



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>)

4. De woordtraining

4.1 Focus op vlotheid

De onderzoeksvraag van het tweede deel van deze studie was of de mate waarin en de wijze waarop individuele leerlingen op de interventie reageerden, de uitkomst ervan voorspelt. In geval leerlingen onvoldoende van de interventie profiteren, zou dit een sterke aanwijzing voor dyslexie inhouden. In de inrichting van het traject zijn uit literatuur en praktijk bekende condities voorzover mogelijk meegenomen, waaronder het systematisch verzwaren van de taak, een werkwijze die past in het automatiserings-paradigma.

De onderzoeksliteratuur op het gebied van lezen, leesproblemen en dyslexie wordt gedomineerd door onderzoek uit de Verenigde Staten en het Verenigd Koninkrijk, dat gaat over Engelssprekende kinderen, die Engels moeten leren lezen. Hierin wordt in de regel niet stilgestaan bij de typische kenmerken van het Engels en van het Engelse en Amerikaanse leesonderwijs. Verder betreft het onderzoek uiteenlopende aspecten van de leesvaardigheid, die op verschillende manieren worden gedefinieerd en geoperationaliseerd. Binnen deze variatie lijkt één conclusie over oorzaken van leesproblemen boven iedere twijfel verheven: de hypothese van het fonologisch tekort.

In 2000 verscheen *Teaching Children to Read*, het rapport van het Amerikaanse National Reading Panel, met de betekenisvolle ondertitel: *an evidence based assessment of the scientific literature on reading and its implications for reading instruction*. Het Panel rapporteert sterke empirische evidentie voor het zogenaamde *phonological deficit* bij zwakke lezers en voor de zinvolheid om leerlingen expliciete training te geven in *alphabets*, waaronder *phonemic awareness* en *phonics* worden begrepen. *Phonemic awareness* is gedefinieerd als *the ability to focus and manipulate phonemes in spoken words*. *Phonics* betreft *acquisition of letter-sound correspondences and their use to read and spell words*. Het voorgaande zou een samenvatting kunnen zijn van Grossen's (1997) aanbevelingen voor effectieve leesinstructie:

- begin op jonge leeftijd al meteen met onderwijs in fonemisch bewustzijn;
- leer elke klank-tekenkoppeling expliciet en systematisch aan;
- doe precies voor hoe je woorden kunt lezen door het verklanken van letters en auditieve synthese;
- gebruik *decodable texts* om het spellende lezen te oefenen;
- maak gebruik van goede verhalen om taalbegrip te bevorderen;
- wissel het gebruik van goede verhalen en instructie in decodeervaardigheden af, maar vermeng ze niet.

In Nederland wordt, in tegenstelling tot de Angelsaksische landen, een nogal strikt onderscheid gemaakt tussen aanvankelijk en voortgezet technisch lezen. Eind jaren zeventig en begin jaren tachtig beleefden we nog een heuse methodestrijd, maar deze is inmiddels opgelost. Moderne leesmethoden verschillen niet meer wezenlijk van elkaar

in opzet en effectiviteit (Groeneveld e.a. 1992, Hol e.a. 1995, Harskamp e.a. 1995). Vrijwel alle basisscholen in Nederland maken sinds jaar en dag gebruik van dergelijke effectief bevonden methoden. Alle gangbare methoden voor aanvankelijk lezen hebben tegenwoordig de kenmerken die zowel Grossen als het Amerikaanse National Reading Panel aanbevelen. Ze zijn, in termen van Spear & Sternberg (1994), sterk gericht op de *phonetic cue reading* en schrijven expliciete oefening voor in de deelvaardigheden auditieve synthese en klank-tekenkoppeling en directe instructie in de toepassing daarvan, voor het overgrote deel met zogenaamde klankzuivere woorden, zodat er in die beginfase sprake is van een vrijwel volmaakt regelmatige orthografie. Met andere woorden, in alle moderne aanvankelijk leesmethoden komen zowel *phonemic awareness* als *phonics* uitgebreid aan de orde met daarbij aanwijzingen voor differentiatie in instructie en hoeveelheid leertijd die aan de verschillende aspecten besteed wordt en de mate waarin de instructie expliciet is, afhankelijk van de instructiebehoeften van de leerling.

De aanbevelingen van het National Reading Panel en Grossen komen, vertaald naar de Nederlandse context, in hun algemeenheid neer op onze praktijk van het aanvankelijk lezen, maar zijn voor de Engelstalige context nadrukkelijk niet tot de beginfase van het leren lezen beperkt. Op grond daarvan zou men al kunnen vermoeden dat deze Angelsaksische conclusies als remediërende aanpak voor oudere leerlingen met een achterstand in technisch leesvaardigheid in ons taalgebied niet zonder meer zijn over te nemen. De enkele keer dat dat toch gedaan is, leverde dat dan ook zeker geen resultaten op die vergelijkbaar zijn met wat in Engels en Amerikaans onderzoek gerapporteerd wordt, zie bijvoorbeeld Van den Broeck (1997) die een *phonics*-aanpak als controle conditie gebruikte, of Wesseling (1999). Beiden werkten met leerlingen die een technische leesvaardigheid hadden vergelijkbaar met het gemiddelde niveau eind groep 3. Expliciete instructie in *phonics* kan voor het Nederlands klaarblijkelijk grotendeels beperkt blijven tot de fase van het aanvankelijk lezen, omdat de fonologische vaardigheden dan voldoende ontwikkeld zijn, gegeven de relatief beperkte eisen die de Nederlandse orthografie daaraan stelt. Alleen door het uit het oog verliezen van de Engelstalige onderwijscontext van Engels en Amerikaans onderzoek kan men binnen de Nederlandstalige context besluiten tot het kopiëren van een *phonics* training voor lezers die het aanvankelijke niveau gepasseerd zijn (Struiksmā 1998).

Maar in Engels en Amerikaans onderzoek is er wel degelijk evidentie voor de effectiviteit van *phonics*-instructie, ook voor oudere leerlingen met een leesachterstand en voor behandeling van dyslexie. De effectiviteit is aangetoond voor groepen van uiteenlopende leeftijden en leesniveaus. Swanson e.a. (1999) voerden een meta-analyse uit op behandelingsresultaten van interventiestudies. 55 studies waren gebaseerd op groepsvergelijking met als onderwerp *word recognition*. De typische - voor het overgrote deel Amerikaanse - interventiestudie op dat gebied had de volgende kenmerken (standaard deviatie tussen haakjes). De proefpersonen waren 10,0 (SD 1,3) jaar. Hun leesvaardigheidsquotiënt was 80,0 (SD 11,5); dit komt overeen met een gemiddelde score op het 10^{de} percentiel; ruwweg tweederde van de studies betrof leerlingen met

een gemiddelde leesscore tussen het 2^{de} en het 25^{ste} percentiel. De interventie bestond uit 38,6 (SD 42,1) sessies van 39 (SD 29) minuten in een frequentie van 3,5 (SD 1,3) per week. De effectgrootte was .71 (SD .56). De volgende kenmerken bleken van invloed op de gerapporteerde effectgroottes; het eerst genoemde kenmerk resulteerde in een groter effect in vergelijking tot het tweede genoemde kenmerk:

- gebruik van experimentele maten versus gestandaardiseerde instrumenten als afhankelijke variabele;
- steekproefgrootte kleiner dan 25 of groter dan 100 versus tussen 25 en 100;
- geringe achterstand versus grote achterstand;
- klein aantal behandelsessies versus een hoog aantal;
- bottom-up directe instructie van decoderen versus top-down instructie van strategieën;
- verschil in *instructional focus* tussen behandeling en aanpak van de controlegroep; de behandelingsintegriteit werd overigens in slechts 23% van de studies gecontroleerd.

De grootste effecten werden gevonden als de interventie gebaseerd was op directe instructie en als tevens de *instructional focus* van de interventie meer afweek van die in de controlegroep. Aangezien in de Verenigde Staten *whole language* nog altijd de standaard is (Pressley 1998), worden de grootste effecten bereikt met directe instructie in *phonics*. Volgens Foorman e.a. (1997) geldt in het algemeen zelfs hoe meer *phonics* hoe beter, want meer *phonics* leidt tot snellere groei van zowel accuratesse als snelheid, zowel bij woorden met een regelmatige als een onregelmatige spelling.

Blok e.a. (2002) deden een meta-analyse op trainingstudies waarbij gebruik gemaakt werd van de computer. De trainingduur varieerde van 1 tot 32 uur met een gemiddelde van 6 uur, verdeeld over 1 tot 75 weken, met een gemiddelde van 11 weken. De programma's waren gericht op fonologisch bewustzijn, op het lezen van woorden met spraakfeedback, op flitsaanbieding of waren multicomponentieel. De afhankelijke variabele was meestal een accuratessemaat voor het lezen van woorden. Vooral Nederlandse studies vormden hierop een uitzondering door - ook - de leessnelheid als afhankelijke variabele te nemen. Blok e.a. rapporteren bescheiden gemiddelde effectgroottes van .37 (SD .09). Een opmerkelijk resultaat dat de auteurs vermelden, maar schrijven niet goed te kunnen interpreteren, betreft wat zij noemen een regio-effect. Europese studies hadden gemiddeld .19 geringere effectgroottes in vergelijking tot studies van Amerikaanse en - Engelstalige - Canadese origine. Van de 44 geanalyseerde studies waren er 31 Europees, maar daarvan slechts twee Brits, drie Scandinavisch en de overige 26 Nederlands. De regioverdeling Europees/Noord-Amerikaans viel dus vrijwel samen met de verdeling Engelstalig/Nederlandstalig. Bovendien blijkt tweederde van de Engelstalige studies betrekking te hebben op *regular/at risk* leerlingen, terwijl de niet-Engelstalige studies juist voor tweederde gaan over *dyslectic/learning disabled* leerlingen. Van der Leij (1994) vond met betrekking tot dezelfde Nederlandse studies dat multicomponentiële programma's niet effectief waren.

Het voorgaande lijkt te illustreren dat in Amerikaanse studies, in vergelijking tot Nederlandse, doorgaans sprake is van minder extreme populaties, die leesonderwijs ontvingen met weinig systematische *phonics*-instructie. Het lijkt niet onwaarschijnlijk de verklaring voor verschillen in effectiviteit van op *phonics* gerichte training te zoeken in de richting van systematische verschillen tussen onderzoekspopulaties,

zowel qua leeftijd als leesniveau, in combinatie met verschillen tussen de vigerende onderwijspraktijken.

Ook het gebruik in onderzoek van kwalitatief zeer verschillende maten voor leesvaardigheid maakt internationale vergelijking lastig, zowel met betrekking tot prevalentiecijfers, als op het punt van effectgroottes bij interventiestudies. Zo wordt in de Verenigde Staten en in het Verenigd Koninkrijk het leesniveau vrijwel zonder uitzondering uitsluitend vastgesteld met behulp van zogenaamde *graded tests*. Hierbij gaat het slechts om de accuratesse van het lezen van losse woorden of teksten die opklimmen in technische moeilijkheidsgraad (Wagner & Barker 1994). Rasinski (2000) vond dat zwakke lezers teksten op hun *grade level* lezen in een tempo dat slechts 60% was van dat van gemiddelde lezers van hetzelfde *grade level*. In Nederland is het onmogelijk om tot een dergelijke bevinding te komen, vanwege de wijze waarop het leesniveau bij ons wordt vastgesteld. Ook als dat alleen gebeurt in de vorm van een groeps- of didactischeleeftijdequivalent met behulp van *graded* leesteksten, zoals de AVI-kaarten, telt daarin naast accuratesse ook het leestempo mee voor het al of niet halen van de kaart. Een Nederlandse leerling met leesproblemen leest dus *per definitie* niet alleen even accuraat, maar ook even snel als een gemiddelde leerling met hetzelfde leesniveau. Maar meestal wordt het leesniveau bepaald door leerlingen een woordenlijst te laten lezen waarbij er niet of nauwelijks sprake is van een opklimmende moeilijkheidsgraad van de woorden. De score is het aantal goed gelezen woorden in een minuut en dit is in feite een zuivere tempomaat, omdat, met uitzondering van specifieke klinische groepen, de accuratesse altijd hoog is. Omdat de moeilijkheidsgraad van woorden in het Engels zeer sterk uiteenloopt en er zelfs op elk leesvaardigheidsniveau ook voor normale lezers nog moeilijke woorden bestaan, zou een 'Engelstalige EMT' in principe wel te maken zijn, maar deze zou binnen de context van het Engels weer veel minder relevant zijn dan bij ons.

Wimmer e.a. (1993, 1996, 2000) onderzoeken de kenmerken van leerlingen met leesproblemen juist nadrukkelijk tegen de achtergrond van enerzijds de gebruikelijke leesdidactiek en anderzijds de kenmerken van de orthografie. Het leesonderwijs in Oostenrijk en Duitsland wordt door Wimmer getypeerd als klanksynthese didactiek; nieuwe klank-tekenkoppelingen worden steeds maar met enkele tegelijk geïntroduceerd en woordherkenning via de weg van auditieve synthese wordt expliciet aangeleerd en geoefend. De Duitse orthografie is voor het lezen, dat wil zeggen wat betreft de consistentie van de koppelingen van grafeem naar foneem, nog eenduidiger dan de Nederlandse. Duitstalige leerlingen met leesproblemen hebben geen problemen met accuratesse, enerzijds omdat de decodeervaardigheden systematisch en expliciet aangeleerd worden en anderzijds omdat de eisen die aan de fonologische vaardigheden gesteld worden veel lager zijn dan in het Engels door de regelmatigheid van de orthografie van het Duits. Duitse leerlingen met leesproblemen lezen even accuraat als, maar veel langzamer dan hun leeftijdgenoten.

Het feit dat in Amerikaans en Engels onderzoek - soms aanzienlijk - oudere leerlingen met een technische leesvaardigheid ruim boven eind groep 3 niveau, nog zoveel

baat hebben bij een systematische *phonics*-aanpak, is volgens Foorman e.a. (1997) *a consequence of instruction*. De leesdidactiek in het Engelstalige onderwijs is *primarily literature based and guided by a whole-language philosophy, with phonics being taught on an as-needed basis rather than systematically* (Torgesen e.a. 1999). Daarbij is de orthografie van het Engels zeer onregelmatig en zijn de eisen die aan fonologische vaardigheden gesteld worden hoog. Aangezien het standaard onderwijs hieraan maar weinig systematische aandacht besteedt, terwijl de directe woordherkenning al vanaf de vroegste fase van het leren lezen - dat wil zeggen vanaf de kleuterleeftijd - centraal staat, is het niet vreemd dat in de Engelstalige context *phonics*-instructie een sleutel tot succes lijkt.

