan er een analyse uitgevoerd worden waarin de CILT-waarden geregresseerd worden op het gemiddelde van de Leestempocores van de leerlingen die alleen de eerste tekst kunnen lezen, van de leerlingen die ook de tweede tekst kunnen lezen, van de leerlingen die ook de derde tekst kunnen lezen, enzovoort, tot en met het gemiddelde van de Leestempotoetscores van de leerlingen die alle twaalf teksten voldoende goed kunnen lezen. De onderzoekers zochten dus de coëfficiënten *A* en *B*, zo dat:

$$CILT_i = A + B \times Leestempo_i + \varepsilon_i \quad (7.2)$$

waarbij de variantie van de residuen zo klein mogelijk is. Deze analyse resulteerde in een *A* (het intercept van de regressievergelijking) van 81,28448 en een *B* van –8,8716; met een multiële *R* van 0,94 ( $R^2 = 0,88$ ).

Staphorsius & Verhelst (1997, pagina 161) concluderen op basis hiervan: *'Van een leerling met een bepaalde score op de toetsen Leestempo [de toetsen Leestechneik waren destijds nog niet ontwikkeld] (...) kunnen we nu vaststellen voor welke teksten deze vermoedelijk al wel een voldoende leestechneik heeft en voor welke teksten dat naar alle waarschijnlijkheid nog niet het geval is.'* Met deze uitspraak is meteen antwoord gegeven op de vraag waarmee we deze paragraaf begonnen, de vraag naar de criteriumvaliditeit van de toetsen in Leestechneik & Leestempo.

## 8 Normering van de toetsen Leestempo

De toetsen Leestempo zijn ontwikkeld in de context van het normerings- en kalibratie-onderzoek van het Cito-Leerlingvolgsysteem (LVS); een landelijk longitudinaal onderzoek waarmee in het schooljaar 1989/1990 werd gestart en waarin tussen 1991 en 1995 ook de toetsen uit de opgavenbank Leestempo werden afgenomen. Het onderzoek

had tot doel een evaluatie-instrumentarium tot stand te brengen met behulp waarvan de ontwikkelingen van leerlingen te volgen zijn op de leergebieden Taal, Rekenen en Wereldoriëntatie.

Gezien de omvang van het instrumentarium was het onmogelijk alle toetsen aan alle leerlingen voor te leggen. Daarom is ervoor gekozen om de totale steekproef in substeekproeven te verdelen en de af te nemen toetsen over deze substeekproeven te verdelen, maar wel zodanig dat iedere toets minstens één keer met elke andere binnen een substeekproef voorkwam. Hierdoor bleef de mogelijkheid open om de relatie tussen toetsen, met name die tussen 'soortgenoten' te analyseren.

## 8.1 Referentiegroep

Ten behoeve van de samenstelling van het scholencohort dat aan het LVS-onderzoek zou deelnemen, werd het bestand van Nederlandse basisscholen op basis van hun schoolgewicht opgesplitst in drie strata, een procedure vergelijkbaar met het onderzoek in het kader van de Periodieke Peiling van het Onderwijs Niveau (Wijnstra, 1988; zie ook deze publikatie: paragraaf 4.1):

- Stratum 1: scholen met een schoolgewicht  $\leq 1.05$ ; in dit stratum zijn vooral scholen vertegenwoordigd met relatief weinig arbeiderskinderen en relatief veel kinderen uit gezinnen met een hogere sociaal-economische status.
- Stratum 2 scholen met een schoolgewicht  $1.06 - 1.15$ ; de scholen in dit stratum nemen een middenpositie in tussen de beide andere strata.
- Stratum 3 scholen met een schoolgewicht  $> 1.15$ ; scholen in stratum 3 kenmerken zich door een relatief hoog aantal allochtone leerlingen.

