



UNIVERSITY OF AMSTERDAM

## UvA-DARE (Digital Academic Repository)

Lezen gaat voor

Struiksma, A.J.C.

[Link to publication](#)

*Citation for published version (APA):*

Struiksma, A. J. C. (2003). *Lezen gaat voor*. Amsterdam: VU Uitgeverij.

### General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

### Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<http://dare.uva.nl>)

### 3. Het rt-onderzoek, remedial teaching volgens protocol

#### 3.1 Inleiding

*Het doel van het eerste onderzoek was te onderzoeken in welke mate een ongunstige positie op het continuüm van technische leesvaardigheid op het moment van afronding van de fase van het aanvankelijk leesonderwijs - een positie die een risicofactor inhoudt voor het ontwikkelen van dyslexie - door gerichte interventie kan worden omgebogen. Deze interventie had de vorm van een geprotocolleerde vorm van individuele remedial teaching (rt). Het onderzoek beoogde om in de praktijk van het basis-onderwijs de omstandigheden voor de  $\pm 10\%$  zwakste lezers te optimaliseren door een geprotocolleerd, individueel uit te voeren remediërend traject te ontwikkelen en in te voeren, in de verwachting dat hierdoor de technische leesvaardigheid voor een zo groot mogelijk deel van de zwakke lezers weer binnen de normale bandbreedte komen.*

Het rt-programma omvatte zowel het oefenen met losse woorden als met teksten, conform de aanbevelingen van Grossen (1997), het National Reading Panel (2000) en Levy (2001). De training op woordniveau, door projectmedewerkers, werd gedaan met behulp van het computerprogramma woordtraining dat beschreven wordt in hoofdstuk 4. De remedial teaching, uitgevoerd door de remedial teachers van de projectscholen, was toegespitst op het lezen van teksten. Remedial teaching heeft een aantal formele kenmerken: setting, frequentie, duur van de sessies, lengte van het traject en de toegepaste methodiek.

#### 3.2 Overwegingen

##### **methodologie**

De opzet van een dergelijk onderzoek kent echter een aantal problemen. Swanson e.a. (1999) noemen als belangrijkste methodologische valkuil van zogenaamd *school based* interventieonderzoek de *classroom by treatment confounds*, dat wil zeggen *treatment* in de ene klas en de controle in een andere. Dit kan worden opgelost door de klas als eenheid van analyse te nemen en te rekenen met de gemiddelde klassescore in plaats van individuele scores; deze oplossing wordt echter vrijwel niet toegepast. In ons onderzoek is hiervoor en voor een eventueel schooleffect gecontroleerd door de wijze waarop de groepen zijn samengesteld en over de condities zijn verdeeld.

Ook Lyon & Moats (1997) beschreven een aantal methodologische beperkingen van veel interventieonderzoek:

- heterogeniteit van de proefgroep en onduidelijke definiëring;
- slecht beschreven interventies, waardoor deze niet gerepliceerd kunnen worden;
- inadequate controlegroep;
- te geringe duur van de interventie;

- onvoldoende generalisatiemogelijkheid als onderdeel van de interventie;
- onvoldoende controle voor de leergeschiedenis van de leerlingen;
- onvoldoende controle op leerkrachtkenmerken en consistentie in de uitvoering van de methode;
- onvoldoende onderhoud na afloop van de interventie;
- effectmeting op te beperkte variabelen, met name alleen accuratesse en niet vlotheid.

In het remedial teaching onderzoek is geprobeerd deze valkuilen zo goed mogelijk te omzeilen. Als belangrijkste aandachtspunt wat betreft heterogeniteit van de proefgroep en onduidelijke definiëring noemen Lyon & Moats heterogeniteit qua leesniveau bij aanvang, omdat dit een rol zou kunnen spelen in de afhankelijkheid van leerlingen van bepaalde typen instructie en de intensiteit en duur van de instructie. De remedial teaching was erop gericht het technisch lezen te bevorderen bij leerlingen die de eerste fase van het aanvankelijk lezen voorbij waren. Alle leerlingen waren qua leesvaardigheid vergelijkbaar met gemiddelde leerlingen  $\pm$  halverwege groep 3. Ze waren in staat teksten op niveau 1 te lezen, al was dat meestal nog nadrukkelijk spellend en langzaam, overeenkomend met het zogenaamde frustratieniveau (Adams 1991, Wheldall e.a. 1992, Rasinski 1999). Leerlingen van dit leesniveau beheersen de deelvaardigheden van het aanvankelijk lezen, met name klank-tekenkoppeling en auditieve synthese, conform de wijze waarop deze in het onderwijs gewoonlijk worden gevolgd en geïnstrueerd.

Zwakke lezers hebben instructie nodig en uitbreiding van de daadwerkelijk bestede leertijd. Gettinger (1984) schat dat de zwakste 10% van de leerlingen vier tot zes keer zoveel tijd nodig hebben als de beste 10% om dezelfde leerstof op hetzelfde niveau te gaan beheersen. Vergelijking van gerapporteerde effectgroottes valt uit in het voordeel van individuele instructie ten opzichte van klasgebonden interventies (Nicolson e.a. 1999). Om die reden is voor individuele remedial teaching gekozen.

Remedial teaching is vooral intensivering van instructie met toepassing van bewezen effectieve (ortho)didactische principes. Deze zijn: directe instructie, *precision teaching* en toegepaste gedragsleer (Kozloff e.a. 1999). Frequentie op zichzelf lijkt geen thema in evaluatieonderzoek. In trainingonderzoek waarin slechts de vraag aan de orde is of een bepaalde interventie effectief is, wordt een naar het lijkt arbitraire frequentie gehanteerd (Nicolson e.a. 1999, Torgesen 2000). In enkele schoolverbeteringstrajecten (Clay 1993, Slavin 1994) werd individuele remedial teaching dagelijks gegeven, zonder dat de frequentie een experimentele variabele was. De keuze voor een frequentie van viermaal per week is dan ook overwegend een praktische: de hoogst haalbare frequentie met de beschikbare medewerkers. Voorzover in onderzoeksrapportage de sessieduur beschreven wordt, ligt deze vaak rond een half uur. Ook hier lijkt meer sprake van gezond verstand, dan van iets anders.

In trainingonderzoek wordt de lengte van het traject meestal uitgedrukt in het aantal uren dat de training in zijn totaliteit duurde (Blok e.a. 2002), variërend van 1 tot 20 uur in totaal. In schoolverbeteringstrajecten vinden we meestal een trajectduur van een half of een heel schooljaar, zonder dat de trajectduur op zichzelf een experimentele variabele is. Vellutino e.a. (1998) beschrijven een meer flexibele opzet waarbij de vooruitgang

van leerlingen na 25 weken geëvalueerd wordt en de training alleen bij onvoldoende resultaat nog 10 à 15 weken wordt voortgezet met een totale trainingduur variërend van 35 tot 80 uur.

De keuze voor een half schooljaar komt voort uit het gekozen design. Ook de leerlingen van de controlegroep moest binnen het schooljaar een even intensief remediërend traject kunnen worden aangeboden. Gebleken is dat in een schooljaar, ondanks zorgvuldige planning, twee perioden van vijftien effectieve schoolweken op de meeste scholen niet voor alle leerlingen gerealiseerd konden worden. Met een frequentie van viermaal per week en een sessieduur van een half uur was in vijftien effectieve schoolweken de totale trainingstijd 30 uur. Dit aantal uren is relatief hoog in vergelijking met andere trainingstudies, maar gering in het kader van schoolverbetering.

### **inhoud: accuratesse en vlotheid, spellen en raden**

Accuratesse en vlotheid zijns de twee meest in het oog springende kwaliteiten van leesgedrag. Verschillende benamingen voor subtypen dyslexie suggereren dat een gestoorde leesontwikkeling een zaak is van het een of het ander, zoals *speed dyslexia* tegenover *decoding dyslexia* (Wimmer 1993) en de al genoemde *accuracy disabled* versus *rate disabled* (Lovett 1987). Bij een onregelmatige orthografie zoals in het Engels worden accuratesse en vlotheid lange tijd onderscheiden als afzonderlijk aspecten van het lezen. Bij een regelmatige orthografie zoals het Duits en het Nederlands lijkt accuratesse slechts aanvankelijk een onderscheidend aspect.

Bij zwakke lezers wordt ook vaak een spellende of een radende voorkeurstrategie onderscheiden. Het onderscheid tussen spellers en raders, als beschrijvende termen (Van der Leij 1983), is eveneens onder verschillende andere benamingen bekend, bijvoorbeeld *dysphonetic dyslexia* versus *dyseidetic dyslexia* (Boder 1973) of P-type dyslexie versus L-type dyslexie (Bakker 1979). Dit onderscheid loopt niet parallel aan het onderscheid tussen accuratesse en vlotheid, maar heeft er wel mee te maken. Als een leerling een sterk spellende leesstrategie hanteert, dan leidt dat tot een laag leestempo, maar ook zonder overmatig spellend lezen is een laag leestempo mogelijk. Aan de andere kant leidt raden tot een geringe accuratesse. Aan raden zit de connotatie van ook - te - snel lezen, maar raden is ook mogelijk bij een laag leestempo. Deze auteurs gaan verder dan een beschrijving van leesgedrag; bij hen is sprake van subtypering, waarbij spellen of raden een min of meer stabiel kenmerk van de lezer zou zijn.

Rispens e.a. (1994) zijn van mening dat de waarde van subtypering in de gebruikelijke betekenis van het woord twijfelachtig is, omdat de classificatie van leerlingen bijvoorbeeld als speller of rader sterk afhankelijk is van de gekozen criteria en de wijze waarop de categorieën geoperationaliseerd zijn. Een alternatieve interpretatie is, dat het niet gaat om stabiele trek die altijd en in alle leesgedrag tot uiting komt, maar dat spellend of radend lezen - of juist het ontbreken daarvan - worden uitgelokt door taakeisen, zoals de relatieve moeilijkheidsgraad van de leesstof, of specifieke instructie- of aanbiedercondities, zoals flitsen of een wisselrij. Zelfs van de beste lezer kan immers een rader worden gemaakt, als de condities maar extreem genoeg zijn, bijvoorbeeld

door lange, laagfrequente woorden te laten lezen bij een zeer korte aanbestedingstijd. Het is dus niet zozeer, in de woorden van Rispen, de classificatie als speller of rader die een functie is van de gevolgde classificatieprocedure, als wel het leesgedrag, in het bijzonder de kenmerken spellen of raden, dat een functie is van verschillende stimulus- en bekrachtigingscondities en dat binnen de toegepaste gedragsleer in een functie-analyse beschreven kan worden. Dit laat onverlet, dat leerlingen met bepaalde kenmerken een voorkeur kunnen ontwikkelen; zo toonden Van der Schoot e.a. (2000) een verband aan tussen impulsiviteit en de neiging tot raden.