Torgesen (2000) schreef een overzichtsartikel over *individual differences in response to early intervention in reading* van zes preventiestudies uitgevoerd tussen 1990 en 1999. Het betroffen grootschalige, intensieve en langdurige interventies, met verschillende startpunt, maar altijd doorlopend tot eind klas 2, bij geselecteerde zwakke lezers. In alle gevallen betrof de interventie expliciete training van fonologische en decodeervaardigheden. Zijn conclusie luidde dat dit type interventie het aantal uitvallers sterk kan doen verminderen. Maar ook dat er gemiddeld 4-6% *treatment resisters* zijn, terwijl er van de leerlingen die met succes een dergelijk programma doorliepen, bij selectieratio's variërend van 12 tot 18%, ten minste eenderde in latere jaren toch onvoldoende leesvorderingen laat zien. En dat terwijl in de besproken interventiestudies de gebruikte leestoetsen slechts accuratesse maten.

Torgesen e.a. (2001) doen verslag van een succesvol, zeer intensief remediërend programma. Acht- tot tienjarige leerlingen met ernstige leesproblemen ontvingen 80 uur individuele instructie gericht op *phonological awareness* en *phonemic reading skills*. De leerlingen gingen gemiddeld sterk vooruit in accuratesse, maar bleven *substantially below average in fluency*. Lyon & Moats (1997) veronderstellen een drempelwaardemodel voor de relatie tussen leesvaardigheid en de deelvaardigheden auditieve synthese en klank-tekenkoppeling, als zij zich afvragen *is there a point when continued instruction in sound structure and sound-symbol relationships is no longer productive or necessary ... in order for fluent text reading and reading comprehension to reach robust levels?* Tunmer & Nesdale (1985) en Juel e.a. (1986) toonden dit aan voor *phonological sensitivity* in relatie tot *efficient word recognition*. Het onderzoek van Wesseling (1999) lijkt dit ook voor de Nederlandse situatie te ondersteunen. Bij beginnende lezers is er na drie maanden leesonderwijs een sterke relatie tussen het lezen van woorden, letterverklanking en auditieve synthese; bij oudere leerlingen, gemiddeld 8;9 jaar, uit het speciaal basisonderwijs met een ernstige leesachterstand, hun leesniveau is vergelijkbaar met ongeveer halverwege groep 3, blijkt er geen verband (meer) te zijn.

Mayringer e.a. (1998) vonden in Oostenrijk een veel lagere voorspellende waarde van fonologische vaardigheden op toekomstige leesvaardigheid dan gemiddeld in Amerikaans en Engels onderzoek gerapporteerd wordt. Maar ook in Amerikaanse trainingstudies worden slechts geringe effecten gevonden van verbetering bij jonge leer-

lingen van fonologische vaardigheden in het algemeen op de latere leesvaardigheid (Bus & Van IJzendoorn (1999). Duitstalige leerlingen met fonologische problemen ondervinden wel belemmeringen bij het *leren* lezen in klas 1, maar hebben daar in bijna alle gevallen in klas 3 geen last meer van (Wimmer e.a. 2000). In Nederland vonden De Jong & Van der Leij (1999) dat de consequenties van fonologische problemen beperkt bleven tot het leren lezen in groep 3.

Wimmer (1993) onderscheidde *decoding* en *speed dyslexia*. In Engelssprekende landen is men, bij wijze van spreken, aan *speed dyslexia* nog nauwelijks toegekomen en in de Verenigde Staten zal men daar pas echt kennis mee maken, nadat de aanbevelingen van Grossen en het National Reading Panel breed zijn geïmplementeerd. De slag die het National Reading Panel in Amerika hoopt te maken, is in Duitsland en Oostenrijk misschien wel nooit verloren en kan in Nederland inmiddels binnengehaald zijn, terwijl de *decoding dyslexia*, vanwege de hogere taakeisen, voor Engelstaligen altijd een groter probleem zal blijven. Waarschijnlijk is daarom van een verdere verbetering van de decodeervaardigheden in het aanvankelijk lezen voor potentiële uitvallers in Nederland niet zo veel winst (meer) te verwachten, omdat onvoldoende *fluency* - zeker in de niet-Engelstalige context - het meest bepalende kenmerk is van pervasieve leesproblematiek.

Toch is ook bij Engelstalige onderzoekers belangstelling voor het tweede aspect van de leesteknik, de *fluency*. Wolf & Katzir-Cohen (2001) zijn van mening dat er drie redenen zijn voor een hernieuwde belangstelling voor *fluency* en op *fluency* gerichte interventies. Ten eerste heeft systematisch onderzoek naar op fonologische processen bij leesproblemen gebaseerde interventies aangetoond dat deze zowel zeer succesvol als onvoldoende zijn. Voorts wijst onderzoek op een scala van *time-related deficits* die de ontwikkeling van *fluency* kunnen belemmeren. Ten slotte zou er, gerelateerd aan het vorige punt, steeds meer bewijs zijn dat een deel van de kinderen met leesproblemen gekenmerkt kan worden door verschillende tekorten. Maar ook al was het deelrapport van het National Reading Panel over *fluency* qua omvang slechts een kwart van dat over *alphabets* en was het aantal voor de meta-analyse geschikt geachte bronnen minder dan de helft, de conclusie dat pas recent onderzoek het belang van vlotheid onder de aandacht heeft gebracht, lijkt niet gerechtvaardigd.

Wolf & Katzir-Cohen beginnen hun artikel met een verwijzing naar Huey (1908), die zowel het belang van een steeds snellere uitvoering van leestaken beschreven heeft, als het essentiële kenmerk van geautomatiseerd lezen, namelijk dat dit vrijwel geen beroep doet op bewuste aandacht. Het bekende artikel van LaBerge & Samuels uit 1974 gaat juist over onvoldoende automaticiteit als een van de belangrijkste factoren bij leesproblemen en stipuleert het belang van het blijven oefenen van leestaken nadát accuratesse bereikt is. Dowhower (1987) toonde het effect aan van herhaald lezen, niet alleen op de leessnelheid, maar ook op het begrip van het gelezene; Tan & Nicholson (1997) toonden aan dat training op snelheid met behulp van flitsoefeningen van het lezen van woorden juist voor zwakke lezers bevorderlijk is voor hun leesbegrip. Allington (1983) schreef een artikel *Fluency: the neglected reading goal in reading instruction*. Lovett

(1987) onderscheidde *accuracy-disabled* en *rate-disabled* lezers. De *accuracy-disabled* maken veel fouten én lezen langzaam, de *rate-disabled* definieert zij als leerlingen die qua accuratesse op leeftijdsniveau lezen en alléén problemen hebben wat betreft hun leessnelheid. Binder (1996) beschrijft het belang van *fluency* als hét kenmerk van beheersing en citeert Skinner uit 1938 in een pleidooi voor *fluency-based instruction: the rate of responding is the principal measurement of the strength of an operant*.

De conclusie uit het voorgaande is dat het stellen van accuratessedoelen gebruikelijk, maar het stellen van fluencydoelen niet of nauwelijks voorkomt en dit lijkt voor de onderwijssituatie in Nederland evenzeer op te gaan als voor de Amerikaanse. Dat is opmerkelijk, omdat op andere gebieden, zoals sport en muziek, voortdurend sprake is van fluencydoelen. Een uitwerking van toegepaste gedragsleer binnen de didactiek in het algemeen vinden we in het *instructional hierarchy* model (Haring e.a. 1978). In het model worden in het leren vier stadia onderscheiden: *acquisition*, *fluency*, *generalization* en *adaptation*. Elke fase vraagt om toepassing van verschillende instructieprincipes. Aanleren vraagt *modeling* en *prompting*, voor het ontwikkelen van *fluency* zijn *drill* en *reinforcement* nodig, generalisatie wordt ontwikkeld door training in een natuurlijk context en adaptatie ten slotte betekent dat de vaardigheid wordt ingezet voor het oplossen van nieuwe problemen. In termen van dit didactische model lijkt de praktijk van het technisch leesonderwijs te zijn om na het aanleren direct door te schieten naar de fase van de adaptatie. Dat het voor het bereiken van *fluency* nodig is om te oefenen, werd door Huey begin vorige eeuw al geschreven. Maar *drill and practice* zijn in het onderwijs niet meer populair. Er bestaat volgens Binder (1996) een aantal populaire misvattingen over, zoals het idee dat *drill and practice* van deelvaardigheden het verwerven van hogere probleemoplossingsvaardigheden in de weg staat, terwijl juist onvoldoende geautomatiseerde deelvaardigheden het verwerven van samengestelde vaardigheden belemmeren (Kubina & Morrison 2000). Dit is ook weer een van de belangrijke conclusies van het National Reading Panel ten aanzien van de relatie tussen woordherkenning en bijvoorbeeld begrijpend lezen. Ook bestaat het idee dat het perfect uitvoeren van (deel)vaardigheden voor leerlingen vervelend zou zijn. Afgezien van de vraag naar de geldigheid van de 'Sesamstraatdoctrine' dat leren vanzelf hoort te gaan en altijd leuk zou moeten zijn (Lindsley 1992), is het correct én vlot kunnen uitvoeren van vaardigheden juist helemaal niet vervelend, zoals ook weer bij bijvoorbeeld sport en muziek is waar te nemen.

Fluency blijkt uit het korter worden van de responstijd en impliceert dat het beslag op *attentional resources* minder wordt. Is de leesteknik voldoende, dan is de te leveren inspanning gering en komt de inhoud beschikbaar. Is deze interessant genoeg, dan houdt het leesgedrag zichzelf in stand; het lezen als activiteit roept zijn eigen bekrachtiging op, $+C^+$. Met het toenemen van de leesteknik zal de leerling steeds minder afhankelijk worden van externe bekrachtiging. De leesactiviteit krijgt per saldo intrinsiek belonende waarde. In het spraakgebruik wordt ten onrechte niet de (lees)activiteit intrinsiek motiverend, maar de leerling voor wie dit geldt, intrinsiek gemotiveerd genoemd.

Gresham (1988, 1998) onderscheidt vaardigheidstekorten (*skill deficits*), uitvoeringstekorten (*performance deficits*) en *fluency*-tekorten. Als leerlingen gewenst (lees)gedrag niet laten zien, wordt vaak geprobeerd de leerling te motiveren door bijvoorbeeld het (lees)materiaal qua uitvoering of inhoud aantrekkelijker te maken. Dit gaat uit van een uitvoeringstekort, vrijwel altijd impliciet, waardoor verzuimd wordt om deze aanname te controleren. Maar als de (lees)vaardigheid onvoldoende is, als er met andere woorden sprake is van een vaardigheidstekort, kan het gewenste gedrag niet gesteld worden. Gresham wijst op *competing behaviors* als interfererende responsen. Een interfererende respons is een niet gewenste, maar sterkere respons op dezelfde stimulus die als die zich voordoet, belemmerend is voor het stellen van het gewenste gedrag (leidend tot een uitvoeringstekort) of voor het verwerven ervan (leidend tot een vaardigheidstekort). Het begrip faalangst is in dit verband erg populair. Leesstof roept een angstrespons op die interfereert met het stellen van het gewenste gedrag, namelijk het leveren van een adequate (lees)prestatie. Ook hier wordt alleen rekening gehouden met de mogelijkheid van een uitvoeringstekort, de leerling levert een prestatie onder het eigenlijke kunnen (Poulie 1991) en de behandeling is gericht op het laten afnemen van de interfererende angstrespons. Maar ook hier geldt dat als de leesvaardigheid onvoldoende ontwikkeld is, deze niet op het gewenste niveau kan worden gesteld.

Volgens Skinner (1998) maakt *fluency* het verschil uit tussen kunnen en doen, want alleen gedrag dat behalve accuraat ook vlot kan worden uitgevoerd, wordt ook graag uitgevoerd. Alleen gedrag dat *fluent* is, is met andere woorden intrinsiek motiverend. Van der Leij & Van Daal (1999) concluderen, onder verwijzing naar onder andere Reitsma (1983) en Van Daal (1993), dat zwakke lezers in verhouding tot niet-zwakke lezers om te beginnen veel meer trials nodig hebben voor het bereiken van accuratesse en daarna langzamer blijven lezen, ook na veel trials. Als de te leveren inspanning groot blijft, zoals in geval van leerlingen met leesproblemen, wordt de leesactiviteit per saldo weinig belonend of, in geval van veelvuldig falen, zelfs intrinsiek straffend, +C⁻.

Binder (1996) vindt met een van LaBerge & Samuels (1974) afgeleide procedure een sterke relatie tussen responsefrequentie, als maat voor automaticiteit, en afleidbaarheid, gemeten als de onderdrukking van de gewenste respons door de aanwezigheid van externe stimuli. Zo ligt, als het lezen nog moeizaam gaat, de ontwikkeling van taakvermijdend gedrag voor de hand. Door het vertonen van taakvermijdend gedrag wordt ook de negatieve consequentie (C⁻) van het gewenste gedrag vermeden, -C⁻. Dit is een zogenaamde negatieve beloning voor het taakvermijdend gedrag. Ook taakvermijdend gedrag houdt op deze manier zichzelf in stand. De leerling die taakvermijdend gedrag vertoont, wordt in het spraakgebruik ongemotiveerd genoemd, eveneens ten onrechte.

Oefenen is *the repetition of a given response class after it has been accurately established in a repertoire* (Binder 1993). Hiervoor worden ook de termen *overlearning* of *overtraining* gebruikt. Maar zonder een duidelijke doel en zonder feedback is oefenen weinig belonend. Traditioneel staat oefenen in het onderwijs daarom ook nog eens onder aversieve controle, dat wil zeggen dat alleen het *niet* voldoen aan de opdracht wordt gestraft, in plaats dat het *wel* voldoen ook wordt beloond. Ten slotte is het wél

vervelend - in de zin dat de activiteit onvoldoende bekrachtigers oproept - of zelfs uitgesproken aversief om niet-beheerste deelvaardigheden herhaaldelijk te moeten uitvoeren. Juist dan blijkt oefenen kwetsbaar voor niet-taakgericht gedrag, dat onder die condities negatief beloond wordt. Al met al is in de huidige onderwijspraktijk het oefenen van (deel)vaardigheden die al wel accuraat, maar nog onvoldoende vlot kunnen worden uitgevoerd, belast met negatieve consequenties. Niet door de aard ervan, maar door de manier waarop er het in het onderwijs mee wordt omgegaan. Daardoor vormt het bijna een garantie voor storend, niet-taakgericht gedrag. Het niet regelmatig en voldoende inpassen van dit type oefening in het lesprogramma kan dan ook heel goed worden geïnterpreteerd als vermijdingsgedrag ... van de leerkracht.

De conclusie van het National Reading Panel met betrekking tot het oefenen van *fluency* luidt dat dit niet alleen met losse woorden dient te gebeuren, maar ook door het lezen van teksten, omdat *competent reading requires skills that extend beyond the single-word level to contextual reading*. Het belangrijkste daarbij is een sterke uitbreiding van de tijd dat zwakke lezers daadwerkelijk aan het lezen zijn, in de vorm van *guided oral reading* met *one-to-one tutoring*, uitgevoerd door beroepskrachten, met directe instructie en correctieve feedback (ook Wasik & Slavin 1994). Levy (2001) formuleert dezelfde aanbeveling net andersom: trainen moet niet alleen met teksten gebeuren, maar ook met losse woorden. In termen van het *instructional hierarchy* model is het lezen van losse woorden de deelvaardigheid en zijn teksten de natuurlijke context die nodig is voor generalisatie.