Uit deze drie strata werd een aselechte steekproef van scholen getrokken. Deze scholen werden aangeschreven met het verzoek aan het onderzoek deel te nemen. Circa 220 scholen, met in totaal ongeveer 4500 leerlingen in groep 3, reageerden positief. Om budgettaire redenen moest dit aantal zoveel mogelijk worden teruggebracht tot ongeveer de geplande 180 scholen en 3600 leerlingen. Uiteindelijk bleven circa 190 scholen met ongeveer 3500 leerlingen over (zie tabel 8.1). Omdat de omvang van het instrumentarium het onmogelijk maakte om alle toetsen bij dezelfde leerlingen af te nemen, is ervoor gekozen om de geselecteerde scholen te verdelen over zes substeekproeven, ook wel school-afnamegroepen (SAG's) geheten. Hierbij werd rekening gehouden met regionale spreiding, grootte van de school, het aantal allochtone leerlingen, evenredige spreiding over de drie strata en het totaal aantal leerlingen per substeekproef. Indien een school met parallelgroepen aan het onderzoek meedeed, werden deze groepen doorgaans bij verschillende substeekproeven ingedeeld. Het aantal scholen per substeekproef lag tussen de 28 en 33, ieder met in totaal bijna 600 leerlingen.

Voor iedere toetsserie werd binnen dat onderzoek een apart design opgesteld; dat van Leestempo wordt weergegeven in figuur 5.4 in paragraaf 5.6.3 van deze publicatie.

Tabel 8.1 Regionale spreiding van scholen en leerlingen in populatie en steekproef bij de start van het onderzoek

Regio	Percentage in populatie	Percentage in steekproef	Aantal leerlingen in steekproef		
			Stratum 1	Stratum 2	Stratum 3
A Amsterdam e.o.	7	7,4	4	24	226
B Kop Noord-Holland	6	3,1	3	38	67
C Leiden e.o.	6	6,3	55	117	43
D Rotterdam/Den Haag	7	3,4	18	47	52
E Zeeland/W-Brabant	8	10,8	157	152	61
F Gorinchem e.o.	7	7,0	126	59	56
G Utrecht	8	7,6	77	68	117
H Arnhem/Nijmegen	8	10,8	173	134	65
I Midden/Oost N-Brabant	8	11,1	149	106	127
J Limburg	7	15,4	180	224	126
K Twente/Achterhoek	7	5,2	14	58	105
L NW Overijssel/ZW Drente	7	5,7	111	56	30
M Friesland	7	2,2	18	31	26
N Groningen	8	3,8	84	15	32
			1169 (34,1%)	1129 (32,9%)	1133 (33,0%)

Door het wegvallen van een deel van de subsidie moest bij de start van het derde onderzoeksjaar de omvang van het cohort gehalveerd worden. Vanuit overwegingen van representativiteit werd deze halvering zoveel mogelijk per substeekproef doorgevoerd. Vervolgens werden de resterende 'halve' substeekproeven weer samengevoegd tot drie substeekproeven van ongeveer 30 scholen met ieder ongeveer 500 à 600 leerlingen. Deze groep bleef uiteindelijk tot het einde van het onderzoek (medio groep 8) in tact (zie tabel 8.2).

Tabel 8.2 Regionale spreiding van scholen en leerlingen in populatie en steekproef na halvering van het cohort

Regio	Percentage in populatie	Percentage in steekproef	Aantal leerlingen in steekproef		
			Stratum 1	Stratum 2	Stratum 3
A Amsterdam e.o.	7	6,6	4	23	81
B Kop Noord-Holland	6	3,5	0	38	19
C Leiden e.o.	6	6,6	53	44	12
D Rotterdam/Den Haag	7	6,1	19	46	35
E Zeeland/W-Brabant	8	11,7	98	80	15
F Gorinchem e.o.	7	3,9	37	11	16
G Utrecht	8	4,1	0	20	48
H Arnhem/Nijmegen	8	13,0	98	47	69
I Midden/Oost N-Brabant	8	15,7	83	59	117
J Limburg	7	11,1	73	109	0
K Twente/Achterhoek	7	4,7	11	36	30
L NW Overijssel/ZW Drente	7	6,3	80	0	23
M Friesland	7	1,8	0	3	26
N Groningen	8	5,0	35	34	14
			591 (35,9%)	550 (33,4%)	505 (30,7%)