Op grond van welke leergeschiedenis vallen spellen en raden te begrijpen? Confrontatie van een leerling met een woord dat niet herkend wordt levert, zeker in een schoolomgeving waar dit wel verwacht wordt, spanning op. Spellen betekent in dit verband dat de leerling teruggaat naar de route van de *self-teaching*, een arbeidsintensieve, maar veilige weg. Raden betekent wel verwerking op woordniveau, maar er wordt een ander woord gelezen dan er staat. Geen van beide zijn fraaie oplossingen, maar ze leveren in ieder geval een resultaat op en de spanningsvermindering die daarop volgt, bekrachtigt  $(-C^-)^3$  het kiezen van deze oplossing. Daarbij zullen juist leerlingen bij wie de directe woordherkenning slecht op gang komt, aanvankelijk in verhouding veel meer leerervaring opdoen met het spellend lezen. Door negatieve bekrachtiging kan zich een gewoonte (*habit*) vormen die gaat functioneren als vermijdingsgedrag voor het oefenen op een ander niveau van verwerking, maar vooral voor het soepel schakelen tussen verschillende niveaus van verwerking afhankelijk van de taakeisen. Een kenmerk van vermijdingsgedrag is dat het star wordt toegepast, niet afgestemd op de taakeisen; dat zou men een voorkeurstrategie kunnen noemen.

Een spellende strategie is bij moeilijke woorden nog de geëigende, *item-based* oplossing (vergelijk Spear & Sternberg 1994). Ook kan beargumenteerd worden dat een beetje raden geen probleem is en hoort bij de normale leesontwikkeling; niemand leest onder alle omstandigheden volledig foutloos. Maar er kan ook sprake zijn van rigide en niet langer adequate oplossingsstrategieën. Van der Leij (1983) noemt spellen en raden noodoplossingen om onbekende woorden te attaqueren bij gebrek aan interne representatie en meer een gevolg van dat tekort dan de oorzaak. Ook Hendriks & Kolk (1997) spreken over spellen en raden als verschillende oplossingen voor het kernprobleem van een te traag verlopend leesproces. Zij beschreven spellend of radend lezen als een vorm van strategische adaptatie; een oplossing die de leerling kiest ter compensatie van zijn tekort. Ook toonden zij aan dat de gekozen oplossing in zekere mate willekeurig is, althans dat lezers kunnen overgaan van spellend lezen naar radend lezen en andersom, afhankelijk van de vraag of zij de instructie kregen zo snel mogelijk of juist met zo weinig mogelijk fouten te lezen. Met andere woorden een hoger leestempo gaat ten

---

3  $C^+$  staat voor gevolg of consequent,  $C^+$  betekent een positieve,  $C^-$  een negatieve consequent. Het eerste plus- of minteken betekent dat deze wordt toegevoegd/ontvangen of weggenomen/verloren.  $+C^+$  staat zo voor een positieve beloning, dat wil zeggen dat op de respons een positieve consequentie volgt.  $+C^-$  staat voor een positieve straf, dat wil zeggen dat op de respons een negatieve consequentie of straf volgt.  $-C^+$  is een negatieve beloning, een negatieve consequentie wordt weggenomen.  $-C^-$  ten slotte is de negatieve straf, een beloning wordt weggenomen.

coste van de accuratesse en andersom. Stanovich & Pachella (1976) toonden deze *speed-accuracy trade-off* aan bij het benoemen van alfanumerieke stimuli. Gegeven de combinatie van te hoge taakeisen in verhouding tot de mogelijkheden of beperkingen van de leerling, zijn spellen en raden als voorkeursstrategieën voorbeelden van aanpassing en een hervonden evenwichtstoestand, zoals vanuit de dynamische systeemtheorie aangereikt.

In de praktijk van de dyslexiebehandeling is bekend dat hardnekkig spellende lezers via allerlei maatregelen van overmatig spellen afgebracht kunnen worden, maar dan helaas nogal eens radende lezers worden. Op grond van de hiervoor gegeven interpretatie vanuit de leergeschiedenis is te verwachten dat bij alle lezers - in Nederland bij de vigerende leesdidactiek - spellend lezen met de leeftijd van de leerlingen af zal nemen en dat bij zwakke lezers radend lezen toeneemt. Als in de toegepaste didactiek de spellende route alleen maar wordt gestraft, en de leerling is niet in staat het gewenste gedrag, correcte directe woordherkenning, te laten zien, dan is raden nog de enige uitweg om te voldoen aan de eis van vlotheid. Vooral een toename van radend lezen is dan een indicatie dat de leerling zich ontwikkelt in de richting van een nieuw, maar eveneens ongewenst evenwicht. Het betekent dat de toegepaste didactiek tekortschiet en het met de leesontwikkeling van een leerling niet goed gaat. Wanneer leestaken zo zorgvuldig zijn afgestemd op de instructiebehoefte van de leerling dat hij overwegend in staat is de leestaak voldoende accuraat uit te voeren en het snelheidsaspect kan oefenen, zal snelheidswinst *niet* ten koste gaan van de accuratesse.

In Angelsaksisch onderzoek wordt meestal slechts gebruikgemaakt van een accuratessemaat, zoals Lyon & Moats (1997) aangaven. Als accuratesse en vlotheid wel worden onderscheiden, dan worden deze lange tijd gezien als afzonderlijk aspecten van het lezen en als aspecten waartussen mogelijk een zogenaamde *trade-off* bestaat. In het Nederlands lijkt accuratesse slechts in het begin van de leesontwikkeling een onderscheiden aspect. Beide aspecten worden als afhankelijke variabelen meegenomen.

### methodiek

Meyer & Felton (1999) geven een opsomming van effectieve methodiekcomponenten:

- naast decodeerinstructie aparte aandacht voor leestempo;
- kort en frequent trainen op vlotheid;
- herhaald lezen van losse woorden;
- veel lezen van tekst op instructieniveau;
- herhaald lezen van tekst;
- vooruitgang meten;
- hoe zwakker de leesvaardigheid, hoe groter de noodzaak van *adult guidance*;
- vooraf oefenen van woorden uit de te lezen tekst, oefenen van korte passages, voordoen van expressief lezen - *modeling*.

Herhaald lezen kent verschillende vormen, met als meest voor de hand liggende een tekst een aantal malen achterelkaar lezen (Samuels 1979). Sindelar e.a. (1990) toonde een positief effect hiervan aan op accuratesse, vlotheid en *recall*, zowel bij derde- als vijfdeklassers, zowel bij *learning disabled* als bij gemiddelde lezers en zowel met teksten op instructie- als op beheersingsniveau. Dowhower (1987) vond dat zowel

samen lezen, ook *paired reading* of duolezen genoemd, als *independent practise* de leessnelheid, de accuratesse, het leesbegrip en de leestoon van zwakke tweedeklassers bevorderden, maar dat beide vormen niet verschilden in effectiviteit. Homan e.a. (1993) vergeleken herhaald lezen met verschillende *assisted nonrepetitive* oefenvormen: echo lezen (leerkracht lees zin voor zin voor, leerling leest zin na zin na), samen lezen, *cloze reading* (leerkracht leest, laat stiltes vallen waarin de leerling het volgende woord leest, waarna de leerkracht weer verder gaat). Het onderzoek betrof zwakke zesdeklassers; alle vormen waren effectief en verschilden onderling niet qua effectiviteit. Levy e.a. (1997) combineerden het trainen van inhoudswoorden uit een tekst met herhaald lezen van de tekst en rapporteren positieve effecten op accuratesse, tempo en begrip bij zwakke vierdeklassers.

De hier beschreven methodieken betreffen zogenaamde antecedente technieken; het aanbod van het leesmateriaal en de instructie worden zo ingericht dat de kans op een juiste repons zo groot mogelijk wordt. Ook verschillende vormen van *feedback*, consequente technieken, zijn onderzocht. *Feedback* is effectiever dan geen *feedback* (Merrett & Thorpe 1996). Een fout gelezen woord juist voorzeggen (*modeling*) en voordoen van *sounding out* zijn effectiever dan niet-specifieke *feedback*. *Modeling* is effectiever dan *prompting* Perkins (1988). De effectiviteit van de beschreven methodiekcomponenten lijkt groter als de methodiek wordt toegepast door professionals in vergelijking met vrijwilligers of medeleerlingen en ook bij individuele toepassing in vergelijking tot groepjes leerlingen (National Reading Panel 2000).

Wat opvalt is dat het gewenste instructiegedrag zelden op microniveau wordt beschreven. Een uitzondering hierop vormt de methodiek *pause, prompt, praise* (McNaughton e.a. 1987, Glynn e.a. 1992). Deze methodiek is in oorspronkelijke vorm gepresenteerd als een, in TGL-termen, functieanalyse van Clay's programma (1993). Haar programma had weliswaar een sterk *whole language* accent, maar de kern van *pause, prompt, praise* was een consequentie toepassing van gedragstherapeutische principes op het geven van leesinstructie. Om die reden is voor deze methodiek gekozen als basis voor de geprotocolleerde remedial teaching onder de naam *wacht, hint, prijs* (WHP).

De essentie van WHP is de toepassing van effectieve gedragstherapeutische technieken binnen de context van individuele leesinstructie. De functieanalyse is, in vergelijking met het oorspronkelijke *pause, prompt, praise*, veel strikter gebaseerd op de leerprocedure voor woordherkenning. Dat betekende vooral een veel nadrukkelijker toepassen van decodeerhints en een beperkte plaats voor betekenis- en taalhints. De methodiek is in vele toepassingen onderzocht en effectief gebleken: toepassing door een gekwalificeerde remedial teacher, toegepast door niet-gekwalificeerde leeshulpen, binnen *peer tutoring* met goed lezende oudere leerlingen als tutor, met oudere zwakke lezers als tutor of met goede lezers van ongeveer dezelfde leeftijd (Wheldall e.a. 1987, Merrett 1988a, Medcalf 1989, Leach & Siddal 1990, Houghton & Bain 1993). Zonder uitzondering betrof de toepassing in de aangehaalde onderzoeken oudere leerlingen met een achterblijvende leesvaardigheid, vanaf klas 4 tot en met

leerlingen van het voortgezet onderwijs en zelfs volwassenen.