Stevens (1996) noemt de behoefte om competent te zijn een van de basisbehoeften. Of, hoe en wanneer 'iets kunnen' zijn eigen bekrachtiging inhoudt, is een vraag die in dit verband niet beantwoord hoeft te worden. In ieder geval is 'iets kunnen' vooral ook een middel is om toegang te krijgen tot bekrachtigers, zoals bijvoorbeeld leesteknik toegang geeft tot interessante tekstinhoud. Voldoende succes is derhalve essentieel in de fase waarin een leerling zich de leesteknik eigen maakt. Lezen houdt altijd het leveren van inspanning in. Die inspanning is gering als de leesvaardigheid goed en voldoende geautomatiseerd is, maar hoe geringer de leesvaardigheid, des te groter de inspanning. Kinderen blijven lezen zolang dat voldoende bekrachtiging oplevert in verhouding tot de inspanning. Is de inspanning nog groot, dan zal het juiste leesgedrag opgeroepen moeten worden (antecedente gedragscontrole) en de bekrachtiging extern moeten worden aangevuld (consequente gedragscontrole). Dit is, in TGL-termen, de essentie van leesinstructie.

4.2 Van leesmodel naar didactiek

Dit proefschrift gaat over de techniek van het lezen, leesproblemen en dyslexie. Met leesteknik wordt de vaardigheid bedoeld om woorden te decoderen. In de literatuur wordt hiervoor het woord *decoding* gebruikt en soms *encoding*, maar het verschil tussen beide begrippen wordt meestal niet nader gedefinieerd. Foorman (1995) pleit voor

onderscheiden gebruik van de termen: *decoding* gaat in de definitie van Foorman vooraf aan herkennen, het is de *sounding out* route, *encoding* is het herkennen van het woord en beschrijft de - fonologische kant van de - directe route. Hogaboam & Perfetti (1978) spreken van *phonetic encoding* als het toewijzen van de fonetische code: *the translation stage in which letter sequences are assigned phonetic codes*.

Er zijn vele varianten van het zogenaamde duale route model. Duale route modellen onderscheiden het lezen via letter-voor-letter decoderen van directe woordherkenning, waarbij de lezer op basis van het geschreven woord direct toegang krijgt tot een mentaal lexicon. In een longitudinaal onderzoek naar verschillende niveaus van verwerking bij het lezen van woorden, waarbij leerlingen van klas 1 tot 3 gevolgd werden, toonden Cunningham e.a. (2001) aan dat *orthographic processing* naast fonologische vaardigheden een uniek en met de jaren toenemend aandeel verklaart van de variantie van woordherkenning. In onderzoek van Barker e.a. (1992) bleken orthografische vaardigheden in het bijzonder voorspellende waarde te hebben ten aanzien van de *fluency* van het lezen, onafhankelijk van fonologische vaardigheden. Er lijkt dus wel aanleiding verschillende routes te onderscheiden. Misschien zouden modellen waarin expliciet plaats is voor eenheden tussen grafeem en woord best als meervoudige route modellen kunnen worden aangeduid. Variatie tussen de verschillende modellen zit vooral in de mate van samenhang en interactie tussen de verschillende routes, variërend van volledig gescheiden, zoals in het oorspronkelijke *horse race* model van Coltheart (1978) tot een voortdurende ondersteuning van elkaar.

Share (1995) beschrijft het decoderen van woorden die nog niet direct herkend worden, als *self-teaching*. Met het ontsleutelen van een nog niet herkend woord via verklanking van afzonderlijke grafemen, gevolgd door auditieve synthese, meestal aangeduid als spellend lezen, creëren leerlingen voor zichzelf een leermoment omdat geschreven en klankvorm van het woord dan contigu aanwezig zijn. Het is het lezen van een woord via een omweg, dat na voldoende oefening overgaat in directe woordherkenning. In het model van Share is het lezen door middel van verklanking *item-based* en niet *phase-based*. Het betekent dat voor het aanleren van elk nog onbekend woord in welke fase van de leesontwikkeling dan ook *self-teaching* plaatsvindt.

Perfetti (1992) onderscheidt een functioneel en een autonoom lexicon (zie ook Foorman 1994 en Rayner e.a. 2001). Het functionele lexicon is toegankelijk op basis van grafeem-foneem-correspondentieregels. Het autonome lexicon is toegankelijk op basis van abstracte kennis omtrent de orthografie van het woord, die automatische herkenning mogelijk maakt (*encoding*). De opbouw hiervan is *item based* en in de normale ontwikkeling is sprake van een wederkerige relatie tussen fonologische en orthografische representatie, met als resultaat dat orthografische representaties worden ontwikkeld die geschikt zijn voor transfer naar nieuwe woorden *by mapping*. Slechte kwaliteit van de fonologische representatie zou enigszins gecompenseerd kunnen worden door een betere kwaliteit van orthografische representatie.

Morton (1980) geeft een zogenaamd interactief activatiemodel voor woordherkenning in zijn logogen model. Een logogen is een *device* die een woord beschikbaar

maakt als een respons. Dit gebeurt door het vergaren van informatie. Hierbij doet het er niet toe van welke aard de informatie is. Ook kan informatie van verschillende aard bij elkaar worden opgeteld. Komt de totale hoeveelheid informatie boven een zekere grens, dan is het woord herkend, dat wil zeggen het is als interne respons beschikbaar.

In sommige modellen gaat het ook bij directe woordherkenning om een vorm van fonologisch decoderen. Torgesen (2000) bijvoorbeeld formuleert dat fonologische decodeervaardigheden een kritische *ondersteunende* rol spelen als leerlingen de *orthografische* leesvaardigheden beginnen te verwerven, die relatief vlotte en moeiteloze woordidentificatie mogelijk maken. Leerlingen met leesproblemen lijken vooral beperkt in het aantal woorden dat zij als orthografische eenheden kunnen herkennen. De oorzaak van problemen met woordherkenning ligt hoofdzakelijk in een geringe vaardigheid om de fonologische kenmerken van taal te verwerken. Essentieel in duale route modellen is dat er wordt uitgegaan van het bestaan van een systeem van betekenisvolle eenheden - woorden - die in het geheugen liggen opgeslagen, die op onderscheiden wijzen benaderd kunnen worden.

Daartegenover staan de *single* of enkelvoudige route modellen op basis van de theorie van het connectionisme. In connectionistische modellen wordt herkenning van een woord gelijk gesteld aan de gespreide of parallelle activatie van de fonologische code. Het decoderen van een geschreven woord naar de klankvorm kan gebeuren op het niveau van afzonderlijke grafeem-foneem koppelingen en op subwoordniveau zoals lettergrepen of letterclusters (Seidenberg & McLelland 1989).

Foorman (1994, 1995) komt tot een integratieve opvatting. Alle kinderen maken een ontwikkeling door van, in Nederlandse begrippen, globale woordherkenning, de *logographic phase* genoemd, via een spellende leesstrategie, de *alphabetic phase*, naar accurate en snelle directe woordherkenning, aangeduid als de *orthographic phase*, een fase die wordt gekenmerkt door woordspecifieke associaties tussen geschreven woorden en hun klankvorm. Het gaat daarbij niet om het toepassen van bepaalde vaardigheden, maar om de kwaliteit van de verbinding tussen orthografische input en *phonological resonance*. Fonologische en orthografische *encoding* zijn wel te onderscheiden en, door taken slim te kiezen, ook afzonderlijk te meten, maar het zijn principieel twee kanten van dezelfde zaak. Ook Ehri (1998) gaat uit van meerdere fasen in de ontwikkeling van directe herkenning van wat zij noemt *sight words*, maar ook deze is gebaseerd op een volledige verbinding tussen de grafemen in het geschreven woord en de foneemen in de klankvorm van het woord.

Elbro's (1996) *distinctness of phonological representation* hypothese betreft vooral een verschuiving in oriëntatie van stimuluscondities, zoals bij de hypothese van het fonologisch tekort, naar de kwaliteit van de representatie, niet in relatie tot automaticiteit, maar veeleer als een specificatie van het model van de fonologische verwerking dat de basis is van de hypothese van het fonologisch tekort. De kern van automatisering is gelegen in de associatie van fonologische en orthografische representaties op lexicaal en sublexicaal niveau. In tegenstelling tot het model van de fonologische verwerking of van het fonologisch bewustzijn, gaat het bij de kwaliteit van de fonolo-

gische representatie om lagere orde aandachtsvrije processen. De fonologische tekorten van dyslectici zijn een gevolg van onvoldoende kwaliteit van de fonologische representaties van te lezen woorden. De ontwikkeling van de fonologische representaties verloopt van hele woord naar steeds kleinere eenheden en een steeds grotere mate van detaillering. Dyslectici haken af in dit proces van toenemende segmentering. De problemen van dyslectici zouden dan het gevolg zijn van *imprecise representations* (Swan & Goswami 1997). Van Orden e.a. (1997) spreken van de noodzaak van *fine-grain knowledge of phonology* voor een goede leesvaardigheid. Leerlingen met leesproblemen onderscheiden zich hierin blijvend van andere leerlingen. Hun prestaties op tests die meer subtiele aspecten van de fonologische representatie meten, zijn zwakker. Ook op latere leeftijd en ook als zij inmiddels een hoger leesniveau bereikt hebben (De Jong en Van der Leij, 2003).

Dat de fonologische code van de taal bij het lezen een rol speelt, lijkt wel zeker, wat daarvan ook de precieze betekenis is. Maar wat betekenen de verschillen tussen de modellen voor de didactiek? Op basis van duale of meervoudige routemodellen wordt in de eerste fase van het aanvankelijk lezen expliciet de omweg van de zelfinstructie aangeleerd door middel van het aanleren van klank-tekenkoppelingen en auditieve synthese: de *phonics*. Worden deze eenmaal voldoende beheerst om eenvoudige woorden te ontsleutelen, dan komt de nadruk te liggen op de directe woordherkenning.

Het connectionistische leesmodel gaat uit van het zogenaamde covariate leren (Van Orden e.a. 1990). Een didactische uitwerking van het covariate leren is beschreven door Foorman (1994) en toegepast door Van den Broeck (1997). Leerlingen kregen woorden aangeboden die zij moesten lezen. Een specifiek covariate notie is de samenstelling van de trainingwoorden. Deze dienen overlap te vertonen en in een willekeurige volgorde aangeboden te worden. Dit principe kan in een experimentele opzet systematisch en gecontroleerd worden vormgegeven. Als algemeen didactisch principe is het nauwelijks mogelijk om zich, met uitzondering van zogenaamde wisselrijtjes, leestof van enige omvang voor te stellen die niet aan deze eisen voldoet. Vanuit het principe van het covariate leren zou ook al in de eerste fase van het leren lezen uitsluitend met woorden gewerkt moeten worden, beginnend met woorden van de allereenvoudigste km- en mk-structuur (vergelijk Feys 1995). Deze aanpak in de allervroegste fase van het leren lezen is echter slechts bij wijze van remediëring bij leerlingen beschreven (Verhagen e.a. 1997). Bij een training op basis van het covariate principe dienen leerlingen ten slotte bij correct lezen stimulusspecifieke feedback te krijgen. Van den Broeck voegde daar bij correct lezen nog het verkorten van de aanbestedingsduur aan toe. Deze laatste condities kunnen echter moeilijk als specifiek voor het covariate leren worden geclaimd (voor een overzicht zie bijvoorbeeld Van der Leij 1994).

Na de fase van het aanvankelijk lezen is er in de didactiek geen verschil, welke theoretische oriëntatie men ook kiest. Snelle directe woordherkenning zal geoefend moeten worden door uitbreiding van de tijd dat zwakke lezers daadwerkelijk aan het lezen zijn, met directe instructie en correctieve feedback.

woordherkenning als discriminatieleren

In de literatuur die handelt over lezen, leesproblemen en dyslexie treft men de leertheorie of het toepassen van gedragstherapeutische technieken aan waar het gaat over het inrichten van de leeromgeving waarbinnen de leerling bezig is te leren lezen, zoals beschreven in hoofdstuk 1. Maar als het gaat over het leren lezen zelf is het begrippenkader uit de leertheorie opvallend afwezig, om het even of het experimenteel toetsingsonderzoek betreft of interventiestudies. Toch heeft technisch lezen hoe dan ook te maken met het leren van associaties, spreken we bij dyslexie over een *leerstornis* en laat leesdidactiek zich uitstekend beschrijven in termen van stimulus (bijvoorbeeld instructie, lesmateriaal, leeromgeving), respons (zowel gewenst leergedrag als ongewenst, niet-taakgericht of vermijdingsgedrag) en consequentie (zoals de feedback van de leerkracht of het gevoel bij de leerling wanneer iets wel of niet gelukt is).

Ook aan het onderzoek naar basale leerprocessen wordt binnen het leesonderzoek niet vaak gerefereerd. Een uitzondering is Van den Broeck (1997); hij brengt de verwachtingsdiscrepantiehypothese van Rescorla en Wagner in verband met de in zijn onderzoek gebruikte covariate leerprocedures. De verwachtingsdiscrepantiehypothese houdt in, dat de associatie tussen twee stimuli alleen geleerd wordt als de ongeconditioneerde stimulus zich onverwacht voordoet. De aanvankelijk neutrale stimulus wordt vervolgens als het ware de verklaring voor het optreden van de ongeconditioneerde stimulus; deze krijgt zo zijn signaalwaarde of betekenis voor de dan geconditioneerde respons (Rescorla & Wagner 1972, Eelen 1988). Ook door Rescorla (1988) is de overeenkomst tussen het connectionisme en klassieke of Pavlovianse conditionering naar voren gebracht. Beide theorieën gaan uit van interacties tussen meervoudige associaties met complexe responsen als resultaat, bijvoorbeeld de relatie tussen spraakperceptie en spraakproductie.

Binnen de leertheorie is het scherpe onderscheid tussen klassieke en operante conditionering inmiddels verlaten en gaat men er veeleer van uit dat binnen alle leren aspecten van beide schema's aanwezig zijn. Ook wordt het onderscheid tussen ongeconditioneerde en geconditioneerde stimulus veel minder sterk benadrukt en wordt veel meer belang gehecht aan het leren van associaties tussen stimuli in het algemeen. Eelen (1988) benadrukt ook het onderscheid tussen *leerproces* en *leerprocedure*⁴. Dit onderscheid kan van dienst zijn bij het verhelderen van de relatie tussen psychologische kennis en de toepassing daarvan in de didactiek.