Aan het onderzoek rond de toetsen Leestempo namen vóór de steekproefreductie ruim 2100 leerlingen (van de  $\pm$  3500) deel. Zij waren afkomstig van 111 scholen. Na de reductie restten nog zo'n 1000 leerlingen (van de  $\pm$  1600), afkomstig van – aanvankelijk – 52 scholen; 'aanvankelijk', want drie scholen trokken zich in de loop van het onderzoek terug. Voor de goede orde: de in figuur 5.4 (paragraaf 5.6.3) vermelde substeekproeven A, B, C en B\*, D, E vallen niet zonder meer samen met een van de zes (vóór de steekproefreductie) of drie (na de reductie) substeekproeven (SAG's) waarin het totale LVS-cohort werd ondergebracht. A en C bevatten elk alle leerlingen uit één zo'n SAG plus een kwart van de leerlingen uit een andere; B is de enige substeekproef in het onderzoek rond Leestempo die compleet samenvalt met een SAG; B\*, D en E bestaan elk uit de helft van een van de drie SAG's die na de reductie samengesteld werden.

Zie voor de exacte leerlingaantallen per afnamemoment tabel 8.3 in paragraaf 8.2.

## 8.2 Referentiegegevens

Op grond van het longitudinale onderzoek in de bovenbeschreven referentiegroep beschikken we over de gegevens (gemiddelden en P10-, P25- en P75-waarden) waarmee we de verdeling van de vaardigheid *technisch lezen* op negen meetmomenten kunnen typeren. Deze gegevens staan in tabel 8.3.

In de gegevens is middels weging rekening gehouden met het feit dat in de referentiegroep de verdeling van de scholen over de strata uiteraard een andere is dan die in de populatie. De wegingsfactor per stratum is gelijk aan de ratio van de populatieproportie en de proportie in de referentiegroep.

Tabel 8.3 Typering van de vaardigheidsverdeling

Op afnamemoment	P10	P25	GEM	P75	Aantal leerlingen
Medio 4	70,475	75,032	79,740	84,264	2126
Eind 4	74,171	78,607	83,219	87,548	2148
Medio 5	76,847	81,268	85,639	89,526	974
Eind 5	80,016	83,360	88,616	92,421	1002
Medio 6	80,789	85,378	92,081	97,447	993
Eind 6	81,475	87,238	95,161	102,486	1018
Medio 7	81,535	89,010	99,560	108,193	974
Eind 7	87,168	92,979	100,824	106,827	978
Medio 8	87,477	94,035	101,278	107,984	971

Medio 4 tot en met Medio 8: de negen afnamemomenten (steeds halverwege en einde groep)

P10, P25 en P75: het tiende, vijftiengste en vijfenzeventigste percentiel

GEM: het gemiddelde

Aantal leerlingen: het aantal leerlingen per afnamemoment

In de handleiding bij de toetsen zijn deze waarden gekoppeld aan de in het Cito-Leerlingvolgsysteem gebruikelijke vaardigheidsniveaus A tot en met E (Krom, 2001, pagina 79). In tabel 8.4 geven we deze koppeling weer.

Tabel 8.4 Referentiegegevens Leestempo

---

Vaardigheidsniveau	Medio 4	Eind 4	Medio 5	Eind 5	Medio 6	Eind 6	Medio 7	Eind 7	Medio 8
A	> 84	> 87	> 89	> 92	> 97	> 102	> 108	> 106	> 107
B	80-84	84-87	86-89	89-92	93-97	96-102	100-108	101-106	102-107
C	76-79	79-83	82-85	84-88	86-92	88-95	90-99	93-100	95-101
D	71-75	75-78	77-81	81-83	81-85	82-87	82-89	88-92	88-94
E	< 71	< 75	< 77	< 81	< 81	< 82	< 82	< 88	< 88

A: de leerlingen in het bovenste kwartiel (Q1)