### ***wacht, hint, prijs***

Als een leerling bij het lezen hapert of een fout maakt, is de gebruikelijke reactie van leerkrachten om onmiddellijk in te grijpen (Allington 1980, Shake 1986, Wheldall e.a. 1992). Dit heeft twee negatieve effecten. Ten eerste bevestigt het de incompetentie van de leerling; het falen van de leerling krijgt nadruk. Ten tweede ontnemt het de leerling de mogelijkheid om op eigen kracht toch het moeilijke woord te vinden en daarmee voor zichzelf een oefenmoment te creëren. Dit kan door bijvoorbeeld het verklanken van het woord in delen en het woord daarna via auditieve synthese alsnog te vinden. De andere mogelijkheid is de zin uit te lezen en zo een semantische en/of syntactische context te laten ontstaan, waardoor de leerling het woord toch kan lezen. *W o o r d e n* worden in zinsverband beter, dat wil zeggen sneller en met minder fouten gelezen (Van der Leij 1983), mits de moeilijkheidsgraad van de tekst aansluit bij het niveau van de leesvaardigheid (Adams 1991).

In plaats van direct ingrijpen schrijft de WHP-methodiek voor dat de leerkracht de eerste vijf seconde helemaal niets doet: *wacht*. Afgezien van het effect dat het wachten de leerling de kans geeft om op eigen kracht het juiste woord te vinden, brengt het rust en ontspanning in de interactie tussen leerkracht en leerling. Lukt het de leerling niet binnen vijf seconden een bepaald woord te lezen, dan geeft de leerkracht een *hint*. De essentie van een hint is dat het een discriminatieve stimulus (Sd) is, die het aantal responsalternatieven beperkt, waardoor de kans op een juiste respons toeneemt. Op deze begrippen uit de toegepaste gedragsleer zal in paragraaf 4.2 nader worden ingegaan. Ten slotte krijgt de leerling feedback omtrent de juistheid van zijn respons: *prijs*. De methodiek is in bijlage 1 nader beschreven en in schema gezet.

De leerling leest teksten op zogenaamd instructieniveau, dat wil zeggen een niveau waarop de leerling 80 à 90 % van alle woorden voldoende snel en direct goed leest en dus met 10 à 20% van de woorden nog moeite heeft (Merrett 1988b, Adams 1991), met als vuistregel: gemiddeld nog één moeilijk woord per zin. Alleen dan leveren het lezen van tekst en het uitlezen van een zin de context die het de leerling mogelijk maakt om een te moeilijk woord toch op eigen kracht te vinden. Maar hoe zorgvuldig de leesboeken ook gekozen zijn, een tekst kan toch te moeilijk zijn. Niet omdat deze verkeerd gekozen is, in dat geval neemt men een andere, maar omdat een bepaalde passage moeilijk is of omdat de leerling moe is of wordt, bijvoorbeeld in de tweede helft van de remedial teaching sessie. Daarom is de remedial teachers ook een reeks mogelijkheden aangereikt waarvan uit onderzoek bekend is dat deze effectieve orthodactische maatregelen zijn, zodat zij op elk moment de moeilijkheidsgraad van de leestaak naar de behoefte van de leerling aan kunnen passen. Deze zijn ontleend aan de hiervoor beschreven andere methodieken: een passage meer dan één keer laten lezen, eenmaal voorlezen, moeilijke woorden voorzeggen, zinnen om-en-om lezen, duo-lezen, moeilijke woorden vooraf geïsoleerd oefenen. De remedial teacher kan daarbinnen zelf een keuze maken, of overlegt de keuze met de leerling, wat over het



algemeen de motivatie ten goede komt (Van Bilsen 1992).

De verwachting was dat de zwakke lezers de neiging zouden hebben tot taakvermijgend gedrag. Hieronder wordt verstaan passief vermijdingsgedrag of hulpeloosheid, actief vermijdingsgedrag of taakirrelevant gedrag en gebrek aan doorzetting (Onatsu-Arvilommi & Nurmi 2000). Voldoende *time-on-task* was daarom steeds een belangrijk aandachtspunt.

Essentiële kenmerken van de remedial teaching waren setting: individueel, frequentie: viermaal per week, duur: een half uur, lengte: een half (school)jaar, *time-on-task*: minstens 80% en adequate toepassing van *wacht*, *hint*, *prijs*.

### 3.3 Opzet

Het onderzoek naar de effectiviteit van de remedial teaching vond plaats binnen de context van de gewone basisonderwijspraktijk. In het remedial teaching onderzoek ontving de helft van alle zwakke lezers in de groepen 4 naast *care as usual* gedurende een half (school)jaar extra viermaal per week een half uur individuele remedial teaching in de periode september tot februari. De andere helft was in die periode controlegroep.

#### instrumenten

Het instrument dat primair gebruikt is voor screening en selectie en voor het vaststellen van vooruitgang in technische leesvaardigheid is de Drie-Minutentoets (DMT), kaart 2, parallelversies A,B,C (Verhoeven 1993). Leerlingen lezen een rij losse woorden gedurende één minuut, de score is het aantal goed gelezen woorden.

Voor de selectie van leerlingen is tevens gebruik gemaakt van de AVI-kaarten 1 en 2 (Visser e.a. 1994). Dit zijn korte teksten met respectievelijk uitsluitend MKM-woorden en ongeveer 80% MKM- en 20% MMKM/MKMM-woorden.

Behalve de gebruikelijke ruwe score op de DMT-2 hebben de leerlingen ook standaard de eerste 25 woorden gelezen. Hiervan zijn zowel het aantal goed gelezen woorden gescoord als de tijd die de leerling er voor nodig had. Het aantal goed gelezen woorden is omgerekend naar het percentage goed gelezen woorden als accuratessemaat. De benodigde tijd is omgerekend naar het aantal gelezen woorden per minuut, dus inclusief de fouten, als tempomaat.

Naast de DMT zijn vier gangbare RAN-taken afgenomen: benoemen van letters, cijfers, plaatjes en kleuren (Denckla & Rudell 1974). De toetskaarten bestonden steeds uit 9 x 4 elementen die zo snel mogelijk benoemd moesten worden: een traditionele toetskaart letterbenoemen met de 36 verschillende grafemen, de cijfers 1-9 in willekeurige volgorde, vier verschillende plaatjes en vier verschillende gekleurde balkjes. De tijd die de leerlingen voor elke kaart nodig hadden is omgerekend naar het aantal benoemde eenheden per minuut.

De remedial teaching is afgesloten met het maken van een bandopname van 15 minuten lezen door elke leerling uit de getrainde groep van een niet eerder gelezen passage

uit het boek waar hij of zij aan het einde van de onderzoeksperiode mee bezig was. Daarna las de leerling uit de controlegroep met wie hij een *matched pair* vormde hetzelfde stuk tekst. Op deze wijze zijn accuratesse, het percentage goed gelezen woorden, en tempo, het gemiddelde aantal gelezen woorden per minuut, vastgesteld van het lezen van tekst aan het einde van de trainingperiode.

Ten slotte zijn ongeveer halverwege de onderzoeksperiode nog eenmalig de woordenschattoets uit het CITO-leerlingvolgsysteem en de Schoolgedrag Beoordelingslijst (SCHOBL-R, Bleichrodt e.a. 1993) afgenomen.

### **selectie van de scholen**

Om te kunnen werken met een controlegroep en daarbij te kunnen controleren voor de factor school, kwamen alleen grote scholen in aanmerking met minimaal 50 leerlingen in de toekomstige groep 4. Daarbij werden ook op voorhand uitgesloten de grote scholen die uit een fusie zijn ontstaan. Om te beginnen is een uitdraai gemaakt van alle scholen op basis van hun totale aantal leerlingen op de meest recente teldatum. Vervolgens is aan schoolcontactpersonen van het Centrum Educatieve Dienstverlening gevraagd welke scholen in groep 3 een effectieve methode voor aanvankelijk lezen gebruikten (Hol e.a. 1995) en daarbij een adequaat systeem van leerlingzorg voerden. Als belangrijkste indicator daarvoor is genomen of de school reeds enkele jaren een leerlingvolgsysteem hanteert, de resultaten daarvan in overzichten verwerkt en ook gebruikt als basis voor de leerlingzorg (Houtveen 1992). Dit resulteerde in een lijst van achttien potentiële projectscholen. De scholen zijn in volgorde van grootte benaderd.

In een kennismakingsgesprek is uitvoerig gesproken over de voorwaarden voor deelname en de wederzijdse verplichtingen. Aan de kant van de scholen ging dit om:

- het informeren van ouders en regelen van toestemming van ouders voor deelname aan het onderzoek op een wijze die paste binnen het beleid van de school;
- het aanleveren van de meest recente DMT-gegevens van alle leerlingen uit groep 3;
- er zouden minimaal zes leerlingen moeten zijn die aan de criteria voldeden;
- scholen - en ouders - moesten accepteren dat leerlingen door loting zouden worden toegewezen aan de eerste of de tweede experimentele periode, respectievelijk deelonderzoek 3 of 4;
- de leerlingen van de tweede experimentele groep vormden de controlegroep voor onderzoek 3;
- het onderwijs aan alle leerlingen uit de experimentele en de controlegroep gedurende het eerste halfjaar, onderzoek 3, diende gelijk te zijn en volgens de normale gang van zaken te verlopen;
- het organiseren van individuele remedial teaching voor de leerlingen uit de experimentele groepen, zowel onderzoek 3 als 4, tweemaal per week gedurende 30 minuten;
- garanderen dat de remedial teaching in minstens twaalf van de vijftien weken tweemaal per week door zou gaan;
- de remedial teaching diende volgens protocol te verlopen;
- de remedial teachers in de gelegenheid stellen deel te nemen aan acht curcus/coachingsbijeenkomsten in de loop van het projectjaar.

Het Pedologisch Instituut verplichtte zich:

- de school, indien daaraan behoefte bestond, te voorzien van informatie die gebruikt zou kunnen worden voor het informeren van de ouders;

- de leerlingen in het kader van de deelonderzoeken 3 en 4 een additionele experimentele computergestuurde woordtraining te geven, tweemaal per week gedurende een half uur door daartoe specifiek opgeleide logopedisten;
- gratis een cursus- en coachingstraject aan te bieden voor twee deelnemers per school, hetzij twee remedial teachers, dan wel een remedial teacher en de intern begeleider;
- bij positief resultaat de scholen een exemplaar aan te bieden van het computergestuurde woordtrainingprogramma.

De eerste tien benaderde scholen bevestigden allemaal mee te willen doen, waarna de werving is gestopt, omdat met deze tien scholen het maximaal aantal plaatsbare leerlingen al bereikt zou zijn. Hiervan is een school later alsnog uitgesloten, omdat er dat jaar in groep 3 slechts drie uitvallers bleken te zijn.