⁴ Een door Eelen gegeven voorbeeld van het leren van de relatie tussen een lichtschakelaar en het aangaan van het licht is illustratief. De beschrijving van het leren van de werking van een lichtschakelaar binnen het operante paradigma komt erop neer dat een toevallige handeling, het omzetten van de schakelaar, wordt gevolgd door een interessante consequentie, $+C^+$, als het licht aangaat. Als gevolg hiervan neemt het schakelgedrag toe. Deze gang van zaken is echter erg onwaarschijnlijk. Veeleer zal het plotseling aangaan van het licht, een situatie die verrast en discrepant is met de verwachting, worden 'verklaard' uit een handeling die men zojuist iemand anders heeft zien uitvoeren, het omzetten van de schakelaar. Aldus is niet zozeer de relatie tussen een operant of handeling en het gevolg daarvan geleerd, maar veeleer een associatie tussen twee stimuli. In het voorbeeld is de operant, het schakelgedrag, zeker geen ongeconditioneerde respons; wel verliep de *leerprocedure* niet via het operante, maar via het klassieke conditioneringsschema.

Zwakke lezers hebben in verhouding tot niet-zwakke lezers veel meer trials nodig voor het bereiken van accuratesse, maar blijven daarna langzamer lezen, ook na veel trials (Van der Leij & Van Daal 1989). Hierin liggen twee maten voor - de kwaliteit van - het leren besloten: het aantal benodigde trials om iets tot een bepaald criterium te leren en de responssnelheid (zie ook Samuels 1979). Wimmer & Mayringer (2001) schrijven dat er overlap in tijd moet zijn tussen de activatie van de fonologische representatie van een woord (en de daarin besloten fonemen) en de activatie van de letterherkenningseenheden. Als vanuit de gebieden in de hersenen die betrokken zijn bij de verwerking van letterpatronen het toegang krijgen tot de fonologische woordrepresentatie te langzaam gaat, zijn de visuele gebieden mogelijk niet langer voldoende actief en treedt er weinig associatievorming op. Als daardoor de visueel-fonologische associaties weinig stabiel zijn of er zijn veel meer trials nodig zijn om stabiliteit te bereiken, dan zal het orthografisch lexicon verarmd zijn. Nicolson e.a (2002) voerden een klassiek conditioneringsexperiment uit waarbij zij gebruik maakten van de ongeconditioneerde *eyeblick* response. Dyslectici bleken significant slechter conditioneerbaar. Dit werd door de auteurs opgevat als een bevestiging van hun *cerebellar deficit* hypothese, maar zou mogelijk ook kunnen wijzen in de richting van een meer algemene geringere conditioneerbaarheid. Anders gezegd, de opbouw van orthografische woordrepresentaties kan belemmerd worden door beperkte associatievorming tussen orthografische en fonologische representaties.

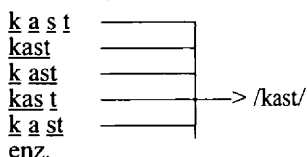
een woord is vele stimuli

Uit onderzoek naar aandachtsprocessen is het zogenaamde *Stroop*paradigma zeer bekend (Stroop 1935), onder andere in de vorm van de Stroop kleur-woord test. Op een kaart staan namen van kleuren gedrukt in verschillende kleuren inkt, bijvoorbeeld het woord rood staat gedrukt in *blauwe* inkt. Het woord is dus tegelijkertijd twee verschillende stimuli: het geschreven woord - rood - en de kleur van de inkt - blauw. De stimuli zijn geassocieerd met verschillende responsen, in het voorbeeld /rood/ en /blauw/. In onderzoek naar selectieve aandacht gaat het om de vaardigheid van de proefpersoon de verwerking van de meest geautomatiseerde stimulus-respons associatie, gewoonlijk is dat het lezen van het geschreven woord rood, te onderdrukken en alleen op de *blauw*-stimulus te reageren. Naarmate iemand beter in staat is selectieve aandacht te geven aan een bepaald aspect van een stimulusconfiguratie, is hij ook beter in staat soepel om te schakelen op een ander aspect, als de taak dat vraagt.

Het lezen van woorden kan in een aantal opzichten met de Strooptaak vergeleken worden. Voor vlotte lezers is het geschreven woord niet slechts één stimulus. Een geschreven woord *kán* verwerkt worden als één stimulus: bijvoorbeeld kast. Maar ook als een serie van vier stimuli: k a s t. Als er sprake is van een woord dat uit meerdere lettergrepen bestaat, is het aannemelijk dat ook de lettergreep als een eenheid kan worden waargenomen (Wentink 1997). Veel Engelstalig onderzoek is gedaan naar de verdeling van een woord in zogenaamde *onset* en *rhyme* of *word body*. Ook deze lijken, althans in het Engels, als subwoordeenheden een rol te spelen. Ten slotte zijn

ook medeklinkerclusters, de st in kast, voor de geoefende lezer waarschijnlijk mogelijke verwerkingseenheden of *chunks* (Stanovich 1992, Van Daal 1993). Het geschreven woord is daarin gelijktijdig verschillende stimuli:

Sd: Il ziet (interne) R: Il herkent



Het geschreven woord is zo een ambigue stimulusconfiguratie die, vergelijkbaar met de Strooptaak, op verschillende manieren gecodeerd kan worden, maar met als belangrijk verschil met de Strooptaak, dat bij het lezen van een woord verwerking op elk niveau is geassocieerd met dezelfde interne respons, althans bij klankzuivere woorden. Volgens Adams (1990) is het kenmerk van goede lezers dat verwerking ook inderdaad volledig geautomatiseerd en gelijktijdig op alle mogelijke niveaus plaatsvindt, inclusief het benutten van contextinformatie zoals bijvoorbeeld in Morton's logogenmodel.

Het lezen van een woord is beschreven als het toewijzen van een fonologische code, conform Hogaboam & Perfetti (1978). Het begrip toewijzen zou wellicht teveel kunnen worden gelezen als een activiteit die met inzet van bewuste aandacht en dus relatief langzaam wordt uitgevoerd, terwijl het kenmerk van een geautomatiseerd verlopende activiteit juist is dat deze snel verloopt en niet of nauwelijks een beslag legt op de aandachtscapaciteit. Juist een goede, dat wil zeggen snelle prestatie op bijvoorbeeld de Strooptaak, houdt in dat de proefpersoon zijn gedrag flexibel weet af te stemmen op de eisen van de taak. Een onderzoek van Monsel e.a. (1992) met volwassen, ervaren lezers is illustratief. Pseudowoorden worden sneller gelezen als ze gemixed zijn met hoogfrequente woorden, dan wanneer ze gemixed zijn met laagfrequente woorden. Hoogfrequente woorden worden langzamer gelezen als ze gemixed met pseudowoorden worden aangeboden in *pure blocks*, of wel wisselrijtjes. Ook als het leesproces geautomatiseerd is, is het niet uitsluitend *data driven*, maar kan er sprake zijn van strategische anticipatie van bijvoorbeeld de lexicale status van een woord. Het onderzoek Monsel e.a. laat ook zien hoe bepalend stimuluscondities zijn voor het gedrag dat wordt opgeroepen.

Goede lezers lezen laagfrequente woorden nauwelijks langzamer dan hoogfrequente woorden en ook pseudowoorden nauwelijks langzamer dan bestaande woorden (Van der Leij 1983, Seymour 1986, Van der Leij & Van Daal 1999), afhankelijk van leeservaring en opleidingsniveau. Kenmerkend voor zwakke lezers lijkt vooral dat zij veel minder in staat zijn om, als de taak daarom vraagt, soepel en efficiënt te schakelen tussen verschillende niveaus van verwerking van geschreven materiaal. Aangezien pseudowoorden niet als eenheid bekend kunnen zijn, staat het vlot decoderen op subwoordniveau klaarblijkelijk goede lezers ook in het gevorderde lezen ter beschikking. Uiteindelijk onderscheiden oudere zwakke lezers zich niet van goede lezers in het lezen van hoogfrequente woorden. Ook zwakke lezers blijven, zij het na veel meer

oefening, woorden als eenheid voldoende vlot en accuraat te kunnen verwerken. Bij zwakke lezers zijn er wel verschillen tussen woorden en pseudowoorden (Yap 1993, Smeets 1997, Van der Leij & Van Daal 1999). Yap (zie ook Yap & Van der Leij 1993) toonde aan dat het pseudowoordeffect bij dyslectici versterkt wordt door beperking van de presentatieduur. Snel waarnemen van sublexicale eenheden is klaarblijkelijk van belang voor het vlot lezen van onbekende woorden. Niet alle verwerkingsmogelijkheden lijken voor zwakke lezers in gelijke mate open te staan, vooral niet die waarbij eenheden op subwoordniveau efficiënt verwerkt worden (vergelijk Logan 1988, 1990).

beperking van de presentatieduur

De orthodidactische maatregel die zwakke lezers in een cruciale fase van hun leesontwikkeling moet helpen, zal in de eerste plaats het stellen van een ongewenste respons, te langzaam, spellend lezen, moeten voorkomen. In het model van Gresham (1998) is er sprake van interfererende responsen, de verwerking van het woord op een te laag niveau, die het *leren* of verwerven van de gewenste respons - directe woordherkenning - belemmert. Presentatieduurbeperving of flitsen lijkt daarvoor de aangewezen weg (Yap 1993, Das-Smaal e.a. 1996). Uit onderzoek van Van den Bosch (1991) bleek presentatieduurbeperving effectiever dan responsversnelling: het flitsen van woorden leidde tot een verkorting van de responstijd, bij een onbeperkte aanbiedingsduur lukte het veel minder de responstijd te laten afnemen, zelfs als de instructie was sneller te lezen en snel antwoorden beloond werd. In TGL-termen uitgedrukt, differentiële bekrachtiging van het gewenste gedrag was minder effectief dan antecedente controle.

Presentatieduurbeperving zorgt ervoor dat het woord niet langer op een te laag niveau kan worden verwerkt, waardoor een zwakke respons die het gewoonlijk verliest, de kans krijgt. Maar gedragsinstructie in combinatie met differentiële bekrachtiging blijkt onvoldoende om bij zwakke lezers verwerking op een hoger niveau uit te lokken. De training moet ook zo worden ingericht, dat naast het voorkomen van interfererende responsen, het uiteindelijk gewenste gedrag in stappen wordt benaderd - *shaping*. Daartoe wordt de moeilijkheidsgraad van de leertaak zo gemanipuleerd, dat deze voor de leerling steeds *nét* te doen is. Ook moet de leerling het gewenste (leer)gedrag veelvuldig kunnen laten zien. Op dat punt toonden Belfiore e.a. (1995) aan dat in een woordtrainingsexperiment met behulp van een computer niet de frequentie van de juiste respons de meest bepalende factor was, maar de vraag of sprake was van *trial repetition*. Van een *learning trial* is alleen sprake als op een discriminatieve stimulus, gevolgd door een juiste respons, ook contingent bekrachtiging volgt (Skinner 1998). Het alleen maar frequent stellen van de juiste respons, zonder dat deze bekrachtigd wordt, betekent dat op de respons geen enkele consequentie volgt, met andere woorden dat het gewenste gedrag genegeerd wordt. En negeren, ook wel aangeduid met C^0 , leidt tot afname van de frequentie van gedrag.

beperking van het aantal responsalternatieven

Voor het leren van associaties tussen stimuli en responsen, is de moeilijkheid van de

taak ook een functie van het aantal te leren associaties, of wel: hoe beperkter het aantal associaties en daarmee het aantal responsalternatieven, des te gemakkelijker de taak. Het beperken van het aantal responsalternatieven is een veelvuldig toegepast didactisch principe. Methodes voor aanvankelijk lezen introduceren per kern steeds een beperkt aantal nieuwe klanktekenkoppelingen. Daarmee kan maar een beperkt aantal nieuwe woorden gemaakt worden. Die worden, als het goed is, net zo lang geoefend tot ze accuraat én snel gelezen worden. Pas daarna volgen nieuwe klanktekenkoppelingen. Hiermee wordt tijdens het leerproces het aantal responsalternatieven beperkt.

Hoogfrequente woorden worden beter gelezen dan laagfrequente woorden. Ook dit zogenaamde frequentie-effect kan begrepen worden op basis van beperking van het aantal responsalternatieven. De kans dat een willekeurig woord een hoogfrequent woord is, is - per definitie - groter dan de kans dat het een laagfrequent woord is. Door in het kader van een training met een beperkte set van woorden te werken, worden die woorden, binnen dat kader, tot hoogfrequente woorden, of dat in de gebruikelijke betekenis van het woord nu hoogfrequente woorden zijn of niet.

Woorden in zinsverband worden met minder fouten en sneller gelezen dan losse woorden. Bij een woord in de context van een zin is het aantal responsalternatieven beperkt; het woord moet in de zin passen. Meyer & Felton (1999) pleiten voor het gebruik van redundante oefenteksten. Redundantie in een tekst staat gelijk aan een nog verdere beperking van responsalternatieven.

Ook het gebruik als oefenwoorden binnen een trainingsexperiment van zogenaamde orthografische buurwoorden, woorden die qua spelling minimaal van elkaar verschillen, of *friends*, orthografische buurwoorden waarvan het overlappende deel tevens verwijst naar eenzelfde klankpatroon, is een vorm van redundantie. Soms worden de termen alleen gebruikt voor woorden die maar één letter verschillen (Van den Broeck 1997), anderen stellen dat woorden meer of minder burens of vriendje kunnen zijn, al naar gelang ze een of meer letters van elkaar verschillen (Reitsma 1997). *Kast-last-mast-vast* is als set trainingwoorden redundanter dan *kast-lust-mest-vist*. De eerste set van woorden voldoet tevens aan het *onset-rhyme* principe, dat in het Engels van groot belang lijkt, maar waarvan de betekenis in Nederlands onderzoek niet eenduidig kon worden gerepliceerd. Hogaboam & Perfetti (1978) konden transfer aantonen naar woorden met gelijk rijm. Maar Van den Bosch (1991) kon geen *onset-rhyme* effect aantonen en Assink e.a. (1998) vonden verschillende effecten voor normale versus zwakke lezers en voor hoogfrequente versus laagfrequente clusters.