B en C: de leerlingen in Q2 en Q3 (B: de leerlingen boven het gemiddelde; C: de leerlingen onder het gemiddelde)

D en E: de leerlingen in Q4 (E: de 10% laagstscorende leerlingen; D: de 15% leerlingen daarboven)

---

Zeker na de halvering van het cohort was het niet langer zinvol om voor Leestempo afzonderlijke referentiegegevens ten behoeve van allochtone leerlingen te verzamelen. Het aantal allochtone leerlingen dat aan het normerings- en kalibratieonderzoek in het kader van de ontwikkeling van de toetsen Leestempo deelneemt is daarvoor te gering: vóór de steekproefreductie beschikten we per afnamemoment over de gegevens van een kleine 250 allochtone leerlingen; daarna nog slechts over 88 à 113 allochtone leerlingen.



predictor). Criterium en predictor correleren .67. Een correlatie die 'gedrukt' zal zijn door, ten eerste, het plafondeffect dat we aantreffen in de scores op de ankertoets. Dit effect treedt op omdat Leestechnekitems – bedoeld immers voor leerlingen eind groep 3 – veelal te gemakkelijk zijn voor het merendeel van de leerlingen halverwege groep 4. Zelfs de selectie van de twintig moeilijkste Leestechnekitems ten behoeve van de ankertoets heeft dit niet kunnen verhinderen. Ten tweede is het natuurlijk zo dat de toetsinhoud van Leestempo een andere is dan die van Leestechneik (teksten vs. woorden) en dat Leestempo de technische leesvaardigheid *onder tijdsdruk* meet en Leestechneik de technische leesvaardigheid zonder meer; de toetsen meten met ander woorden niet exact dezelfde vaardigheid (zie ook deze publicatie: paragraaf 9.3).

De regressievergelijking die het latente criterium 'het best' voorspelt luidt:

$$\hat{\xi}_{Leestempo} = 62.688 + 12.112 * \hat{\xi}_{Leestechneik}$$

Met behulp van deze vergelijking worden de vaardigheids- of schaalscores op de toetsen Leestechneik ten slotte getransformeerd en overgebracht op de Leestemposchaal.

### 9.3 Meten van de technische leesvaardigheid (z)onder tijdsdruk

De introductie van tijdsdruk in de meting van de technische leesvaardigheid na de fase van het aanvankelijk lezen – het in deze context belangrijkste verschil tussen Leestechneik en Leestempo – is overigens onvermijdelijk. De mogelijkheden van een 'tijdloze' toetsvorm raken dan namelijk uitgeput: zo goed als alle leerlingen hebben alle (lees-) technische kennis wel vergaard, zij onderscheiden zich vanaf dat moment onderling eigenlijk alleen nog in de mate van geautomatiseerdheid van de leeshandeling, respectievelijk de snelheid waarmee zij lezen (kunnen).

Aan de andere kant – waarom dan immers geen leessnelheidsinstrument voor alle groepen van de basisschool ontwikkeld? – is het meten van de leestechneik onder tijdsdruk vóór het einde van de aanvankelijke fase naar onze mening minder geschikt. Wij verwachten dat een aantal leerlingen in die periode waarschijnlijk nog te veel moeite zal hebben met het herkennen van de letters en met het synthetiseren van de klanken om deze deelvaardigheden onder tijdsdruk correct uit te kunnen voeren.

Vandaar dat wij in groep 3 nog geen leessnelheidsinstrument willen inzetten en gekozen hebben voor de ontwikkeling van een 'dubbel'-instrument: Leestechneik (voor leerlingen in groep 3; zonder tijdsdruk) & Leestempo (voor leerlingen in de hogere groepen; met tijdsdruk).

### 9.4 Stabiliteit van de positie van leerlingen in de vaardigheidsverdelingen

Gelet op de functionaliteit van de toetsserie Leestechneik & Leestempo – het *volgen* van leerlingen – gaat het bij het verantwoorden van de koppeling van de beide constituerende elementen vooral om de vraag naar de stabiliteit van de positie van de leerlingen in de verdelingen van achtereenvolgens de scores op Leestechneik en die op Leestempo. Het antwoord op deze vraag geeft tabel 9.1.