Een onderzoek in de dagelijkse praktijk van scholen levert vele praktische overwegingen en problemen op. Op een aantal punten was het daarom niet mogelijk op alle scholen dezelfde lijn te trekken. Voor het onderzoek kwamen leerlingen in aanmerking die 10 maanden leesonderwijs hadden genoten, didactische leeftijd (DL) 10. Dat betekende dat leerlingen die een jaar langer hadden gekleuterd mee mochten doen. Ook leerlingen die zouden blijven zitten in groep 3 mochten meedoen, hun didactische leeftijd was immers ook 10 en ze behoorden tot de doelgroep. Maar leerlingen die al eerder in groep 3 of groep 4 waren blijven zitten werden uitgesloten van deelname, omdat zij inmiddels 20 maanden leesonderwijs achter de rug hadden. Twee scholen kozen er echter nadrukkelijk voor het onderzoek te beperken tot leerlingen uit de groepen 4. Men meende dat deze leerlingen door het doubleren al veel extra leesinstructie zouden krijgen. Deze acht leerlingen zijn wel gevolgd met het oog op deelname in deelonderzoek 4.

Ten slotte verschilden de organisatorische invulling van de remedial teaching nog op de volgende wijze:

- 1 remedial teacher deed alle leerlingen 2 x per week;
- de leerlingen waren verdeeld over 2 remedial teachers, elk deed haar eigen leerlingen 2 x per week;
- 2 remedial teachers zagen alle leerlingen ieder 1 x per week.

### **kenmerken van de scholen**

Enkele gegevens van de negen scholen die in het onderzoek participeerden, zijn weergegeven in tabel 4. De scholen hadden de volgende kenmerken: 4 scholen waren overwegend 'witte' scholen, 3 scholen hadden een gemengde populatie en 2 scholen waren overwegend 'zwarte' scholen. Er waren een jenaplanschool en twee montessorischolen. Zeven scholen werkten met de methode Veilig leren lezen (Zwijzen 1991), een school met de methode Leessleutel (Malmberg 1990) en een school met het Zwaluwprogramma (Bekadidact 1993).

De leerlingen zijn getoetst met de DMT, kaart 2. Voor de verschillende toetsmomenten geeft de normtabel een scorebereik voor het zogenaamde E-niveau, dat wil zeggen de scores die behaald worden door de zwakste 10% van alle leerlingen in Nederland op dat moment. De leerlingen die op E-niveau scoren worden in het algemeen gezien als uitvallers. Als de geselecteerde scholen representatief zouden zijn voor de basisscholen

in Nederland, dan zou 10% van de leerlingen eind groep 3 op E-niveau moeten scoren. Het gemiddelde aantal E-leerlingen (tabel 4), was 12,2 ofwel 14,6%. Dit wijkt significant af van de verwachte 10% ( $t(8) = 4,440$ ,  $p < .01$ ). Het percentage uitvallers op technisch lezen eind groep 3 ligt op de deelnemende scholen boven het landelijk gemiddelde.

Tabel 4: Aantallen en leesniveaus van de leerlingen in de negen geselecteerde scholen.

	Min.	Max.	Som	M (SD)
totaal aantal leerlingen	410	859	5576	620 (165)
totaal aantal leerlingen in groep 3	52	125	780	87 (26)
aantal E-leerlingen in groep 3	7	16	110	12,2 (3,0)
percentage E-leerlingen in groep 3	10%	22%		14,6% (3,1)

### selectie van leerlingen

Medewerkers van de scholen hebben tegen het einde van het schooljaar aan alle leerlingen met een didactische leeftijd van 10 de DMT-2 afgenomen. Hieruit is het zwakste eenderde deel geselecteerd voor toetsing eind augustus door de medewerkers van het project. Op voorhand werden van verdere deelname uitgesloten de acht zittenblijvers in groep 3 op twee van de negen scholen, drie leerlingen van wie al bekend was dat zij in de loop van het projectjaar van school zouden gaan wegens verhuizing en twee leerlingen die reeds aangemeld waren voor plaatsing in het speciaal (basis)onderwijs.

Bij een score in augustus op de DMT boven D-niveau werd het onderzoek gestaakt. Aan leerlingen met een D- of E-score werd een uitgebreide serie toetsen afgenomen, waaronder de AVI-kaarten 1 en 2. Uiteindelijk is de selectie van leerlingen voor het onderzoek gemaakt op basis van de DMT-score in augustus. Als aanvullend selectie-criterium gold een AVI-beheersingsniveau van maximaal 1. AVI-1 beheersing kan, met de nodige voorzichtigheid, bij een didactische leeftijd van 11 maanden als een score op D-niveau worden aangemerkt, AVI-2 als een score op C- of B-niveau. De combinatie van een DMT-score op E-niveau en een AVI-beheersingsniveau 2 kwam niet voor en. Slechts vier leerlingen haalden AVI-kaart 1, de overige haalden kaart 1 nog niet.

Per school werden de zwakste acht of tien lezers gekozen voor deelname in het onderzoek, in totaal 78 leerlingen. Daarnaast zijn op acht van de negen scholen de vier of vijf leerlingen geselecteerd die op de DMT net boven de experimentele groepen scoorden, in totaal 34 leerlingen. Deze leerlingen werden als een quasi-controlegroep gedurende de projectperiode gevolgd met het oog op de follow up en om de kwaliteit van de selectie begin groep 4 vast te kunnen stellen, met name de vraag of het aantal zogenaamd vals negatieven acceptabel was, zie hoofdstuk 2.

### leerlingenmerken

De 78 leerlingen die werden geselecteerd, waren afkomstig uit minimaal twee en maximaal vijf parallelgroepen per school. Zij zijn ingedeeld in 39 *matched pairs* op basis van hun ruwe score op de DMT-2 in augustus, DL 11. Per school werden twee vergelijk-

bare groepen samengesteld door de leerlingen per matched pair aan een van beide toe te wijzen. 29 koppels waren van hetzelfde geslacht. Daarvan hadden 10 ook dezelfde leerkracht. Zij werden door loting aan een groep toebedeeld. De overige leerlingen werden zo toegewezen dat de groepen per school op het punt van de leerkracht en op het punt van de sekse van de leerlingen zo vergelijkbaar mogelijk waren. Uiteindelijk werd door loting bepaald welke groep in de periode september-februari de training zou krijgen. Deze groep wordt aangeduid met de code T11; zij ontvingen remedial teaching vanaf een didactische leeftijd van 11 maanden. Halverwege de onderzoeksperiode is een van de leerlingen vertrokken, waardoor groep T11 uiteindelijk 38 leerlingen telde. De groep die in de periode september-februari controlegroep was, zou vanaf februari voor training in aanmerking komen, dus vanaf een didactische leeftijd van 16 maanden. Daarom is deze groep aangeduid als T16.

Tabel 7: Beschrijvende informatie over de deelnemende leerlingen aan het rt-onderzoek.

VARIABLE	groep	N	M	(SD)		
SEKSE (0=meisje; 1=jongen)	T11	39	0,67	0,48		
	T16	39	0,62	0,49		
	C4	34	0,50	0,51		
DMT-2, ruwe score, DL=11	T11	39	7,03	2,96		
	T16	39	7,21	2,96		
	C4	34	12,47	2,00		
CITO-woordenschat, ruwe score	T11	36	41,17	4,93		
	T16	38	42,11	5,30		
	C4	34	40,06	7,28		
Woordenschat, niveau E=1, D=2, C=3, B=4, A=5	T11	36	3,03	1,18		
	T16	38	3,32	1,12		
	C4	34	2,97	1,38		
SCHOBL-extraversie, Z-scores	T11	29	0,17	0,94		
	T16	27	0,10	0,95		
	C4	25	0,00	1,01		
SCHOBL- werkhouding, Z-scores	T11	29	-0,30	0,68		
	T16	27	-0,49	0,71		
	C4	25	-0,36	0,63		
SCHOBL-aangenaam gedrag, Z-scores	T11	29	0,11	1,10		
	T16	27	-0,15	1,05		
	C4	25	0,32	0,94		
SCHOBL- emotionele stabiliteit, Z-scores	T11	29	0,21	0,98		
	T16	27	0,18	0,90		
	C4	25	0,01	0,83		
	leerling- gewicht	T11	GROEP T16	C4	N	Totaal %
SE-score	1,00	21	22	19	62	65,3%
	1,25	7	5	5	17	17,9%
	1,90	6	5	5	16	16,8%
Totaal		34	32	29	95	100%

Ten slotte waren er nog de 34 leerlingen uit groep 4 van de quasi-controlegroep. Zij worden aangeduid met de code C4. In alle gevallen waren deze leerlingen *op hun school* de net iets betere lezers dan de leerlingen uit de groepen T11 en T16, maar 17 van de 34 leerlingen uit groep C4 hadden in augustus eveneens een DMT-score op E-niveau. De leerlingen waren op het moment van screening en selectie uitsluitend bekend onder een code. De volgende beschrijvende informatie is over de leerlingen verzameld (tabel 7).

- In groepen T11 en T16 bleek de sekseratio ongeveer 2:1, in quasi-controlegroep C4 1:1.
- De drie groepen zijn vergelijkbaar qua SE-score, het zogenaamde leerlinggewicht in het kader van het onderwijsachterstandsbeleid.
- De woordenschattoets uit het CITO-leerlingvolgsysteem is afgenomen om vast te stellen dat bij geen van de deelnemende leerlingen sprake is van een te geringe Nederlandse woordenschat. De normering van de woordenschattoets volgt de schaal die bij het CITO-LVS gebruikelijk is. Meer leerlingen haalden een E/D-score, 12,0 + 20,3% in plaats van 10 + 15%, 25% een C-score, 29,6% een B-score en 13,1 % een A-score in plaats van de te verwachten 25%. De gemiddelde woordenschatscore ligt iets onder het landelijk gemiddelde ( $t(107) = -2,666; p < .01$ ).
- Ten slotte is op zeven scholen de SCHOBL-R afgenomen. Het afnemen van deze gedragsvragenlijst in dit kader had tot doel ernstige verdenking uit te sluiten van co-morbide problematiek waarvan bekend is dat deze dikwijls met leesproblemen samengaat (Van Leeuwen & Vieijra 1996, Koot 2000, Vieijra e.a. 2002) en de aanpak van de betreffende leerlingen sterk kunnen bepalen. In tabel 7 zijn de standaardcores als z-scores weergegeven op de vier factoren die de SCHOBL onderscheidt: extraversie, werkhouding, aangenaam gedrag en emotionele stabiliteit. Een lage score geeft steeds de negatieve pool van de dimensie aan. De handleiding van de SCHOBL geeft geen criteria voor de interpretatie van de factorscores. Gebruikelijk is een afwijking van meer dan twee standaarddeviaties van het gemiddelde als klinische range aan te duiden (Verhulst e.a. 1996) en een score in de klinische range of meer dan een score bij de meest extreme 10% op te vatten als een indicatie voor een stoornis (Resing e.a. 2002). De drie groepen verschilden onderling op geen van de factoren. Maximaal drie leerlingen per subgroep scoorden op een van de factoren bij de meest extreme 10%, geen enkele leerling op meer dan een factor en geen enkele leerling behaalde een klinische score op enige factor.