Gottardo e.a. (1999) lieten negenjarige dyslectici en qua leesvaardigheid vergelijkbare zevenjarigen woorden en pseudowoorden lezen van vier typen:

regelmatig consistent: woorden waarvan de klinkers geassocieerd zijn met de meest frequent voorkomende klank en de *word bodies* - het rijm - de meest voorkomende regelmatige uitspraak hebben (*stiff, press, blame*);

regelmatig inconsistent: woorden waarvan de klinkers geassocieerd zijn met een frequente klank, maar waarvan de *word bodies* niet de meest voorkomende regelmatige uitspraak hebben (*said* tegenover de uitspraak als in *paid* die geldt als de regelmatige, *have* tegenover *slave*, *give* tegenover *dive*);

onregelmatig consistent: woorden waarvan de klinker met een ongebruikelijke klank geassocieerd is, maar die wel veel orthografische burens hebben, woorden met eenzelfde *word body*, die tevens *friends* zijn (*bind*: *find*, *mind*, *kind*);

onregelmatig inconsistent: woorden waarvan de klinker met een ongebruikelijke klank geassocieerd is, en waarvan de meeste orthografische burens geen *friends* zijn, dat wil zeggen dat de *word body* anders wordt uitgesproken (*pint*: *hint*, *mint*, *lint*);

Zwakke lezers maakten in vergelijking met jongere gemiddelde lezers met hetzelfde leesniveau vooral meer fouten in de klankwaarde die wordt toegekend aan de klinker. Deze is in het Engels sterk contextafhankelijk en hoe sterker de afhankelijkheid van de context, zoals bij woorden uit de categorie *onregelmatig consistent*, hoe groter het verschil is tussen dyslectici en niet-dyslectici. Een dergelijk onderzoek is in het Nederlands niet te repliceren. Vrijwel alle eenlettergrepige (pseudo)woorden in het Nederlands zijn *regelmatig consistent*. Verder zijn *regelmatig consistente* woorden in het Nederlands al consistent op grafeemniveau. De categorieën *regelmatig inconsistent* en *onregelmatig consistent* vallen samen en bestaan, voorzover het eenlettergrepige woorden betreft, slechts uit het beperkte aantal woorden dat eindigt op een enkele klinker, *sta* tegenover *stak*. Bij de meerlettergrepige woorden zijn vooral woorden met een zogenaamde open lettergreep *regelmatig inconsistent*, maar hierbij overschrijdt de contextafhankelijkheid de grens van de *word body*. *Onregelmatig inconsistent* zijn alleen zogenaamde vreemde woorden, *flat* tegenover *plat*. Orthografische burens zijn in het Nederlands dus vrijwel altijd ook *friends* en het ligt voor de hand dat een verschil in betekenis van het *onset-rhyme* principe in het Nederlands en Engels samenhangt met het feit dat in het Engels de klankwaarde van de klinker pas op het niveau van de *rhyme* invariant is.

Ten slotte is ook de wijze van aanbieden van oefenwoorden in zogenaamde wisselrijen (*is*, *vis*, *mis*, *sis*) ook een manier om het aantal responsalternatieven te beperken. Beginnende en zwakke lezers lezen woorden die in structuurrijen worden aangeboden, beter dan bij aanbieding in een willekeurige volgorde. Dit geldt zowel voor het Nederlands (Reitsma 1988, Van den Broeck 1993) als voor het Engels (Perfetti & Bell 1991, Levy e.a. 1999).

automaticiteit

Kenmerken van een geautomatiseerde vaardigheid zijn: snelheid, autonomie, onbewust verloop en moeiteloosheid, in de zin van een gering het beslag op *mental resources*. Dit is geen zaak van alles-of-niets, maar een relatief begrip. De index voor *automatized responding* is de *speed of processing* (Perfetti 1992). Noodzakelijk voor de ontwikkeling van automaticiteit is oefening. Door een taak voldoende vaak te herhalen, zal een leerling deze steeds beter, sneller en met minder aandacht kunnen volbrengen (Binder 1993). Leerlingen laten grote verschillen zien, niet alleen in het aantal trials dat ze nodig hebben om iets te leren, zoals de directe herkenning van een woord (Reitsma 1983), maar ook om daarin vervolgens automaticiteit te bereiken (Van der Leij & Van Daal 1989).

Het effect van extra oefentrials op responstijden na het bereiken van maximale accuratesse op de responstijd werd in algemene zin aangetoond bij *paired associate* leertaken. Daarbij gaat het om de snelheid van het optreden van de juiste respons ofwel de latentietijd, die een functie is van de sterkte van de associatie (Peterson 1965). Eenzelfde positief effect van voortgezette oefening op responstijden is er bij het lezen van woorden (Judd & Glaser 1969, Lemoine e.a. 1993); extra trials na bereiken van accuratesse hebben ook een positief effect op de responstijden bij retentiemeting; zwakke lezers bereiken een plafond pas na meer extra trials dan niet-zwakke lezers, maar het effect van extra trials na bereiken van het plafond is relatief sterker voor zwakke lezers.

Er is discussie over de mate waarin automatisering, de ontwikkeling van automatisiteit, domeinspecifiek is. Opvattingen variëren van automatisering als een meer algemeen kenmerk (Wolf 1991, Nicolson & Fawcett 1990), via specificiteit voor het domein van leestaken, gebonden aan categorieën van woorden, tot uiteindelijk zelfs betrekking hebbend op afzonderlijke woorden. In de standaard opvatting van het begrip, zoals met name door LaBerge & Samuels (1974) gepresenteerd, wordt automatisering gezien als het steeds efficiënter verlopen van de uitvoering van een bepaalde vaardigheid, waardoor voor de uitvoering steeds minder bewuste aandacht nodig is. De standaard opvatting kapitaliseert op generalisatie; het gaat immers om het steeds efficiënter uitvoeren van eenzelfde proces, dat ook kan worden toegepast op nieuwe stimuli. Om die reden noemt Logan (1988) de standaard opvatting van automatisering *process-based*. De verwerving van automatisiteit houdt in het versterken van de connectie tussen stimulus en respons. Oefening kan ook leiden tot *chunking*, het samen nemen van stappen en daardoor een reductie van het aantal te nemen stappen.

Rapid automatized naming, als operationalisering van het aspect benoemsnelheid, verklaart een niet groot, maar wel, ook over een lange reeks van jaren, stabiel deel van de variantie in leesvaardigheid (Van den Bos 2000, De Jong & Van der Leij, 2003), terwijl de rol van fonologisch bewustzijn aanvankelijk groter is, maar afneemt. Wolf & O'Brian (2001) trokken de *unexpected conclusion* dat de rol van *rapid naming* groter is, naarmate de orthografie van een taal regelmatigiger en daarmee de rol van fonologische vaardigheden geringer. Maar in het Engels blijft de maat voor leesvaardigheid in hoofdzaak een accuratessemaat, terwijl EMT en DMT aanvankelijk nog wel ontsleutelen meten, maar al snel en in toenemende mate *speed tests* zijn, wat ook zou gelden voor leesvaardigheidstests in het Duits (Wimmer e.a. 2000). Het is met andere woorden niet vrij van tautologie om te concluderen dat *rapid automatized naming* met de jaren relatief belangrijker wordt en belangrijker is naarmate de orthografie van een taal regelmatigiger is; het is een gevolg van de verschuiving in betekenis van de maat voor technische leesvaardigheid, in relatie tot de regelmatigheid van de orthografie van een taal.

Abbott & Berninger (1999) vonden in een interventieonderzoek met zwakke lezers uit klas 4-7 dat alleen de benoemsnelheid van letters significant voorspelde in welke mate de *real word reading efficiency* van deze leerlingen zich gedurende de training ontwikkelde. Levy e.a. (1999) werkten met zwakke lezers uit klas 2, allemaal zwak in

fonologische vaardigheden, onderverdeeld in snelle en langzame *rapid automatized namers*; de laatsten komen overeen met wat Wolf & Bowers (1999) aanduiden met *double deficit*. Ze ontvingen een van drie verschillende trainingen: *phonics* (lezen door volledige verklanking), *onset/rhyme* segmentering - in uitvoering overeenkomend met het oefenen van wisselrijtjes - en hele woord training met dezelfde woorden als in de *onset/rhyme* training, maar in willkeurige volgorde aangeboden, dus conform de uitwerking van het principe van het covariate leren. De beste generalisatie trad op na de *phonics* training en de minste na hele woord training; maar dit verschil kwam voor rekening van de langzame RAN-ers, wat zou wijzen in de richting van een algemeen automatiseringstekort bij althans een deel van de zwakke lezers. Van der Leij & Van Daal (1999) vergeleken zwakke vierdeklassers met gemiddelde tweedeklassers met eenzelfde leesniveau, en met gemiddelde vierdeklassers. Hun conclusie is dat dyslexie vooral een snelheidsprobleem is, maar zij vonden geen verschil bij het benoemen van letters en cijfers, op grond waarvan zij concluderen dat het automatiseringstekort niet een algemeen tekort is, maar gebonden aan *reading related material*. In zijn *instance* theorie maakt Logan (1988) een kwalitatief onderscheid tussen automatisch en niet-automatisch verlopende taakuitvoering. Niet-automatische taakuitvoering is gebaseerd op het volgen van een algoritme. Automatische taakuitvoering is gebaseerd op *single-step direct-access retrieval* uit het geheugen. Iedere gebeurtenis - *instance* - laat een nieuw geheugenspoor achter. Een nieuwe confrontatie met de stimulus levert een wedstrijd tussen algoritmische en automatische verwerking en op een bepaald moment gaat de automatische het winnen van de algoritmische. Vanaf dat moment is de vaardigheid geautomatiseerd. Binnen de *instance* theorie is automatisering niet *process-* maar *item-based* en automatisering betekent het leren van specifieke responsen bij specifieke stimuli. Onderzoek van Van Aarle (1988) toont aan dat het expliciet oefenen van decoderen met behulp van een algoritme niet zondermeer tot automaticiteit leidt. Combineren we deze noties met het *self-teaching* model van Share, dan kan het lezen van woorden via grafeem-foneem omzetting en auditieve synthese de algoritmische verwerking worden genoemd, en de directe woordherkenning het automatische ophalen van het woord uit het geheugen, wat ook de aard van de codering is.

generalisatie

Toepassing van het voorgaande zal ertoe leiden dat leerlingen het oefenmateriaal zullen kunnen *lezen*. Of ze als gevolg daarvan ook beter hebben *leren* lezen, is van een andere orde, maar ook afhankelijk van de vraag hoe beter leren geoperationaliseerd wordt. Als de meest gangbare maat voor leersucces geldt de mate van transfer of generalisatie. Deze lijkt vooral op twee manieren geoperationaliseerd te worden, in leertheoretisch termen aangeduid als stimulusgeneralisatie en responsgeneralisatie (Sulzer-Azaroff & Mayer 1991). Bij stimulusgeneralisatie komt een onveranderlijke respons onder controle van nieuwe stimuli, vergelijk de generalisatiefase in het *instructional hierarchy* model van Haring e.a. (1978), zoals beschreven in paragraaf 4.1. Daarvan is sprake als woorden die in een bepaalde oefencontext geleerd zijn,

bijvoorbeeld als losse woorden, ook vlot worden gelezen in een andere context, bijvoorbeeld een tekst waarin deze woorden voorkomen. Hiervan wordt als beproefd didactisch principe gebruik gemaakt door bijvoorbeeld uit een nieuwe tekst de moeilijke woorden vooraf te oefenen (Levy e.a. 1997).

Meestal wordt in onderzoek waarin generalisatie wordt nagestreefd, het vermogen bedoeld om juist niet eerder geoefende woorden toch te kunnen lezen, ofwel respons-generalisatie in leertheoretische termen. Bij responsgeneralisatie zien we dat gedeeltelijk nieuw gedrag mogelijk wordt, als variatie op en aanpassing van gedrag dat reeds in het repertoire aanwezig is. Een gedeeltelijk nieuwe respons is gebaseerd op een gedeeltelijk nieuwe combinatie van elementen van beschikbare responsen, zodanig dat de nieuwe respons is afgestemd op de eisen die de nieuwe situatie stelt. De respons verandert en het gedragsrepertoire breidt uit, vergelijk de adaptatiefase in het model van Haring e.a.. Onvoldoende responsgeneralisatie is een belangrijk kenmerk van een niet goed verlopende (lees)ontwikkeling.

In onderzoek blijkt responsgeneralisatie moeilijk aan te tonen. Van der Leij (1998) is van mening dat training die gericht is op het hele woord een positiever effect heeft op het leren van de geoefende woorden (stimulusgeneralisatie), terwijl training gericht op sublexicale eenheden leidt tot meer (respons)generalisatie naar niet geoefende woorden. In de standaard opvatting van automatisering gaat het om het steeds efficiënter uitvoeren van eenzelfde proces, dat ook kan worden toegepast op nieuwe, nog onbekende stimuli. Maar binnen de *instance* theorie is automatisering niet *process* maar *item based* en betekent automatisering het leren van specifieke responsen bij specifieke stimuli. Er is dan noodzakelijkerwijze nauwelijks sprake van generalisatie tussen items. De vraag is echter wat bij het leren lezen de relevante *items* zijn; zijn dat de afzonderlijke woorden of eenheden op subwoordniveau? Als subwoordeenheden de relevante items zijn, dan moeten die weliswaar afzonderlijk geleerd worden, maar is (respons)generalisatie tussen woorden weer wel mogelijk als die op subwoordniveau overlap vertonen, de zogenaamde *friends*.

Volgens Hogaboam & Perfetti (1978) wordt woordspecifieke informatie verworven tijdens training in het benoemen van woorden. Of de woorden wel of geen betekenis hadden maakte geen verschil. Experimenten van Logan (1988) waarin steeds nieuwe woorden en pseudowoorden werden getraind, leverden geen generalisatie op in de zin van een steeds snellere performance. Maar voortgezette training met dezelfde woorden en pseudowoorden na bereiken van accuratesse liet wel een generalisatie-effect zien, waarbij het geen verschil maakte of gewerkt was met woorden of met pseudowoorden. Als automatiseren *process based* is, zou zowel bij standaard training als bij voortgezette training generalisatie moeten optreden. Van den Bosch e.a. (1993) zijn van mening dat training met een gesloten verzameling echte woorden niet tot generalisatie leidt. Er is wel - enige - generalisatie als gewerkt wordt met pseudowoorden in flitsaanbieding, omdat dit dwingt tot fonologisch decoderen en voorkomt dat leerlingen de woorden leren op basis van orthografisch geheugen. Leerlingen van 7-13 jaar werden in een onderzoek van Lovett e.a. (2000) verdeeld over drie verschillende interventies: *phonics* in

combinatie met *direct instruction*, *metacognitive phonics*, dat wil zeggen het aanleren van strategieën op subwoordniveau (vergelijk Van Aarle 1988) en een controle conditie waarin de standaard aanpak van de klas geïntensiveerd werd. Alle leerlingen gingen in alle condities vooruit. Maar de *metacognitive phonics* training gaf de meeste transfer naar woorden, *phonics* in combinatie met *direct instruction* naar pseudowoorden.

Het lijkt erop dat responsgeneralisatie alleen, en dan nog maar in beperkte mate, aangetoond kan worden als die gemeten wordt binnen taken die noodzakelijk via decodering op subwoordniveau worden uitgevoerd.