Tabel 9.1 Percentage leerlingen per vaardigheidsniveaucombinatie

		Leestempo				
Leestechneik	A	B	C	D	E	
A	57,6	24,6	14,6	2,6	0,6	
B	17,7	33,8	30,2	13,7	4,6	
C	10,1	23,7	36,6	19,0	10,6	
D	2,4	12,2	29,8	29,3	26,3	
E	1,5	1,5	13,1	27,7	56,2	

Leestechneik A tot en met E: vaardigheidsniveaus ankertoets Leestechneik  
 Leestempo A tot en met E: vaardigheidsniveaus Leestempo M4 (-Extra)

Tabel 9.1 laat bijvoorbeeld zien dat ongeveer 58% van de leerlingen die op de ankertoets Leestechneik scores behalen die in vaardigheidsniveau A vallen, dat wil zeggen de leerlingen die op die toets tot de 25% hoogst-scorende leerlingen behoren, ook op Leestempo M4 of M4-Extra scores behalen die 'goed zijn' voor niveau A. Daarnaast komt

nog eens ongeveer 25% van de A-leerlingen op Leestechiek bij de afname van Leestempo M4(-Extra) in het (onmiddellijk aangrenzende) B-niveau terecht, zodat de positie van in totaal zo'n 83% van de 'oorspronkelijke' A-leerlingen qua vaardigheidsniveau gelijk blijft of ten hoogste één niveau opschuift. Dit beeld treffen we op alle niveaus aan: steeds blijft ongeveer 80 à 85% van de leerlingen, gaande van Leestechiek naar Leestempo, in hetzelfde niveau of schuift op naar een van de onmiddellijk aangrenzende niveaus. In de praktijk zal de overstap, halverwege groep 4, van het ene naar het andere instrument derhalve een stabiel beeld geven.

<dummy links van literatuur>

## Literatuur

- Allington, R.A. (1984). Oral reading. In: P.D. Pearson, R. Barr, M.L. Kamil en P. Mosenthal (Eds.), *Handbook of reading research* (pp. 829-863). White Plains, NY: Longman.
- Armbruster, B.B. en I.A.G. Wilkinson (1991). Silent reading, oral reading, and learning from text. *The Reading Teacher*, 45, 154-155.
- Bechger, T., M. van der Lubbe en R. Krom (2001). *Een vergelijking tussen de scores op de papieren versie en de omputerversie van Leestempo E6*. Publicaties Primair Onderwijs. Arnhem: Citogroep
- Brus, B.Th. en M.J.M. Voeten (1973). *Eén-Minuut-Test vorm A en B. Verantwoording en handleiding*. Nijmegen: Berkhout Testmateriaal BV.
- Burge, P.D. (1983). Comprehension and rate: oral vs. silent reading for low achievers. *Reading Horizons*, 23, 201-206.
- COTAN (1999). *Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: COTAN, Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen, NIP.
- Curtis, M.E. en R. Glaser (1983). Reading theory and the assessment of reading achievement. *Journal of Educational Measurement*, 20, 133-147.
- Engelen, R.J.H. en T.J.H.M. Eggen (1993). Equivaleren. In: T.J.H.M. Eggen en P.F. Sanders (Red.). *Psychometrie in de praktijk*. Arnhem: Cito.
- Glas, C.A.W. en N.D. Verhelst (1993). Een overzicht van itemresponsmodellen. In: T.J.H.M. Eggen en P.F. Sanders (Red.). *Psychometrie in de praktijk*. Arnhem: Cito.
- Goldsmith-Phillips, J. (1989). Word and context in reading development: a test of the interactive-compensatory hypothesis. *Journal of Educational Psychology*, 81, 299-305.
- Groot, H. de, M. Huizing en A. Noteboom (1999). *Het Leerlingvolgsysteem en speciale leerlingen*. Arnhem: Cito.
- Kim (Yoon), Y.H. en E.T. Goetz (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.
- Kohnstamm, G.A., A.M. Schaerlaekens, A.D. de Vries, G.W. Akkerhuis en M. Froominckx (1981). *Nieuwe Streeflijst Woordenschat voor 6-jarigen*. Lisse: Swets en Zeitlinger.
- Kramers, H. (1990). *Het CLA model uitgebreid voor onvolledige data met een toepassing*. Rapport 335. Arnhem: Cito.
- Krom, R. (2000). *De grondwoorden in Leestechneek voorzien van frequentiegegevens, -posities en wenselijkheidspercentages*. Intern rapport Project Leestechneek & Leestempo.
- Krom, R.S.H. (1988). *Technisch lezen 3*. Arnhem: Cito.
- Krom, R.S.H. (1990). *Wenselijke woordenschat en feitelijke frequenties. De Nieuwe Streeflijst Woordenschat getrancheerd naar verwervingsleeftijd en voorzien van frequentiegegevens*. Arnhem: Cito.
- Krom, R.S.H. (1996). *Leestempo*. Arnhem: Cito.
- Krom, R.S.H. (2001). *Leestechneek & Leestempo*. Arnhem: Citogroep.
- Krom, R.S.H., N.D. Verhelst en F. Kamphuis (2000). *De invloed van de afname-instructie op leessnelheidsmetingen met Leestempotoetsen*. Paper Onderwijs Research Dagen.
- Krom, R.S.H., N.D. Verhelst en F. Kamphuis (in voorbereiding). *Does item-hopping pay?* (werktitel).