Afgezien van een verschil in technische leesvaardigheid zijn de geselecteerde leerlingen op de gekozen variabelen niet afwijkend van de norm. Dit geldt ook voor de woordenschatscores, als rekening wordt gehouden met het feit dat de leerlingen afkomstig zijn uit een van de grote steden. Voorts verschilden de experimentele groepen T11 en T16 op geen van de onderzochte variabelen van elkaar.

De opzet was dat alle leerlingen zouden meedoen die bij aanvang van het nieuwe schooljaar op E-niveau scoorden op de DMT. De aanduiding E-niveau is conform de normering die gebruikt wordt binnen het CITO-leerlingvolgsysteem. De zwakste 10% van de leerlingen worden E-leerlingen genoemd, de 15% daarboven D-leerlingen. De scholen hadden echter op voorhand aangegeven wat de maximale remedial teaching capaciteit was die zij voor het onderzoek konden vrijmaken; voor zes scholen lag dit op vier leerlingen per halfjaar, voor drie scholen op vijf leerlingen. Dit aantal bleek meestal lager en bij geen van de scholen hoger dan het aantal leerlingen dat op grond van hun leesscore in aanmerking zou komen. De leerlingen die uiteindelijk voor deelname geselecteerd zijn, haalden aan het begin van de onderzoeksperiode een gemiddelde score op de DMT-2 van 7,12 (SD 2,94). Dit komt overeen met een z-score

van  $-1,93$  (SD  $.42$ ) en een percentielscore van  $.03$ , dit is significant lager dan de verwachte waarde van  $-1,65$  ( $t(76) = -75,424$ ,  $p < .001$ ).

### 3.4 Ontwikkeling

Ook in het ontwikkelingsproces van de remedial teaching is waar mogelijk gebruik gemaakt van elementen ontleend aan de toegepaste gedragsleer.

#### cursus

In het schooljaar dat het rt-onderzoek liep zijn acht instructie- en coachingbijeenkomsten van  $2\frac{1}{2}$  uur gehouden. Deelnemers waren dertien remedial teachers die in het onderzoek als zodanig participeerden, van wie er twee tevens intern begeleider waren en naast remedial teacher ook deeltijds klassenleerkracht, en vier intern begeleiders die geen remedial teacher waren of niet als zodanig in het onderzoek participeerden. Ook namen de vier logopedisten deel die de woordtraining verzorgden. Twee remedial teachers die wel als zodanig in het onderzoek participeerden, hebben, om verschillende redenen, niet aan de cursus deelgenomen.

Door deze samenstelling hebben de inzichten, handreikingen en bejegening van de leerlingen binnen de scholen een bredere verspreiding gekregen dan alleen de remedial teaching sessies. Deze varieerde van toepassing bij andere leerlingen tot het doorgeven van de *wacht*, *hint*, *prijs* methodiek in het kader van *peer tutoring*. Dit is aangemoedigd en waar mogelijk ondersteund, omdat het eraan bijdroeg te voorkomen dat de manier van werken een geïsoleerde activiteit zou worden.

Als mensen zeggenschap krijgen over veranderdoelen, bevordert dat de motivatie tot verandering (Van Bilsen 1992). Tijdens de eerste bijeenkomst zijn daarom persoonlijke leerdoelen geïnventariseerd. Deze zijn thematisch gegroepeerd en vormden zo de basis voor het cursusprogramma. De deelnemers is een certificaat in het vooruitzicht gesteld op voorwaarde van minimaal 80% aanwezigheid en voldoende beoordeling van de uitvoering van de *wacht*, *hint*, *prijs* methodiek.

#### logboek en nieuwsbrief

Het Reading Recovery Programma (Clay 1993) is een van de eerste onderzochte systematische invoeringen van intensieve individuele remedial teaching in de reguliere onderwijspraktijk. De achilleshiel van haar opzet bleek de moeite die het kost om de remedial teaching voldoende vaak door te laten gaan en de tijd te laten duren die ervoor staat. Individuele zorg aan leerlingen vervalt dikwijls als eerste wanneer op een school organisatorische problemen moeten worden opgelost, zoals het invallen bij ziekte van klassenleerkrachten. Om het niet doorlaten gaan van de afgesproken remedial teaching en/of het niet volmaken van het afgesproken half uur tegen te gaan, is om te beginnen voor het hele schooljaar een rooster gemaakt met werkbare weken, dat aan alle betrokkenen is gegeven. Verder is gewerkt met een logboek. Dit had de simpele

vorm van een scheurblok, per leerling moest per sessie een blaadje worden ingevuld. Tijdens de cursusbijeenkomsten werden deze ingeleverd. Om het blijven invullen van het logboek te bekrachtigen is, bij wijze van monitoring, in het begin van het projectjaar een samenvatting van de resultaten kort na iedere cursusbijeenkomst in de vorm van een nieuwsbrief aan de deelnemers toegestuurd. Het logboek bood ook ruimte om inhoudelijke vragen stellen of opmerkingen te maken. Dit gaf de gelegenheid om in de nieuwsbrief op inhoudelijk thema's in te gaan. Regelmatig kon op deze wijze stof die tijdens de cursusbijeenkomst aan de orde was geweest, herhaald worden in de vorm van concrete gedragsinstructies. Het logboek is ook gebruikt om de remedial teachers scherp te houden op de cruciale elementen in elke remedial teaching sessie: het bewaken van voldoende *time-on-task* en toepassing van de WHP-methodiek.

### observaties

Er zijn twee observatierondes gehouden. Gedurende de eerste ronde, halverwege de onderzoeksperiode, zijn bij elk van de dertien remedial teachers twee sessies met twee verschillende leerlingen geobserveerd, in totaal 26 sessies. De observatiegegevens van deze ronde zijn gebruikt voor nadere analyse. Een tweede observatieronde had slechts als doel het volhouden van de verworven vaardigheden te bekrachtigen. De gegevens uit deze observaties zijn alleen gebruikt voor het geven van persoonlijke feedback.

Tevoren is verteld dat de observaties gericht zouden zijn op de *time-on-task* en de juiste toepassing van *wacht*, *hint*, *prijs*. Voor het meten van *time-on-task* was een eenvoudige methode beschikbaar (Vonk 2000), zie bijlage 2. De observator turft per minuut eenmaal of de leerling *on task* dan wel *off task* was. Deze vorm van *time sampling* bleek binnen de context van dyslexiebehandelingen een betrouwbare meting op te leveren, met als criterium *event sampling* tijdens continue observatie.

De interbeoordelaarbetrouwbaarheid van de observaties is ook voor de remedial teaching vastgesteld. Drie remedial teachers meldden zich in de eerste weken van de onderzoeksperiode vrijwillig met een door henzelf voorgedragen moeilijke leerling voor het laten maken van een video-opname. De opnamen zijn door twee observatoren afzonderlijk tweemaal gescoord. Eenmaal op basis van een continue observatie en eenmaal op basis van *time sampling*: een score op elke hele minuut. Bij de continue observatie zijn elke minuut twee perioden van 20 seconde beoordeeld op een vierpuntschaal: 0 geheel *off task*, 1 minder dan de helft van de tijd *on task*, maar niet 0, 2 meer dan de helft van de tijd *on task*, maar niet 3, 3 geheel *on task*. Voor de vergelijking tussen de twee methoden van observeren is de gemiddelde score gedichitomiseerd, gemiddelde score 0 of 1 *off task*; score 2 of 3 *on task*. Op basis van de continue observatie waren deze leerlingen gemiddeld 68,3% van de tijd *on task*, de *time sampling* methode leverde 73,2% van de tijd *on task* op. De beoordelingen op basis van continue observatie stemden voor 97% overeen, op basis van *time sampling* voor 96%; de overeenstemming tussen de scores volgens de twee verschillende methodes was 70,7% ( $\chi^2(1) = 7,242$ ;  $p < .01$ ).



Het tweede doel van de observaties was vast te stellen in welke mate de *wacht*, *hint*, *prijs* methodiek volgens protocol werd toegepast. Daartoe werd een observatieformulier ontwikkeld met instructie op basis van Wheldall e.a. (1992), bijlage 2. Dezelfde video-opnamen zijn ook met dit observatieschema door twee observatoren onafhankelijk van elkaar gescoord. Als maat is genomen het aantal adequate toepassingen van een aspect van de methodiek, bijvoorbeeld de toepassing van *wacht*, als percentage van het aantal keren dat dat aspect naar het oordeel van de observator aan de orde was. De interbeoordelaarbetrouwbaarheid bij de verschillende categorieën varieerde van .64 voor het aspect 'prijs aan het einde van de zin' tot .77 voor het aspect 'wacht'. Het percentage adequate toepassingen van het prijzen aan het einde van elke zin was bij alle observaties zeer hoog, minimaal 85%, zodat de relatief lage correlatie voor deze gemakkelijk te beoordelen gedragscategorie waarschijnlijk vooral het gevolg is van een geringe spreiding van de scores.

Tabel 5: Indicatoren voor de mate van behandelingsintegriteit van de remedial teaching volgens protocol zoals gemeten bij twee observaties per deelnemer, bij de elf cursusdeelnemers en de twee niet-deelnemers afzonderlijk.