4.3 De deelonderzoeken

4.3.1 Inleiding

De deelonderzoeken hadden betrekking op de woordtraining. Ze waren gericht op de vraag in welke mate leerlingen van elkaar verschillen in de mate waarin en de wijze waarop zij op de interventie reageerden en op de vraag of deze verschillen het resultaat van de interventie voorspellen. In geval leerlingen onvoldoende van de interventie profiteren houdt dit een sterke aanwijzing voor dyslexie in. Training in kader van het voortraject voor het opsporen van dyslexie zoekt de grens op van wat onder optimale omstandigheden binnen het reguliere onderwijs gerealiseerd kan worden. Kenmerken daarvan zijn in ieder geval een sterke uitbreiding van de leertijd, het optimaliseren van de antecedente gedragscontrole zodat juiste responsen - succes ervaringen - zo vaak mogelijk optreden en het zorgen voor volledige trials. Van der Leij (1983) introduceerde als maat voor leerbaarheid de term differentiële instructiegevoeligheid, de leertijd die een leerling nodig heeft om zich bepaalde leerstof eigen te maken. De uitwerking van dit begrip was globaal, bijvoorbeeld de leerling heeft een schooljaar nodig om een AVI-niveau te vorderen. Een geringe instructiegevoeligheid betekent meer leertijd waar andere leerlingen met minder toekunnen. Vellutino e.a. (1996) beschreven het begrip *response to treatment* in vergelijkbare termen. Deze werd bijvoorbeeld uitgedrukt als vooruitgang in leesvaardigheid van een jaar na een half jaar remedial teaching.

Op microniveau zou het begrip instructiegevoeligheid kunnen worden gebruikt om individuele verschillen in responsegeneralisatie aan te duiden, niet noodzakelijk als een toenemend vermogen om niet eerder geoefende woorden te kunnen *lezen*, maar als een toenemend vermogen om niet eerder geoefende woorden te *leren* lezen. Dit kan gezien worden als een primaire maat voor - de kwaliteit van - het leervermogen en als een specificatie van responsgeneralisatie. Er wordt niet de eis gesteld dat ook woorden die nog niet eerder gezien zijn, ook zonder oefening goed en snel gelezen kunnen worden, maar wel dat ze in minder trials geleerd worden. In het begrip instructiegevoeligheid wordt in deze zin gebruikt als equivalent van *response to treatment*.

4.3.2 Principes van de woordtraining

Een volwaardig remediërend programma oefent niet alleen de toepassing van de leestechiek in het lezen van teksten, maar ook de deelvaardigheid van het correct en vlot herkennen van woorden. Voor (ortho)didactische toepassingen wordt daarbij gewezen op de mogelijkheden van de computer (National Reading Panel 2000), omdat die het 'perfecte model' en consequente feedback kan leveren, zich bij uitstek leent voor het aanbieden van veel herhalingen waarbij, als sprake is van een interactief programma, de moeilijkheidsgraad van de leerstof op het vaardigheidsniveau van de individuele leerling kan worden afgestemd. Daarbij kan een computer grote aantallen responstijden nauwkeurig registreren.

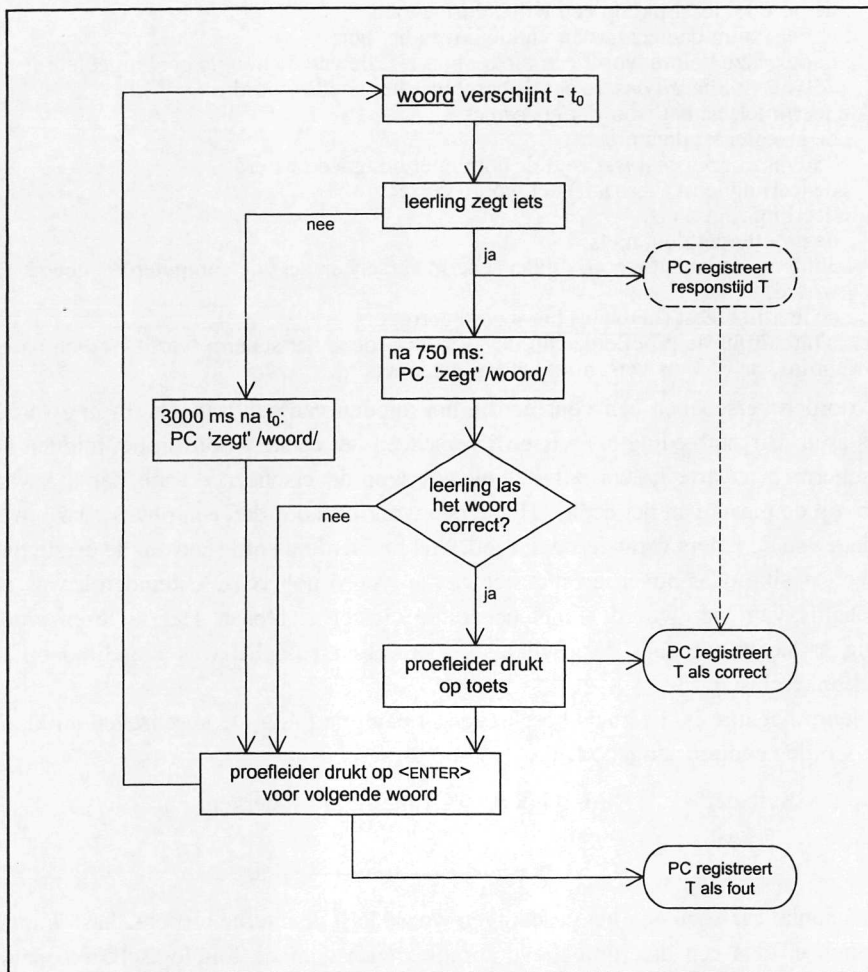
Aan de vorige paragrafen ontleen we de belangrijkste ingrediënten voor een computergestuurde woordtraining. Deze moet gericht zijn op één component, de vlotheid van de woordherkenning, zonder dat dit ten koste gaat van de accuratesse. De basis van de woordtraining in elk van de deelonderzoeken was dat elke sessie getraind werd met een gesloten, maar wel steeds nieuwe set woorden. Verder werden alle woorden, dus ook die reeds goed gelezen werden, herhaald aangeboden.

De moeilijkheidsgraad van een leestaak wordt bepaald door de combinatie van het niveau van de te lezen woorden, de omvang van de set van woorden waarmee getraind wordt en de aanbiedingsduur. De moeilijkheidsgraad dient zo gemanipuleerd te worden dat het gewenste gedrag, vlotte, correcte directe woordherkenning volgens de gedragstherapeutische vuistregel in meer dan 80% van de oefenitems mogelijk is. Gebruikmaking van beperkte aanbiedingsduur zorgt er tevens voor dat spellend lezen wordt tegengegaan. Als alternatief voor een training die alleen succesvol is als er sprake is van voldoende responsgeneralisatie, richten we deze in met in principe alle woorden, maar zodanig dat de kans op generalisatie maximaal is. Kenmerken zijn:

- veel herhalingen en voortzetting van de training nadat accuratesse bereikt is;
- leerlingen gedragsinstructie te geven voor directe woordbenoeming om tegen een mogelijk reeds ontstane gewoonte van spellend lezen in te gaan.
- werken met klankzuivere woorden, dus ook geen woorden met een eind-b, eind-d of eindigend op -eer, -oor, -eur;
- het manipuleren van de moeilijkheidsgraad van de training door:
 - . met woorden van verschillende structuur te werken (MKM, MMKM/MKMM);
 - . verschillende mate van overlap van de trainingwoorden te hanteren, bij woorden met een medeklinkercluster variërend van KMM/MMK, via MM naar geen overlap;
 - . het gebruik van wisselrijaanbieding als orthodidactische optie;
 - . met verschillende taken te werken (benoemen of matchen);
 - . geleidelijke verhoging van de moeilijkheidsgraad door flitsaanbieding;
- volledige *learning trials*.

Het trainen van alle woorden lijkt misschien een grotere taak dan het in werkelijkheid is. Er zijn nog geen 700 klankzuivere zogenaamde MKM-woorden, waarvan een kleine 600 drie of meer *friends* hebben waarmee ze de KM-woordkern delen. Van de ruim 400 klankzuivere MKMM-woorden hebben er zo'n 250 drie of meer *friends* waarmee ze de KMM-woordkern delen. Van de bijna 500 klankzuivere MMKM-woorden ten slotte hebben er ±180 drie of meer *friends* waarmee ze de MMK-woordkern delen. Bijna alle

MKMM- en MMKM-woorden hebben hun dubbele medeklinker gemeen met drie of meer andere woorden.



Figuur 2: Stroomschema van de woordtraining in deelonderzoek 1 en 2.

4.3.3 Uitwerking

Er is gebruik gemaakt van eenvoudige notebook computers met ingebouwde geluidskaart, luidsprekers en microfoon. De basisprocedure, schematisch weergegeven in figuur 2, was als volgt:

- midden op het scherm verschijnt een woord, bijvoorbeeld stok;
- linksonder in beeld staat een grijze kolom;
- de leerling reageert (of niet);

markering voor het criterium. Als een trial is afgerond, wordt deze herhaald. Een volgende grijze kolom verschijnt naast de vorige. De kolommen blijven tijdens de sessie in beeld.

Tabel 12: Voorbeeld van een sessierapport van de woordtraining.

Resultaten van sessie nummer	13
Leerlingcode: 9103129won000	
Leerlingnummer: 8	
Datum; tijd: 12-11-1998; 14:38	
Max. reactietijd: 3000	
Criterium: 10 woorden	
Conditie: RB	
Maximum aantal trials:	12
Aantal trials tot criterium :	9
slaaf	1 1649 0 2 2020 1 3 1274 1 4 148 1 5 588 1 6 2333 1 7 1649 1 8 213 1 9 3019 1
sleep	1 3022 0 2 1088 1 3 2023 1 4 1651 0 5 92 1 6 401 1 7 1837 1 8 1459 1 9 2146 1
slijt	1 2648 1 2 2398 1 3 2862 1 4 1522 1 5 1275 1 6 1276 1 7 2150 1 8 277 1 9 2024 1
slop	1 1021 1 2 1361 1 3 1275 0 4 1524 0 5 1895 1 6 1337 1 7 85 1 8 1525 1 9 272 0
bloos	1 1709 1 2 2023 1 3 1833 1 4 1275 1 5 1150 1 6 1837 1 7 1708 1 8 776 1 9 1709 1
blijk	1 2456 0 2 2524 0 3 2712 1 4 1649 1 5 1150 1 6 1961 0 7 1586 1 8 1774 1 9 1460 0
bloes	1 3040 0 2 2831 0 3 963 1 4 151 1 5 2082 1 6 963 1 7 1272 0 8 1524 0 9 1587 1
blok	1 1982 1 2 1710 1 3 2523 1 4 1271 1 5 523 1 6 1645 1 7 1151 1 8 339 1 9 1649 1
gries	1 340 0 2 1650 1 3 1837 0 4 2145 0 5 1336 0 6 1026 0 7 1523 1 8 960 0 9 526 1
groet	1 3019 0 2 2523 1 3 2955 1 4 2148 1 5 526 1 6 958 1 7 897 1 8 1647 1 9 1770 1
graf	1 2335 1 2 1957 1 3 1334 1 4 1396 1 5 336 1 6 1458 1 7 152 1 8 1770 1 9 2212 1
grit	1 3042 0 2 1650 0 3 3041 0 4 1459 1 5 1276 1 6 1272 0 7 1213 1 8 1274 1 9 1649 1

Legenda: Woord, trialnummer, tijd, score, trialnummer, tijd, score, etc.

Het gestelde criterium voor het halen van een sessie in deelonderzoek 1 was laag: op drie trials moesten minimaal tien van de twaalf woorden binnen de tijd goed gelezen zijn. Het gevolg was dat sommige leerlingen weinig training kregen. Wel was het zo dat drie achtereenvolgende trials gehaald moesten zijn. Het verliezen van reeds gehaalde trials werd door de meeste leerlingen als straffend ervaren en was derhalve niet bevorderlijk voor de motivatie. In de deelonderzoeken 2, 3 en 4 is het criterium voor het halen van een sessie daarom aangepast. In plaats van drie moesten zes trials gehaald zijn, maar niet langer noodzakelijk achtereenvolgend. Het maximaal aantal trials is gesteld op twaalf. Als de sessie wordt afgesloten, omdat het criterium gehaald is of omdat er 12 trials zijn geweest, klinkt een grappig geluidsfragment ten teken dat de sessie klaar is en als beloning. Er waren 24 verschillende geluiden waarvan er steeds één willkeurig geselecteerd werd.

De grens van 3000ms is gekozen op grond van de volgende overwegingen. De training betreft het lezen van losse woorden. Volgens de normtabel van de DMT-kaart 2 lezen de beste D-leerlingen met een dl 11, de didactische leeftijd van de leerlingen aan het begin van de training, ongeveer 20 woorden goed per minuut. Drie maanden later, aan het einde van de trainingperiode, lezen de beste E-leerlingen ongeveer 20 woorden goed per minuut. Voorts kan uit onderzoek van Van Daal (1993) worden afgeleid dat het volledig spellend lezen mogelijk wordt vanaf ongeveer 4000 ms. De procedure is

dus straffend voor spellend lezen, omdat de PC dan het goede woord zegt voordat de leerling klaar is, en belonend voor directe woordherkenning. Per sessie maakte de computer een rapport, voor een voorbeeld zie tabel 12.

4.3.4 Opzet van de deelonderzoeken

Het woordtrainingprogramma zijn vier deelonderzoeken uitgevoerd. Deelonderzoek 1 betrof leerlingen in de eerste helft van groep 4, didactische leeftijd tussen 11 en 16 maanden. Er is gewerkt met woordensets van niveau 2, dat wil zeggen MKMM en MMKM, die in beperkte mate *friends* waren, ze overlaptten in het medeklinkercluster. In het experiment zijn tevens enkele condities nader onderzocht: de positie van het medeklinkercluster, de aanbiedingsvorm als wisselrij of in willekeurige volgorde, de samenhang tussen accuratesse en vlotheid en de relatie tussen *response to treatment* of differentiële instructiegevoeligheid en de technische leesvaardigheid na beëindiging van de training.

Deelonderzoek 2 is uitgevoerd met jongere leerlingen aan het eind van groep 3, didactische leeftijd tussen 7 en 10 maanden, met woorden van niveau 1, dat wil zeggen MKM. De woorden overlaptten in KM-woordkern. Dezelfde condities zijn bekeken; voor het vaststellen van responsgeneralisatie zijn de te trainen woordensets bij wijze van voortoets apart afgenomen en gescoord.

Aan deelonderzoek 3 namen de leerlingen deel die in het remedial teaching onderzoek dat in hoofdstuk 3 is beschreven de experimentele groep T11 vormden. Het betrof evenals in deelonderzoek 1 leerlingen in de eerste helft van groep 4, didactische leeftijd tussen 11 en 16 maanden. Er is gewerkt met woordensets van niveau 2, dat wil zeggen MKMM en MMKM, waarbij de mate van overlap is gevarieerd van medeklinkercluster plus klinker, via alleen medeklinkercluster, naar geen overlap tussen de trainingwoorden van een sessie. Voorts is de aanbiedingstijd variabel gemaakt, teneinde de moeilijkheidsgraad optimaal af te stemmen op het beheersingsniveau van de leerling. In deelonderzoek 3 zijn met andere woorden de kenmerken van de het stimulusmateriaal verder gevarieerd.