- Leu, D.J., Jr. (1982). Oral reading error analysis: a critical review of research and application. *Reading Research Quarterly*, 17, 420-437.
- Leu, D.J., Jr., L-J.C. DeGroot en H.D. Simons (1986). Predictable texts and interactive-compensatory hypotheses: evaluating individual differences in reading ability, context use, and comprehension. *Journal of Educational Psychology*, 78, 347-352.
- Loon-Vervoorn, W.A. van (1985a). *De Nieuwe Streeflijst als maat voor de verwervingsleeftijd van woorden*. Intern rapport Psychologisch Laboratorium, Rijksuniversiteit Utrecht.
- Loon-Vervoorn, W.A. van (1985b). De Nieuwe Streeflijst als maat voor de verwervingsleeftijd van woorden. *Nederlands Tijdschrift voor de Psychologie*, 40, 44-47.
- Loon-Vervoorn, W.A. van (1985c). De Nieuwe Streeflijst als maat voor de verwervingsleeftijd van woorden II. *Nederlands Tijdschrift voor de Psychologie*, 40, 503-518.
- Loon-Vervoorn, W.A. van, A. Ridderbos en A. Rietveld van (1985). *De Nieuwe Streeflijst als maat voor de verwervingsleeftijd van woorden (adjectieven en verba) II*. Intern rapport Psychologisch Laboratorium, Rijksuniversiteit Utrecht.
- Lord, F.M. en M.R. Novick (1968). *Statistical theories of mental test scores*. New York: Addison Wesley.
- Moosbrugger, H., en H. Müller (1982). A classical latent additive test model (CLA model). *German Journal of Psychology*, 6, 145-149.
- Mosenthal, P. (1976-1977). Psycholinguistic properties of aural and visual comprehension as determined by children's abilities to comprehend syllogisms. *Reading Research Quarterly*, 12, 55-92.
- Mosenthal, P. (1978). The new and given in children's comprehension of presuppositional negatives in two modes of processing. *Journal of Reading Behavior*, 10, 267-278.
- 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., C. Lillas en M.A. Rzoska (1988). Have we been misled by miscues? *The Reading Teacher*, 42, 6-10.
- Raatz, U. (1985). Better theory for better tests? *Language Testing*, 2, 60-75.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rosier, W. (2001). *Computerprogramma Leerlingvolgsysteem*. Versie 3.0. Arnhem: Citogroep.
- Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.
- Stanovich, K.E., R.F. West en D.J. Freeman (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.
- Staphorsius, G. (1992). *Clib-toetsen. Een serie toetsen voor het volgen van de ontwikkeling van de leesvaardigheid van leerlingen met behulp van de Cito Leesindex voor het Basis- en speciaal onderwijs*. Arnhem: Cito.
- Staphorsius, G. (1994). *Leesbaarheid en leesvaardigheid. De ontwikkeling van een domeingericht meetinstrument*. Academisch proefschrift, Enschede: Universiteit van Twente.
- Staphorsius, G. en R.S.H. Krom (1985). Predictie van leesbaarheid. *Tijdschrift voor Taalbeheersing*, 7, 192-211.
- Staphorsius, G. en R.S.H. Krom (1998). *Toetsen Begrijpend Lezen*. Arnhem: Cito.
- Staphorsius, G. en N.D. Verhelst (1997). Indexering van de leesteknik. *Pedagogische Studiën*, 74, 154-164.