	Cursusdeelnemer	N	Minimum	Maximum	M	(SD)
% juist wachten	ja	22	92	100	99,1	2,0
	nee	4	73	100	90,8	12,7
% juist prijzen na wacht	ja	22	52	93	71,1	12,5
	nee	4	23	100	63,8	42,0
% juiste hints	ja	22	83	100	96,5	5,7
	nee	4	36	100	70,3	34,6
% juist prijzen na hint	ja	22	78	100	94,1	6,7
	nee	4	86	100	91,8	7,0
% juist prijzen aan einde van zin	ja	22	51	100	86,4	15,8
	nee	4	19	98	65,8	38,3
sessieduur, minuten	ja	22	29	34	31,3	2,1
	nee	4	19	36	27,8	3,8
% time-on-task	ja	22	65	100	87,4	7,9
	nee	4	48	94	71,3	18,8

Tijdens de eerste ronde van observaties duurden de sessies gemiddeld 28 minuten met een spreiding van slechts 4 minuten. De gemiddelde *time-on-task* was 85%, met een spreiding van 11,4%, maar bij de elf remedial teachers die ook aan de cursus deelnamen kwam deze uit op een gemiddelde van 87,4% (SD 7,9). Uitgaande van de vuistregel dat doelgedrag in 80% van de gevallen aanwezig moet zijn, waren zowel de duur van de sessies als de *time-on-task* in een ruime meerderheid van de gevallen voldoende. Bij de twee remedial teachers die de cursus niet volgden, was de *time-on-task* gemiddeld 72%. Zij bleken ook verantwoordelijk voor de uitbijters bij de meeste andere maten. In tabel 5 zijn daarom de resultaten afzonderlijk weergegeven voor de elf deelnemers en de twee niet-deelnemers. Van elke methodiekcomponent (wacht, prijzen na

wachten, hint, prijzen na een hint, prijzen aan het einde van elke zin) is steeds gescoord of deze terecht en juist werd uitgevoerd, niet terecht en/of niet juist werd uitgevoerd of dat deze uitgevoerd had moeten worden, maar niet is gedaan.

Omdat de remedial teachers met verschillende leerlingen, van verschillend leesniveau, met verschillende leesstof werkten, zijn alleen de percentages juist uitgevoerde interventies ten opzichte van het totaal van juist, niet juist en ten onrechte niet uitgevoerde elementen met elkaar vergelijkbaar. Op basis van de gegevens wordt geconcludeerd dat behandelingsintegriteit van de remedial teaching voor het overgrote deel ruim voldoende was; de methodiek werd volgens het boekje uitgevoerd.

Het gemiddelde resultaat van de deelnemers is in de volgende cursusbijeenkomst besproken. Uitvoerig is stilgestaan bij het snelle en boven verwachting goede resultaat. De resultaten van de twee niet-cursisten zijn besproken met de intern begeleiders van de betreffende scholen. Met deze remedial teachers zijn tevens aparte individuele instructiebijeenkomsten georganiseerd.

Enkele deelnemers meldten tijdens een cursusbijeenkomst moeite te hebben met het wachten. Op een van opnames van de remedial teachers viel op dat zij dit voortreffelijk deed. De rust die daarvan uitging en de ruimte die de leerling daardoor kreeg waren als het ware voelbaar. Van deze band zijn clips waarop het gewenste leerkrachtgedrag duidelijk te zien was, met toestemming van de betreffende remedial teacher, leerling en ouder, in de volgende cursusbijeenkomst aan allen getoond en besproken, bij wijze van modelleren.

Tabel 6: Principale componentenanalyse met varimaxrotatie op verschillende geobeserveerde gedragscategorieën bij de toepassing van de *wacht, hint, prijs* methodiek.

Gedragscategorie	Component	
	1	2
%-age juist wachten	<b>.94</b>	-.07
%-age juist prijzen na wachten	-.43	<b>.81</b>
%-age juiste hints	<b>.91</b>	-.22
%-age juist prijzen na hint	.05	<b>.77</b>
%-age juist prijzen aan einde van elke zin	-.19	<b>.73</b>

Verschillende deelnemers meldten dat zij specifiek moeite hadden met een bepaald aspect van de methodiek. Sommigen zeiden nog steeds het wachten moeilijk te vinden, anderen hadden juist moeite met het geven van veel positieve feedback, het prijzen. In de literatuur wordt de aard van de methodiek wel beschreven, maar niet de behandelingsintegriteit, en deze dus ook niet op een gedifferentieerde manier. Het idee dat de WHP-methodiek aan verschillende vaardigheden appelleert is uit de samenwerking met de remedial teachers naar voren gekomen. Met behulp van een factoranalyse is geëxploreerd of een dergelijke structuur in de data is terug te vinden. Op de accuratessematen is een principale componenten analyse uitgevoerd met varimaxrotatie. Dit leverde, op basis van het gebruikelijke criterium van een Eigenwaarde > 1, twee goed interpreteerbare factoren op; het percentage verklaarde variantie varieerde van 57%

voor juist prijzen aan het einde van een zin tot 90% voor juist wachten. De eerste factor wordt gevormd door de variabelen wacht en hint, de drie vormen van prijzen vormden de tweede factor (tabel 6).

De scores van de remedial teachers die in het onderzoek participeerden, lijken inderdaad twee relatief onafhankelijke factoren te reflecteren. Binnen deze groep hing de vaardigheid om adequaat te wachten wel sterk samen met de vaardigheid om de juiste hints te geven. Het geven van adequate positieve feedback, prijs, was voor de leden van deze groep echter iets anders. In termen van het basisschema van operante conditionering, zie paragraaf 4.2, zijn 'wacht' en 'hint' beide zogenaamde antecedente technieken, het prijzen is een consequente techniek. Voor het trainen in de WHP-methodiek en in het beoordelen van iemands bekwaamheid in het toepassen ervan, lijkt het aan te bevelen de verschillende deelaspecten afzonderlijk te behandelen.

### 3.5 Resultaten

#### algemeen

Allereerst werd de hypothese getoetst dat de intensieve remedial teaching in aanvulling op het reguliere onderwijsaanbod de technische leesvaardigheid bevordert. Voorafgaand aan en volgend op de training werd de DMT-2 afgenomen (kolom 1 van tabel 8). 18 van de 38 leerlingen uit groep T11 behaalden aan het einde van de trainingperiode een score op D-niveau, tegen 9 van de 39 leerlingen van groep T16. De vooruitgang van de twee groepen T11 en T16 werd getoetst in een variantieanalyse met herhaalde meting en wel/geen remedial teaching als *fixed factor*. De interactie van DMT-score x traininggroep is significant ( $F(1,75) = 6,532$ ;  $p < .01$ ). Leerlingen die de extra training ontvingen zijn meer vooruit gegaan in technische leesvaardigheid dan leerlingen die geen remedial teaching ontvingen.

Tabel 8: Herhaalde meting op DMT-2 ruwe score, aantal gelezen woorden per minuut (wpm) en percentage goed gelezen woorden (accuratesse).

Variabele:	DMT-2 ruwe score	DMT-2 wpm	DMT-2 accuratesse	
	M (SD)	M (SD)	M (SD)	
Groep T11, dl=11	7,2 (3,0)	9,5 (3,1)	69,2	(18,7)
Groep T11, dl=16	23,3 (11,1)	27,6 (12,9)	83,4	(12,1)
Groep T16, dl=11	7,2 (2,9)	10,2 (3,3)	66,1	(20,9)
Groep T16, dl=16	18,3 (6,9)	20,6 (7,5)	84,4	(10,6)

De vooruitgang op de DMT-2 is een globale maat voor het totaaleffect. Het gemiddelde verschil in ruwe score op de posttest tussen de twee groepen was 5,04. Ten opzichte van de spreiding van 6,93 op de nameting binnen de controlegroep levert dit een effectgrootte op van .73 (Kavale & Forness 1996); ten opzichte van de spreiding van 9,51 op de nameting in de experimentele en controlegroep samen is de effectgrootte

.53 (Cohen 1969). Effectgroottes van rond de .50 tot .80 worden als *moderate* beschouwd.

### lezen van tekst

De leerlingen in de T11-groep kregen instructie met teksten op hun instructieniveau. De remedial teaching periode is afgesloten met een bandopname van 15 minuten lezen door elke leerling van een tekst op het instructieniveau van iedere T11-leerling op dat moment. Daarna las ook de T16-leerling met wie hij een *matched pair* vormde hetzelfde stuk tekst. Door het uitvallen van één leerling uit de T11-groep resteerden 38 paren.

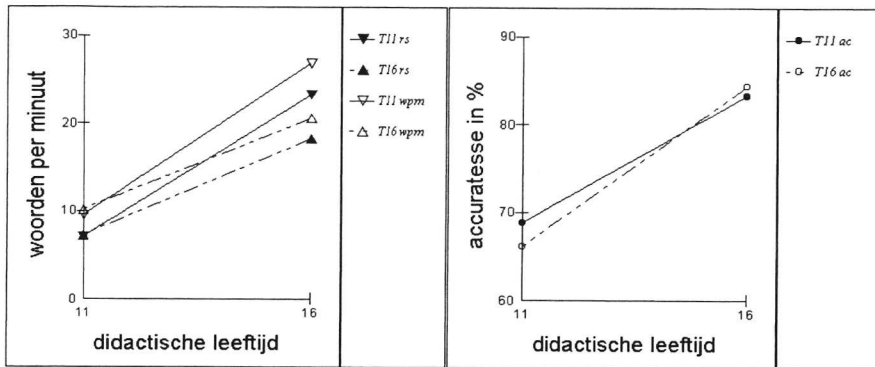
Een eerste observatie van de remedial teachers was dat de getrainde leerlingen zonder uitzondering in staat waren een half uur geconcentreerd te lezen, terwijl vele van de leerlingen uit de T16 groep grote moeite hadden om het lezen een kwartier vol te houden. Daarom zijn alleen de eerste 10 minuten van de opnamen van de T11 leerlingen geanalyseerd. Van de T16 leerlingen is vervolgens het zelfde stuk tekst genomen. Voor het lezen van dezelfde teksten hebben de T16-leerlingen gemiddeld 767 sec. (SD 422) of gemiddeld ruim 25% meer tijd nodig. Dit verschil is statistisch significant ( $t(37) = 2,44$ ;  $p < .05$ ).

De teksten die de leerlingen lezen waren voor de T11-leerlingen steeds teksten op instructieniveau. Hoe moeilijker een tekst voor een leerling is, hoe sterker met name radend lezen wordt uitgelokt, zie paragraaf 3.2. Daarom is, door een onafhankelijke beoordelaar, de accuratesse gescoord. De eerste respons op elk woord van de leerlingen is steeds beoordeeld als goed of fout, waarbij spellend lezen goed gerekend werd, mits het geheel of gedeeltelijk verklanken juist was en direct werd gevolgd door het juiste woord. De accuratessemaat is het percentage goed gelezen woorden. Gemiddeld lezen de leerlingen uit de T11-groep 93% (SD 4,1) van alle woorden de eerste keer goed, tegen 89,7% (SD 8,1) in groep T16. De spreiding in accuratesse in groep T11 half zo groot als in groep T16, waarschijnlijk als gevolg van een plafondeffect; dit verschil is statistisch significant ( $F(1,74) = 5,740$ ;  $p < .05$ ). Ook het verschil in accuratesse is, hoewel niet groot, statistisch significant onder de aanname van ongelijke variantie ( $t(54,7) = 2,26$ ;  $p < .05$ ). De ondergrens van de accuratesse op beheersingsniveau kan worden gesteld op 90%. Van de 38 leerlingen uit groep T16 lezen er 16 (42,1%) met een accuratesse lager dan 90%, in groep T11 waren dat er 6 (15,8%). Ook dit verschil was significant ( $\chi^2(1) = 6,397$ ;  $p < .01$ ). De getrainde leerlingen lezen teksten gemiddeld sneller en accurater dan de niet getrainde leerlingen.