Deelonderzoek 4 ten slotte betrof zeer zwakke lezers in de tweede helft groep 4 met een didactische leeftijd tussen 16 en 20 maanden. Voor de training met woorden op niveau 2 is gebruik gemaakt van een matchingtaak van eenheden op subwoord- en woordniveau.

afhankelijke variabelen

De vlotheid van lezen als afhankelijke variabele is geoperationaliseerd in de gemiddelde responstijd per sessie op de goed gelezen woorden. Responstijden van minder dan 900 ms kwamen zeer weinig voor, met de laagste frequentie rond de 600 ms. Responstijden tot 600 ms zijn geïnterpreteerd als ongeldige registraties, bijvoorbeeld van toevallig geluid, en niet in de analyses betrokken.

Het aantal trials verschilde per proefpersoon en per sessie, maar was minimaal drie in deelonderzoek 1 en minimaal zes in deelonderzoek 2, 3 en 4. Daarom is als afhankelijkke variabele gekozen de gemiddelde responstijd per sessie van de eerste drie respectievelijk zes trials van de items die zowel *geldig* als *goed* waren: Tggd. In het voorbeeld van tabel 12 waren er in de eerste drie trials 23 responstijden die zowel geldig als goed waren met een gemiddelde tijd van 1952 ms. Leerling 8 haalde zo voor sessie 13 een score van 1952.

Op dezelfde wijze werd in deelonderzoek 1 als accuratessemaat het gemiddelde aantal goed gelezen woorden berekend, in het voorbeeld van tabel 12 was dit 7,7. In de deelonderzoeken 2 en 3 zijn per sessie de te oefenen woorden bij wijze van voortoets eenmaal gelezen en is hiervan de accuratesse geregistreerd. In deelonderzoek 4 ten slotte zijn per sessie de te oefenen woorden als voor- en natoets afgenomen en zijn hiervan zowel de accuratesse als de responstijden geregistreerd. In deelonderzoek 4 is tevens parallel aan de training viermaal een woordleestoets afgenomen waarvan responstijden en accuratesse zijn gescoord.

voorspellen op basis van individuele leercurve

In hoofdstuk 2 is gewezen op het voordeel van interventie op het moment dat de eerste problemen in het doelgedrag reeds naar voren komen. De selectie maat voor het vaststellen welke leerlingen voor de interventie in aanmerking komen, deelt dan zoveel mogelijk variantie met de criteriummaat. In paragraaf 4.2 is voorgesteld om op micro-niveau het begrip instructiegevoeligheid, als equivalent van *response to treatment*, te gebruiken om individuele verschillen in responsegeneralisatie aan te duiden, als een toenemend vermogen om niet eerder geoefende woorden te *leren* lezen. Een individuele leercurve op een gecontroleerd traject dat inhoudelijk sterk overeenkomt met het criterium, houdt daarom de belofte in van een goede predictie voor toekomstig leren. De verwachting is dat de respons van de leerling op de training differentieert tussen de leerlingen die aan het einde van het remediërend traject niet meer of nog steeds tot de uitvallers gerekend moeten worden.

Verschillende auteurs hebben procedures beschreven om naast analyses op basis van vergelijking van groepsgemiddelden, inzicht te krijgen in individuele verschillen door in herhaalde metingen groeicurves te bepalen met behulp van zogenaamde *multilevel* methoden (Francis e.a. 1991, Hoeksma & Koomen 1992, Goldstein e.a. 1994, Van den Boom & Hoeksma 1995, Francis e.a. 1996, Foorman e.a. 1997, Crijnen e.a. 1998, Hoeksma & Knol 2001).

De woordtrainingsexperimenten leverden een groot aantal herhaalde metingen op van responstijden en accuratesse. Deze zijn als metingen op niveau 1 genest binnen het individu, niveau 2. Met de techniek worden polynome functies opgebouwd tot dat een maximale *fit* is bereikt. Experimentele condities worden in het model als zogenaamde *dummy* variabelen opgenomen, waarvan op groepsniveau het effect en vervolgens eventuele interactie-effecten kunnen worden bepaald. Ook kunnen zo subgroepen worden onderscheiden die mogelijk een verschillend ontwikkelingsverloop laten zien.

Ten slotte kan worden berekend of er betekenisvolle individuele afwijkingen zijn in groeicurve ten opzichte van het groepsgemiddelde.

Met de *multilevel* techniek op basis van het programma MLwiN (Rasbash e.a. 2000) werden in de vier deelonderzoeken de functies in de beschreven volgorde opgebouwd en zo het groeps- en individuele verloop geschat van de responstijden en maten voor de accuratesse als operationalisering van instructiegevoeligheid.

4.3.5 Deelonderzoek 1

specificatie van de onderzoeksvraag

In deelonderzoek 1 is de invloed nagegaan van wel of geen overlap van het medeklinkercluster tussen training- en generalisatiewoorden op de generalisatie naar nog niet geoefende woorden en van de wisselrij als aanbiedingsconditie. Voorts is de differentiële instructiegevoeligheid onderzocht in relatie tot de leesvaardigheid van de leerlingen aan het einde van de trainingperiode. Leerlingen worden tot de uitvallers gerekend als zij op de DMT een E-score behalen, dat wil zeggen dat zij behoren tot de 10% zwakste lezers volgens de landelijke norm. Het nog wel of juist niet meer behoren tot de groep E-leerlingen is gekozen als uitkomstmaat.

proefpersonen

Op acht middelgrote scholen is aan alle leerlingen met een didactische leeftijd van 9 de DMT-2 afgenomen als onderdeel van het leerlingvolgsysteem. Hieruit is het zwakste eenderde deel van de leerlingen geselecteerd voor toetsing begin september. Bij een score op de DMT boven D-niveau werd het onderzoek gestaakt. Aan de leerlingen met een D- of E-score werd een uitgebreide serie toetsen afgenomen, waaronder de AVI-kaarten 1 en 2. Slechts twee leerlingen combineerden een DMT-score op E-niveau met een AVI-beheersingsniveau 2. Vijftien leerlingen lazen AVI-kaart 1 op beheersingsniveau, de overige haalden kaart 1 nog niet. Na uitsluiting van leerlingen van wie bekend was dat ze binnen de onderzoeksperiode zouden gaan verhuizen of naar een school voor speciaal onderwijs zouden gaan zijn de 46 zwakste lezers voor het onderzoek geselecteerd op basis van de DMT-score in september (tabel 13). Als aanvullend selectie criterium gold een AVI-beheersingsniveau van maximaal 1; beheersing van AVI-1 bij een didactische leeftijd van 11 maanden is te interpreteren als D-niveau.

Tabel 13: Beschrijvende informatie over de proefpersonen (N=46) van deelonderzoek 1.

Sekse totaal	Jongens 20		Meisjes 26	
SE-score totaal	1 17	1,25 21	1,9 8	
	Min.	Max.	M (SD)	
DMT-2, ruwe score	2	12	8,0 (3,3)	

procedure

Deze 46 leerlingen ontvingen in de periode september tot februari groep 4 de woordtraining. Deze begon met een oefensessie, hierin werd de werking van het programma uitgelegd. Er werd geoefend aan de hand van een set van twaalf woorden. In de training werden steeds twee sessies achterelkaar gedaan. Dit duurde maximaal een half uur. Er waren twee trainingen per week op verschillende dagen. In totaal ontvingen leerlingen 36 sessies. Hiervoor waren twaalf effectieve schoolweken uitgetrokken met het oog op uitval van trainingdagen wegens afwezigheid van de leerling.

inhoud

Er zijn 18 (trainingdagen) * 3 (aanbiedingscondities) sets van 12 woorden samengesteld. Deze 648 woorden bestonden allen uit vier grafemen, de helft met een medeklinkercluster aan het begin (B-conditie), de andere helft aan het eind (E-conditie). De 12 woorden van een set hadden steeds alleen een begin- of een eindcluster. Een set bestond uit 3 subsets van 4 woorden met hetzelfde cluster. Twee sets hoorden steeds bij elkaar in die zin dat ze dezelfde drie clusters hadden, zie tabel 14 als voorbeeld.

De woorden waren bestaande klankzuivere woorden, dus geen woorden met een eind-d of eind-b en geen klinker-r combinaties als -eer, -oor, -eur. Elke subset had zoveel mogelijk een of twee woorden met een korte klinker, een of twee woorden met een lange klinker en een woord met een tweetekenklank. Er waren voor negen dubbelsessies 324 woorden nodig in negen maal negen subsets van vier woorden die aan al deze eisen voldeden. Vooral voor de woorden met een eindcluster was dit aantal moeilijk te halen. Niet alle woorden waren dientengevolge de zeven- en achtjarige proefpersonen even gewoon.

Tabel 14: Voorbeeld van drie sets trainingwoorden uit deelonderzoek 1.

set 1	set 2	set 3
T (training)	G (generalisatie)	C (controle)
trek	trap	kroop
trog	trim	kreek
truus	trein	krak
traan	troep	kruim
grijp	grif	smal
grot	gras	smuk
greep	griep	smaak
graaf	groef	smeet
sluis	slap	wrat
slik	slot	wraak
slaan	slijk	wrijf
sloot	sluit	wroet

trainingcondities

Er werden drie verschillende trainingcondities gerealiseerd. In de eerste trainingconditie werden de woorden van set 1 aangeboden in de vorm van een wisselrij. Deze conditie werd aangeduid met de letter W. In de tweede conditie werden de woorden van

set 1 in een willekeurige volgorde een voor een aangeboden. Deze *random*conditie werd aangeduid met R. De derde conditie was een controleconditie C. Hierin werden in plaats van de woorden van set 1, de woorden van set 3 als eerste getraind en in een willekeurige volgorde aangeboden zoals in de *random*conditie. Direct daarna volgde in alle gevallen een sessie met de woorden van set 2. Deze werden uitsluitend een voor een in een *random* volgorde aangeboden. Als er sprake zou zijn van generalisatie, dan zou die bij set 2 moeten blijken. Set 2 werd daarom aangeduid met G.

Om te beginnen werden de 12 woorden van de sessie eenmaal geoefend. De woorden verschenen op het scherm. De leerlingen hoefden alleen maar te kijken, hoewel ze toch, en zeker naarmate de training vorderde, begonnen te lezen. Na 3000 ms zei de computer het woord, waarna de leerling het woord moest lezen (herhalen). De wijze van aanbieden bij het oefenen, wisselrij of een voor een in willekeurige volgorde, kwam overeen met die van de trainingconditie die volgde, zodat de leerling wat dat betreft wist wat er komen zou.

volgorde

De ene helft van alle trainingsets bestond uit woorden met een medeklinkercluster aan het begin, de ander helft met een medeklinkercluster aan het eind. Deze werden om en om aangeboden. Met iedere helft van 324 woorden waren 9 x 3 vaste sets gemaakt volgens de hiervoor beschreven uitgangspunten. Hiermee zijn met behulp van een generator van *random*getallen zonder teruglegging 46 verschillende aanbiedingsvolgordes bepaald en willekeurig aan de proefpersonen toegewezen. Vervolgens is aan elke leerling een conditievolverde toegewezen. Met de drie condities zijn zes verschillende volgordes te maken: WRC, RCW, CWR, WCR, RWC en CRW.

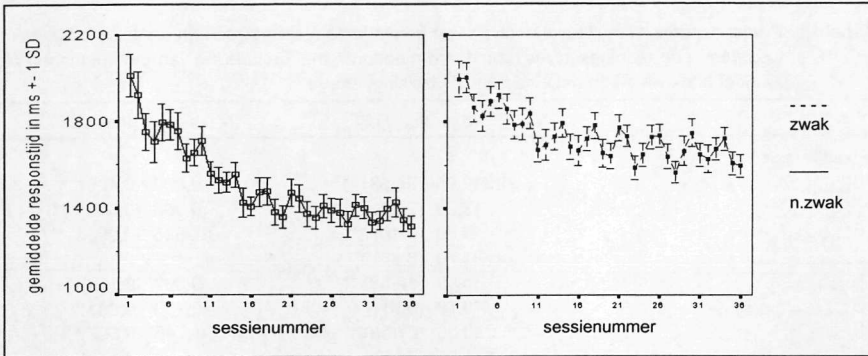
bepaling van trainingseffecten

Zowel de vlotheid van lezen (Tggd) als de accuratesse (Nggd) zijn in de analyse opgenomen. Voor het bepalen van verschillen in trainingseffecten werden de resultaten van de leerlingen in de trainingconditie (set 1 of set 3) vergeleken met de resultaten op de woorden van de generalisatieconditie G (set 2). Een specifieke set van woorden is in de generalisatieconditie door de verschillende randomisering en even vaak op de verschillende momenten in de training aan de orde gekomen (het sessienummer) en tevens even vaak na de verschillende trainingcondities: W, R of C. De generalisatieconditie werd gecodeerd op grond van de conditie die eraan vooraf ging, dus: WG, RG en CG. Met een letter B of E werd aangegeven of het om woorden met een begin- dan wel een eindcluster ging: WE, WEG, WB, WBG, RE enz.

resultaten

De gemiddelde responstijden per sessie van de items die zowel geldig als goed waren, Tggd, waren de maat voor de vlotheid van lezen. Binnen de proefpersonen is onderscheid gemaakt op basis van hun DMT2-score na afloop van de training. 29 leerlingen scoorden aan het einde van de trainingperiode nog steeds op E-niveau scoorden, zij zijn

aangeduid als 'zwak'. De 17 relatief minder zwakke leerlingen die aan het einde van de trainingperiode op D-niveau of hoger scoorden zijn aangeduid als 'niet-zwak'. Gemiddelde en spreiding van de responstijden zijn in figuur 3 voor deze twee onderscheiden groepen apart weergegeven.



Figuur 3: Deelonderzoek 1, gemiddelde en spreiding van de gemiddelde responstijden per sessie met de indeling van proefpersonen in E-niveau (zwak) en D-niveau of hoger (n.zwak) aan het einde van de trainingperiode.

Op de gemiddelde responstijden per sessie is een multilevel analyse uitgevoerd met herhaalde meting. De trainingssessie met *random* aanbieding is gedefinieerd als de referentieconditie. Andere condities worden daarmee vervolgens vergeleken. De middelste sessie is ingesteld als sessienummer nul. De rationale daarvoor is dat de prestatie in de middelste sessie min of meer de gemiddelde responstijd van iedere leerling weergeeft. De regressievergelijking is weergegeven in tabel 15. Hierin zijn de significante componenten en interacties opgenomen; deze is als volgt te lezen. Het verloop van de responstijden wordt het best beschreven met een tweedegraads functie. De gemiddelde voorspelde waarde voor de responstijd - de intercept - in de referentieconditie is halverwege de training 1503 ms. Gemiddeld neemt de responstijd elke sessie met 13 ms af. Echter ook de kwadratische component van het sessienummer van 0,5 ms verbetert het model significant. Naarmate leerlingen verder kwamen in de training, wordt de afname van de gemiddelde responstijd per sessie minder.