- Staphorsius, G., R.S.H. Krom en K. de Geus (1988). *Frequenties van woordvormen en letterposities in jeugdlectuur*. Arnhem: Cito.
- Staphorsius, G., N.D. Verhelst en F.G.M. Kleintjes (1996). De ontwikkeling van een domeingerichte index voor leesbaarheid en leesvaardigheid. *Taalbeheersing*, 18, 116-132.
- Staphorsius, G., N.D. Verhelst en F.G.M. Kleintjes (2000). *De Clib-toetsen. Verslag van kalibratie-, validerings- en normeringsonderzoek, Wetenschappelijke verantwoording*. Arnhem: Citogroep.
- Verhelst, N.D. (1992). *Het één parameter model (OPLM). Een theoretische inleiding en een handleiding bij het computerprogramma*. Arnhem: Cito.
- Verhelst, N.D. (1993). Itemresponstheorie. In: T.J.H.M. Eggen en P.F. Sanders (Red.), *Psychometrie in de praktijk*. Arnhem: Cito.
- Verhelst, N.D. (1994). *Zwarte dozen*. Enschede: Universiteit Twente.
- Verhelst, N.D. en T.J.H.M. Eggen (1989). *Psychonomische en statistische aspecten van peilingsonderzoek* (PPON-rapport, nr. 4). Arnhem: Cito.
- Verhelst, N.D. en C.A.W. Glas (1995). The one parameter logistic model. In G.H. Fischer en I.W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 215-239). New York: Springer.
- Verhelst, N.D. en F.G.M. Kleintjes (1993). Toepassingen van itemresponstheorie. In: T.J.H.M. Eggen en P.F. Sanders (Red.), *Psychometrie in de praktijk*. Arnhem: Cito.
- Verhelst, N.D., C.A.W. Glas en H.H.F.M. Verstralen (1994). *OPLM: One Parameter Logistic Model. Computer Program and manual*. Arnhem: Cito
- Verhelst, N.D., H.H.F.M. Verstralen en T.J.H.M. Eggen (1991). *Finding starting values for the item parameters and suitable discrimination indices in the one-parameter logistic model*. Measurement and Research department reports 91-10. Arnhem: Cito.
- Verhoeven, L. (1995). *Drie-Minuten-Toets*. Arnhem: Cito.
- Verhoeven, L.T. (1980a). *Technisch lezen 1*. Arnhem: Cito.
- Verhoeven, L.T. (1980b). *Technisch lezen 2*. Arnhem: Cito.
- Verhoeven, L.T. (1982). *Verantwoording toetsen Aanvankelijk lezen en spellen*. Specialistisch Bulletin 13. Arnhem: Cito.
- Verstralen, H.H.F.M. (1997). *OPTAL: Inverse OPLAT and item and test characteristics in populations*. Arnhem: Cito.
- Wijnstra, J.M. (red.), (1988). *Balans van het rekenonderwijs in de basisschool*. Arnhem: Cito.
- Wijnstra, J.M. (red.), (1990). *Verantwoording van de rekenpeiling medio en einde basisonderwijs 1987*. Arnhem: Cito.