### accuratesse en vlotheid

De ontwikkeling van de technische leesvaardigheid van de leerlingen uit de groepen T11 en T16 is ook gevolgd door de leerlingen van de DMT-2 standaard de eerste 25 woorden te laten lezen. Het percentage goed gelezen woorden gold als accuratessemaat, het aantal gelezen woorden, inclusief fouten, per minuut als tempomaat. Op deze twee maten (kolom 2 en 3 van tabel 8) is de variantieanalyse herhaald en in een grafiek uitgezet (figuur 1). Het patroon op de tempomaat is identiek en levert dezelfde signifi-

cante interactie tussen traininggroep en eindmeting op ( $F(1,75) = 12,620$ ;  $p < .001$ ). Beide groepen lieten echter geen verschil zien, noch aan het begin, noch aan het eind van de trainingperiode op accuratesse ( $F(1,75) = .012$ ; n.s.). Beide groepen zijn gedurende de trainingperiode evenveel vooruitgegaan in accuratesse van gemiddeld 68% naar gemiddeld 84%. De hogere leessnelheid bij het lezen van losse woorden van de getrainde leerlingen aan het eind van de training is in ieder geval niet ten koste gegaan van de accuratesse.



Figuur 1: De ontwikkeling van de DMT-2 ruwe score (rs), aantal woorden per minuut (wpm) en accuratesse (ac) bij de getrainde groep (T11) en de niet-getrainde groep (T16).

Aan het einde van de trainingperiode zijn accuratesse en tempo ook bepaald in het lezen van tekst. Om de aspecten accuratesse en vlotheid verder te exploreren is een principale componenten analyse met varimax rotatie uitgevoerd voor de groepen T11 en T16 afzonderlijk (tabel 9). In controlegroep T16 zijn volgens het gangbare criterium van Eigenwaarde  $> 1$  twee factoren te onderscheiden die samen 69,5% van de variantie verklaren. De technische leesvaardigheid zoals standaard gemeten met de DMT van zwakke lezers met een didactische leeftijd van 11 maanden, hangt samen met de verschillende accuratessematen, zowel na 11 als na 16 maanden leesonderwijs en zowel bij het lezen van losse woorden als bij het lezen van tekst.

De technische leesvaardigheid op een didactische leeftijd van 16 maanden hangt daarentegen sterk samen met de verschillende tempomaten na 11 en na 16 maanden leesonderwijs en zowel bij losse woorden als bij tekst. Het betreft hier de 10% zwakste lezers die geen gerichte training ontvingen gedurende de eerste helft van groep 4. Mogelijk is de eerste helft van groep 4 voor E-leerlingen de periode waarin de kwaliteit van de leesteknik verandert van primair een zaak van accuratesse naar primair vlotheid. De leesvaardigheid van E-leerlingen aan het begin en het eind van die periode is te vergelijken met de leesvaardigheid van gemiddelde leerlingen tussen ongeveer vijf en acht maanden leesonderwijs.

In groep T11 verklaren twee factoren een evengroot deel van de totale variantie (69,8%). Groep T11 ontving in de onderzoeksperiode intensieve remedial teaching

gericht op verhoging van de vlotheid van woordherkenning, waarbij al is vastgesteld dat dit niet ten koste in gegaan van de accuratesse. Het patroon van samenhang in groep T11 wijkt af van dat in controlegroep T16. De verschillende maten hangen op de twee meetmomenten onderling sterk samen, terwijl er nauwelijks samenhang bestaat tussen de resultaten op de verschillende meetmomenten.

Tabel 9: Gemiddelden en spreiding, correlaties en principale componentenanalyse met varimax-rotatie van ruwe score (rs), tempo (wpm) en accuratesse (accur.) bij het lezen van losse woorden (DMT-2) en tekst, op didactische leeftijden 11 en 16 maanden van de groep getrainde leerlingen (T11) en de controlegroep (T16) afzonderlijk.

Groep:	T11: wel RT (N=38)								T16: geen RT (N=38)							
	M (SD)								M (SD)							
Dmt2-rs, dl11	7,03	(3,00)							7,37	(2,93)						
Dmt2-rs, dl16	23,32	(11,11)							18,63	(7,13)						
Dmt2-wpm, dl11	9,46	(3,14)							10,32	(3,23)						
Dmt2-wpm, dl16	27,63	(12,99)							20,90	(7,36)						
Tekst-wpm, dl16	32,82	(13,33)							29,64	(12,49)						
Dmt2-accur., dl11	69,16	(18,67)							66,42	(21,01)						
Dmt2-accur., dl16	83,37	(12,07)							84,74	(10,56)						
Tekst-accur., dl16	93,05	(4,09)							89,68	(8,19)						
<b>Correlatiematrix</b>	2	3	4	5	6	7	8	2	3	4	5	6	7	8		
1.Dmt2-rs, dl11	.10	.68***	.12	.20	.71***	.15	.20	.52***	.49**	.45**	.51**	.57***	.46**	.53***		
2.Dmt2-rs, dl16	--	.45**	.96***	.76***	.02	.62***	.39**	--	.39**	.87***	.65***	.21	.60***	.38**		
3.Dmt2-wpm, dl11		--	.45**	.34*	.32*	.31*	.15		--	.47**	.48**	-.11	.20	.47**		
4.Dmt2-wpm, dl16			--	.74***	.03	.43**	.33*			--	.61***	.07	.34*	.32*		
5.Tekst-wpm, dl16				--	.22	.40**	.58**				--	.17	.53***	.52***		
6.Dmt2-accur., dl11					--	.15	.39**					--	.55***	.45**		
7.Dmt2-accur., dl16						--	.44**						--	.43**		
8.Tekst-accur., dl16							--							--		
<b>Componentmatrix</b>	Component				Component				Component				Component			
	1		2		1		2		1		2		1		2	
Dmt2-rs, dl11	.06		.94		.46		.68		.46		.68		.46		.68	
Dmt2-rs, dl16	.97		.00		.80		.35		.80		.35		.80		.35	
Dmt2-wpm, dl11	.41		.64		.76		.02		.76		.02		.76		.02	
Dmt2-wpm, dl16	.92		.02		.87		.13		.87		.13		.87		.13	
Tekst-wpm, dl16	.84		.17		.75		.36		.75		.36		.75		.36	
Dmt2-accur., dl11	.02		.87		-.17		.94		-.17		.94		-.17		.94	
Dmt2-accur., dl16	.67		.15		.31		.74		.31		.74		.31		.74	
Tekst-accur., dl16	.54		.31		.41		.61		.41		.61		.41		.61	

\*\*\* p < .001 \*\* p < .01 \* p < .05

Groep T16 ontving in de periode september tot februari *care as usual*. In deze groep vinden we het gebruikelijke sterke verband tussen pre- en posttest ( $r = .52$ ;  $p < .001$ ): hoe beter een leerling in september las, hoe beter hij ook in februari leest, in overeenstemming met het Mattheuseffect (ook Gollob & Reichardt, 1987). In de getrainde groep is dit verband er niet ( $r = .10$ , n.s.).

Er lijkt ook een kwalitatief verschil te zijn ontstaan tussen de wel en niet-getrainde leerlingen en de werking van het Mattheuseffect lijkt in de getrainde groep met enig succes tegengegaan

### benoemsnelheid

Op de vier gangbare RAN-taken is de variantieanalyse met herhaalde meting herhaald zoals ook uitgevoerd op de DMT- en de daarvan afgeleide scores, weergegeven in tabel 10. De toename in benoemsnelheid op elk van de vier taken is significant (letters  $F(1,75) = 169,99$ ; cijfers  $F(1,75) = 186,43$ ; plaatjes  $F(1,75) = 86,43$ ;  $p < .001$ ; kleuren  $F(1,75) = 7,12$ ;  $p .009$ ), maar op geen van de vier maten is een interactie-effect aantoonbaar tussen voor- en nameting en wel of geen training.

Tabel 10: Herhaalde meting op vier RAN-taken in aantal benoemde eenheden per minuut.

	Variabele: letters per min. cijfers per min. plaatjes per min. kleuren per min.			
	M (SD)	M (SD)	M (SD)	M (SD)
Groep T11, dl=11	38,6 (10,2)	64,3 (16,7)	31,2 (6,7)	30,7 (7,0)
Groep T11, dl=16	59,7 (16,1)	85,5 (18,0)	37,0 (6,0)	37,2 (7,8)
Groep T16, dl=11	39,7 (9,3)	55,6 (13,0)	28,9 (6,1)	28,6 (6,5)
Groep T16, dl=16	60,3 (13,4)	82,3 (15,9)	35,8 (6,4)	36,0 (6,6)

Ook is gekeken naar de samenhang tussen de RAN-taken onderling en met de leesvaardigheid. De vier RAN-taken blijken onderling sterk samen te hangen, een principale componentenanalyse levert één centrale factor op die 67% van de variantie verklaart, maar van samenhang met de technische leesvaardigheid zoals gemeten met de DMT-2 is op het moment van de voormeting geen sprake (correlaties variërend van .02 tot .12). Evenmin hebben de prestaties op de RAN-taken van de leerlingen uit de controlegroep voorspellende waarde voor de leesvaardigheid na een half jaar aan het einde van de onderzoeksperiode. RAN-taken voorspellen ook niet welke leerlingen uit de experimentele groep meer of minder van de training profiteren.

### follow up

Een jaar na afronding van het onderzoek, de leerlingen hadden inmiddels een didactische leeftijd van 29 maanden, is de scholen over de 120 leerlingen die tijdens het r-onderzoek zijn gevolgd, een aantal vragen voorgelegd. Hierin werden de meest recente toetsgegevens uit het leerlingvolgsysteem opgevraagd en voorts een korte beschrijving van eventuele extra maatregelen die het afgelopen schooljaar nog genomen zijn. Op grond van deze informatie is het schooloordeel vastgelegd over de vraag of de betreffende leerling aan het eind van zijn vijfde leerjaar nog tot de uitvallers gerekend moest worden wat betreft technische leesvaardigheid. Toetsgegevens zijn beoordeeld aan de hand van de normen voor eind groep 5, ook al was er sprake van zitten blijven.

De groepen T11 en T16 en een deel van groep C3 ontvingen remedial teaching. De groep C4 was de quasi-controlegroep, de leerlingen die op de deelnemende scholen net

iets beter scoorden dan de geselecteerde leerlingen. Acht leerlingen vormden subgroep C3. Zij waren de leerlingen van twee deelnemende scholen die in groep 3 doubleerden. Zij voldeden aan de criteria voor deelname aan het rt-onderzoek, maar de schoolleiding van deze scholen wenste ze van deelname in het eerste onderzoek uit te sluiten, met als argument dat zij door het doubleren reeds veel extra leesinstructie zouden krijgen. Voorts is de scholen gevraagd welke leerlingen uit de groep met een didactische leeftijd van 29 op dat moment tot de uitvallers gerekend zouden moeten worden, die bij het begin van het onderzoek, DL 11, niet als zodanig zijn gesignaleerd. De resultaten zijn samengevat in tabel 11.

Tabel 11: Follow up gegevens, aantallen uitvallers op technisch lezen eind groep 5.

Subgroep	Leesniveau op DL=11	DL=29		Totaal
		E	D	
T11+T16	E	27	44	71
C3	E	7	1	8
C4	E	1	13	14
	D	2	15	17
C5	D	1		1
TOTAAL		38	73	111

In totaal waren van 111 leerlingen follow up gegevens te achterhalen. Twee leerlingen waren inmiddels verwezen naar een school voor speciaal basisonderwijs. Deze twee leerlingen zijn in het overzicht opgenomen als uitvaller op lezen (E), ook al ontbrak de informatie over het leesniveau op dat moment. Negen leerlingen zaten niet meer op dezelfde school wegens verhuizing.

De scholen beoordelen een jaar later minder leerlingen als uitvaller (38) dan de toetsing aan het einde van de onderzoeksperiode aangaf (55). De sekseratio in de groep uitvallers is onveranderd 2:1. Slechts één leerling, aangeduid als C5, die bij het begin van het onderzoek niet als uitvaller is gesignaleerd, wordt een jaar later door de school als uitvaller aangemerkt. Dit lijkt onwaarschijnlijk laag, maar ook van de zwakste lezers die net niet tot de uitvallers gerekend werden bij het begin van het onderzoek, de D-lezers uit subgroep C4, behoren eind groep 5 slechts twee van de vijftien leerlingen tot de uitvallers. Maar toch, ter relativering van deze gegevens, wordt verwezen naar het onderzoek van De Jong & Van der Leij (2003) waaraan in hoofdstuk 2 reeds is gerefereerd, die een minder rooskleurig beeld van de stabiliteit van de leesscores van zwakke lezers. Hierbij moet wel weer worden opgemerkt, dat tussen eind groep 3 en begin groep 4 een aantal cruciale onderwijsmanden kunnen zitten.



### 3.5 Discussie

#### **implementatie van de methodiek**

De gerealiseerde remedial teaching voorzag in een sterke uitbreiding van de instructietijd rond technisch lezen. De werkwijze was expliciet gebaseerd op de toegepaste gedragsleer, evenals de manier waarop deze is geïmplementeerd. De samenwerking met de remedial teachers heeft geleid tot differentiatie binnen de instructie van het *wacht, hint, prijs* protocol in antecedente en consequentie technieken. Door gebruik te maken van verschillende implementatiebevorderende maatregelen is een hoge behandelingsintegriteit gerealiseerd: gedetailleerde handelingsvoorschriften, directe instructie tijdens cursusbijeenkomsten, observatie van de uitvoering, systematische feedback op basis van de uitgevoerde observaties en de bijgehouden logboeken. De ervaring met twee remedial teachers die wel training gaven, maar, overigens om verschillende redenen, niet aan de cursus deelnamen, suggereren dat het bereiken van voldoende behandelingsintegriteit niet lukt op basis van instructie uit de tweede hand.

De zorg van Lyon & Moats (1997) omtrent controle voor de leergeschiedenis van de leerlingen, lijkt mij begrepen te moeten worden vanuit de Amerikaanse onderwijscontext van de meeste trainingstudies en de gemiddeld hoge leeftijd van proefpersonen. In Nederland krijgen - vrijwel - alle kinderen kleuteronderwijs en de gebruikelijke methoden voor aanvankelijk lezen verschillen niet wezenlijk van elkaar. Als belangrijkste aandachtspunt bij interventieonderzoek noemen Lyon & Moats heterogeniteit qua leesniveau bij aanvang, omdat dit een rol zou kunnen spelen in de afhankelijkheid van leerlingen van bepaalde typen instructie en de intensiteit en duur van de instructie. De leerlingen in ons onderzoek vormden in die zin een homogene groep. Ze zijn allen voor deelname geselecteerd na tien maanden leesonderwijs en behoorden volgens de landelijke norm tot de 10% zwakste lezers. Voor andersoortige verschillen tussen individuele scholen is door de samenstelling van experimentele en controlegroep gecontroleerd.

De interventie duurde een half (school)jaar en bestond in totaal uit 30 uur training, remedial teaching plus computergestuurde woordtraining. Dit is lang ten opzichte van de meeste trainingstudies, maar er zijn voorbeelden van *school based* programma's die een veelvoud hiervan realiseerden (Vellutino e.a. 1998, Torgesen e.a. 2001). Als het waar is dat de 10% zwakste lezers vier tot zesmaal zoveel instructietijd nodig hebben als de 10% beste lezers (Gettinger 1984), dan waren de gerealiseerde 30 uur slechts een aanzet. Wel is maximaal gepoogd generalisatie als onderdeel van de interventie op te nemen, door twee belangrijke aspecten van technisch lezen, het oefenen met woorden en de toepassing daarvan bij het lezen van teksten, naast elkaar en in dezelfde intensiteit aan bod te laten komen.

Met een cursus, het bijhouden van een logboek, observaties en coaching is de controle op leerkrachtkenmerken en consistentie in de uitvoering van de methode, ofwel de behandelingsintegriteit, zo goed mogelijk gegarandeerd. Het is een gegeven dat ook de meest succesvolle implementaties gewoonlijk nog een aantal jaren ondersteuning

vragen alvorens sprake is van volledige incorporatie (Slavenburg 1997). Voor onderhoud na afloop van de interventie was in dit onderzoek helaas geen mogelijkheid.

### **effectiviteit**

Het probleem van te eenzijdige effectmeting is ondervangen doordat Nederlandstalige tests voor technische leesvaardigheid altijd een combinatiemaat zijn voor accuratesse en vlotheid. Daarbij is de accuratesse ook apart geanalyseerd. Bovendien is tekstlezen toegevoegd en is er een follow up meting geweest.

De accuratesse bij het lezen van woorden is aan het einde van de trainingperiode hoog en wordt door de training niet beïnvloed. Het lijkt zo te zijn dat, in de Nederlandstalige context, voor de zwakste lezers in de eerste helft van groep 4 vlotheid het meest bepalende aspect van de technische leesvaardigheid is.

Er kon in dit onderzoek, bij deze jonge, zwakke lezers geen samenhang worden aangetoond tussen de standaardmaten voor *rapid automatized naming* en de technische leesvaardigheid bij aanvang van de trainingperiode. Ook blijkt er geen voorspellende waarde voor het leesniveau na afloop van de trainingperiode. De benoemsnelheid op alle RAN-maten, waaronder het benoemen van letters, ontwikkelt zich gedurende de trainingperiode wel verder, maar training gericht op het snel lezen van woorden heeft geen effect op benoemsnelheid van overbekende stimuli buiten het leesdomein. Een verschil met onderzoek waarin de rol van benoemsnelheid sterker en in meer algemene zin werd aangetoond (Wolf e.a. 2000, Bowers 2001), is dat dit betrekking had op oudere leeszwakke of dyslectische leerlingen of op jonge normale lezers. Daarin was met andere woorden sprake van een gemiddeld aanzienlijk hoger leesniveau en van veel grotere onderlinge verschillen in leesvaardigheid of er is zelfs geen sprake van een selectie van proefpersonen op basis van leesvaardigheid (Van den Bos 2000). Dat zou kunnen betekenen dat *rapid automatized naming* wel betekenis heeft voor het onderscheid tussen wel of geen leesproblemen, maar niet binnen de groep leerlingen met leesproblemen.

De leesvaardigheid van de 10% zwakste lezers begin groep 4 lijkt door intensieve individuele remedial teaching positief beïnvloed te kunnen worden. Losse woorden worden sneller herkend zonder dat dit ten koste gaat van de accuratesse, teksten worden sneller en accurater gelezen. Follow-up gegevens lijken in de richting te wijzen dat in deze vroege fase van de ontwikkeling van de technische leesvaardigheid leerlingen met leesproblemen betrouwbaar geïdentificeerd kunnen worden. De werking van het Mattheuseffect lijkt met vroegtijdige intensieve remedial teaching met enig succes te kunnen worden tegengegaan.

Bij 38 van de 93 gevolgde E-leerlingen blijkt de leesuitval hardnekkig. Deze hardnekkigheid is een sterke aanwijzing dat bij deze leerlingen sprake is van dyslexie. Er zijn verschillende onzekere factoren die het lastig maken om op basis van deze uitkomst een betrouwbare schatting te maken omtrent de incidentie van dyslexie. Sommige leiden tot een neerwaartse, andere tot een opwaartse bijstelling. Leerlingen van wie bekend was dat ze de school zouden verlaten wegens verhuizing of verwijzing naar

speciaal basisonderwijs, zijn op voorhand van deelname uitgesloten. Ook konden voor de follow up tien leerlingen niet meer worden achterhaald. Het percentage uitvallers aan het begin van het traject lag op de geselecteerde scholen met 14,6% significant hoger dan de verwachte 10%. Daardoor kregen niet alle E-leerlingen de interventie, maar gemiddeld slechts het zwakste tweederde deel (gemiddelde percentielscore .03 in plaats van .05). De getrainde groep had derhalve een gemiddeld zwakkere technische leesvaardigheid dan de gemiddelde E-leerling. Als aan dit laatste punt relatief meer gewicht mag worden toegekend, dan lijkt, alles bij elkaar genomen, deze vorm van interventie effectief voor ruwweg twee van de drie uitvallers.