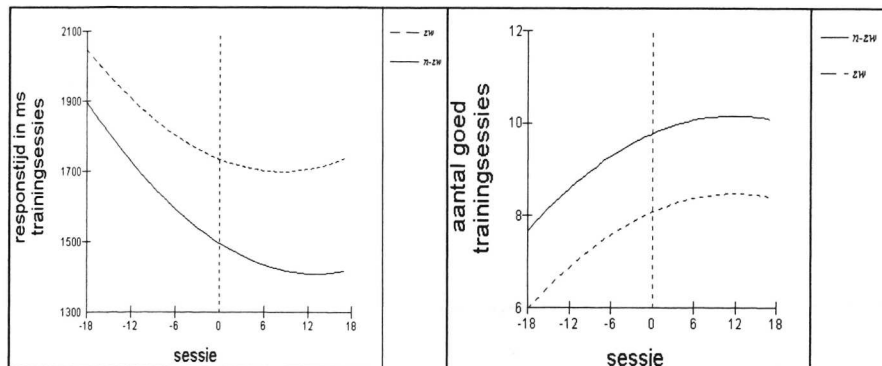
In de training zijn verschillende aspecten systematisch gevarieerd: de positie van het cluster aan het begin of aan het einde van het woord ('eindcluster'), aanbieding van de trainingset met (R) een woord tegelijk of (W) als wisselrij ('wisselrij') en wel of niet hetzelfde medeklinkercluster in training- en generalisatiesessie ('generalisatie w/r-conditie'). Woorden met een eindcluster blijken gemakkelijker dan woorden met een begincluster, gemiddeld werden woorden met een eindcluster 92 ms sneller gelezen. Er was een gering maar significant hoofdeffect van het trainen in de wisselrijconditie van 38 ms. Woorden aangeboden in de vorm van een wisselrij werden gemiddeld iets sneller gelezen dan losse woorden in een willekeurige volgorde. Ten slotte was er sprake

van een marginaal significant extra generalisatie-effect van 17 ms als de woorden in de tweede sessie van dezelfde trainingsdag dezelfde medeklinkerclusters hadden als de trainingwoorden. Er was geen verschil wat betreft generalisatie tussen de twee trainingcondities W en R. Introductie van dit effect leidde niet tot een betere 'fit' van het model en het effect is er daarom niet in opgenomen.

Tabel 15: Deelonderzoek 1, multilevel analyse van gemiddelde responstijden en aantal goed gelezen woorden in de trainingssessies (standaard meetfout), met de indeling van proefpersonen als zwak of niet-zwak aan het einde van de trainingperiode.

	responstijd (SE)	aantal goed (SE)
Fixed effecten		
$\beta_{\text{intercept}}$	1.502,66 (54,53) ***	9,809 (0,420) ***
$\beta_{\text{sessienummer}}$	-13,23 (2,11) ***	0,068 (0,008) ***
$\beta_{\text{sessie-kwadraat}}$	0,50 (0,10) ***	-0,0026 (0,0004) ***
$\beta_{\text{eindcluster}}$	-91,70 (8,88) ***	
$\beta_{\text{training wisselrij}}$	-38,19 (12,58) ***	0,397 (0,103)
$\beta_{\text{generalisatie w/r-conditie}}$	-16,79 (9,91) '	0,160 (0,081)*
β_{zwak}	238,39 (67,58) ***	-1,746 (0,507)***
$\beta_{\text{zwak} \times \text{sessienr.}}$	4,84 (2,59) '	
Random (niveau 2)		
$\text{var}[r_{\text{intercept}}]$	49.476 (10.738)	3,216 (0,682)
$\text{var}[r_{\text{sessienummer}}]$	72,46 (16,89)	0,0025 (0,0006)
$\text{var}[r_{\text{sessie-kwadraat}}]$	0,34 (0,09)	
$\text{cov}[r_{\text{int.}, \text{sessienr.}}]$		0,0397 (0,0156)
Residu (niveau 1)		
$\text{var}[r_{\text{intercept}}]$	32.068 (1.172)	2,014 (0,112)
$\text{var}[r_{\text{sessienummer}}]$		0,0015 (0,0009)
$\text{cov}[r_{\text{int.}, \text{sessienr.}}]$		-0,015 (0,004)

*** $p < .001$ ** $p < .01$ * $p < .05$ ' $p < .05$ (1-zijdig)



Figuur 4: Deelonderzoek 1, gemiddelde responstijd en aantal goed op trainingssessies met de indeling van proefpersonen in E-niveau (zw) en D-niveau of hoger (n.zw) aan het einde van de trainingperiode.

Alle leerlingen behoorden aan het begin van de trainingperiode, didactische leeftijd 11 maanden, volgens de landelijke DMT-norm tot de zwakste 10% lezers. Aan het einde van de training, didactische leeftijd 16 maanden, scoorden 17 leerlingen scoorden op D-niveau of hoger, 29 leerlingen nog steeds op E-niveau. De relatief zwakste lezers, dat wil zeggen een DMT-2 score op E-niveau op didactische leeftijd 16 maanden, hadden gemiddeld 238 ms langer nodig om de trainingwoorden te lezen. Daarnaast was er een significante interactie van 'zwak' met 'sessienummer' van 4,8 ms. Het verschil tussen de zwakste en relatief minder zwakke lezers nam toe met het verloop van de training. Er waren geen andere significante interacties, bijvoorbeeld van 'zwak' met training in de vorm van wisselrij. We berekenden de voorspelde responstijd (Tggd) met de regressievergelijking met daarin de significante factoren. Het verloop illustreert de verschillende gemiddelde differentiële instructiegevoeligheid van de 'zwakke' en 'niet-zwakke' leerlingen (figuur 4).

Tabel 16: Deelonderzoek 1, multilevel analyse van gemiddelde responstijden (standaard meetfout) in de trainingssessies.

	responstijd (SE)
Fixed effecten	
$\beta_{\text{intercept}}$	1.652,89 (38,79) ***
$\beta_{\text{sessienummer}}$	-10,18 (1,39) ***
$\beta_{\text{sessie-kwadraat}}$	0,50 (0,10) ***
$\beta_{\text{eindcluster}}$	-91,75 (8,88) ***
$\beta_{\text{training wisselrij}}$	-37,88 (12,58) ***
$\beta_{\text{generalisatie w/r-conditie}}$	-16,59 (9,90) '
Random (niveau 2)	
$\text{var}[r_{\text{intercept}}]$	65.367 (14.051)
$\text{var}[r_{\text{sessienummer}}]$	81,05 (18,68)
$\text{var}[r_{\text{sessie-kwadraat}}]$	0,34 (0,09)
$\text{cov}[r_{\text{sessienummer, sessie-kwadraat}}]$	-2,37 (0,98)
Residu (niveau 1)	
$\text{var}[r_{\text{intercept}}]$	32.066 (1.170)

*** $p < .001$ ' $p < .05$ (1-zijdig)

Dezelfde analyse op het aantal goed gelezen woorden (Nggd) op de trainingssessies als maat voor accuratesse, levert een in hoofdzaak vergelijkbaar resultaat (tabel 15). We definiëren de trainingssessie met *random* aanbieding weer als referentieconditie en sessie 19 als nulpunt. Gemiddeld lezen de leerlingen in de referentieconditie halverwege de training 9,7 woorden van de 12 goed. Elke volgende sessie wordt gemiddeld 0,07 woord meer goed gelezen, maar er is ook een negatieve kwadratische component. De toename vakt dus af met het verloop van de training. De leerlingen die uiteindelijk de zwakke lezers zullen blijken te zijn, lezen gemiddeld 1,6 woord minder goed. Op de accuratessemaat is geen sprake van een significante negatieve interactie tussen 'zwak' en sessienummer. Ook op de accuratessemaat blijkt het lezen van woorden bij wissel-

ding iets gemakkelijker; gemiddeld wordt 0,4 woord meer goed gelezen. Er is wat betreft accuratesse geen verschil gemeten tussen woorden met een begin of een eind-cluster. Ten slotte wordt in de generalisatiecondities gemiddeld 0,16 woord meer goed gelezen dan in de controleconditie als deze het medeklinkercluster gemeen hadden met de trainingwoorden. Hierbij maakte het geen verschil of in de trainingconditie die eraan voorafging de woorden in *random* volgorde of als wisselrij werden aangeboden.

Een verschil in leesvaardigheid aan het einde van de trainingperiode lijkt samen te gaan met een verschil in respons op de woordtraining. Gezocht is naar een geobjectiverde maat hiervoor. Daartoe zijn uit de predictievergelijking met de Tggd als afhankelijke variabele de factor 'zwak' en de interactie van 'zwak' met 'sessienummer' verwijderd (tabel 16). Behalve de intercept zijn ook de lineaire en de kwadratische component - sessienummer en sessie-kwadraat - zogenaamde *random* factoren. Dat wil zeggen dat er naast een systematisch of *fixed* groepseffect ook betekenisvolle, significante verschillen zijn tussen individuele proefpersonen op deze factoren. De afwijking van individuele leerlingen van het groepsgemiddelde is per factor berekend. De onderlinge samenhang en de samenhang met de leesvaardigheid op didactische leeftijd 16 is weergegeven in correlatietabel 17.

De gemiddelde DMT-2 score aan het eind van de trainingperiode was 22,9 (SD 11,5). Er is een sterke negatieve samenhang tussen de gemiddelde responstijd van de leerlingen, de intercept, en deze DMT-eindscore. De leerlingen met de laagste gemiddelde responstijden zijn de leerlingen met de hoogste DMT-score aan het einde van de trainingperiode. Verder hangen ook de lineaire en de kwadratische component negatief samen, de leerlingen bij wie de responstijden op de trainingssessies het snelst dalen, zijn ook de leerlingen bij wie de afname van die daling het geringst is. Met andere woorden de individuele leercurves waaieren uit en de verschillen tussen leerlingen nemen in de loop van de training toe. De individuele waarden van deze variabelen zijn als predictorvariabelen opgenomen in een multiple regressieanalyse ($\beta_{\text{intercept}} = -.65$, $p < .001$; $\beta_{\text{lin}} = -.18$, n.s.; $\beta_{\text{kwad}} = -.15$, n.s.), gezamenlijk wordt 40% van de variantie van de afhankelijke variabele verklaard ($r = .63$, $F(3,42) = 9,20$, $p < .001$). Alleen de intercept levert een significante bijdrage aan de voorspelling van de afhankelijke variabele, de DMT-score aan het einde van de training.

Tabel 17: Deelonderzoek 1, correlatietabel van DMT-score aan het einde van de trainingperiode en de niveau-2 residuwaarden op de randomfactoren

Correlatiematrix	1	2	3	4
1. DMT 2 op dl 16	--	-.61***	-.09	.11
2. intercept		--	-.03	-.25
3. lineaire component			--	-.53***
4. kwadratische component				--

*** $p < .001$

conclusie

In de training met woorden van niveau 2 die nooit meer dan het medeklinkercluster

overlappen, is het gerealiseerde generalisatie-effect, hoewel statistisch significant, marginaal. Hetzelfde geldt voor het faciliterend effect van aanbieding van trainingwoorden in de vorm van een wisselrij. Leerlingen verschillen van elkaar op het aspect vlotheid van lezen in de wijze waarop ze op de training reageren, aanvankelijk verschillen tussen leerlingen nemen in de loop van de training toe. Wat betreft de accuratesse van het lezen zijn er wel niveauverschillen tussen leerlingen, maar hierin treedt gedurende de training geen verandering op. De differentiële instructiegevoeligheid op de training is gerelateerd aan de uitkomstmaat, maar het verband gaat niet verder dan dat de leerlingen die het op de training gemiddeld beter doen in de zin van een lagere gemiddelde responstijd, in hoofdzaak ook de leerlingen zijn die aan het eind van de periode boven E-niveau scoren.

4.3.6 Deelonderzoek 2

specificatie van de onderzoeksvraag

Ook in deelonderzoek 2 is de invloed nagegaan van de mate van overlap tussen training- en generalisatiewoorden op de generalisatie naar nog niet geoefende woorden en van de wisselrij als aanbiedingsconditie. Hierbij is in vergelijking met deelonderzoek 1 gebruik gemaakt van eenvoudiger MKM-woorden, welke een relatief sterkere overlap vertoonden, waarbij de overlap bovendien tevens de klinker omvatte (de KM-woordkern). Om de mogelijkheid van generalisatie via subwoordeenheden groter dan een enkel grafeem te onderzoeken, is in de generalisatieconditie gebruik gemaakt van pseudowoorden. Evenals in deelonderzoek 1 en op vergelijkbare wijze is de differentiële instructiegevoeligheid onderzocht in relatie tot de leesvaardigheid van de leerlingen aan het einde van de trainingperiode. Het nog wel of juist niet meer behoren tot de groep E-leerlingen is evenals in deelonderzoek 1 gekozen als uitkomstmaat.

Tabel 18: Beschrijvende informatie over de proefpersonen (N=35) van deelonderzoek 2.

Seks totaal	Jongens		Meisjes	
	20		15	
SE-score totaal	1	1,25	1,9	
	10	14	11	
	Min.	Max.	M	(SD)
DMT-1, ruwe score	1	9	6,1	(2,2)

proefpersonen

Op dezelfde scholen van onderzoek 1 is in december aan alle leerlingen met een didactische leeftijd van 4 maanden de DMT-1 afgenomen als onderdeel van het leerlingvolgsysteem. Hieruit zijn de leerlingen met een score op D-niveau of lager geselecteerd voor toetsing in februari. Bij een score op de DMT boven D-niveau werd het onderzoek gestaakt. Aan de leerlingen met een D- of E-score werd een uitgebreider serie toetsen

afgenomen, waaronder de AVI-kaart 1. Geen van de leerlingen combineerden een DMT-score op D- of E-niveau met een AVI-beheersingsniveau 1. Na uitsluiting van leerlingen van wie bekend was dat ze binnen de onderzoeksperiode zouden verhuizen zijn de 35 zwakste lezers met een DMT-score op E-niveau voor deelonderzoek 2 geselecteerd op basis van de DMT-score in februari (tabel 18).

procedure

De 35 leerlingen ontvingen in de periode februari tot juni groep 3 de woordtraining volgens dezelfde opzet als in onderzoek 1, maar met MKM-woorden, passend bij het stadium van ontwikkeling van de leesvaardigheid van deze zwakke lezers uit groep 3. Het criterium voor het halen van een sessie is ten opzichte van onderzoek 1 aangepast. Een sessie was gehaald als de leerling zes trials gehaald had, volgens het criterium van 80% goed gelezen woorden binnen de maximum tijd van 3000 ms. Ieder leerling heeft dus elke sessie minimaal zes trials gehad. Na maximaal twaalf trials werd de sessie afgesloten, of het criterium gehaald was of niet. De training betreft het lezen van losse woorden. Volgens de normtabel van de DMT-kaart 1 lezen de beste D-leerlingen met een dl 7, de didactische leeftijd van de leerlingen aan het begin van de training, ongeveer 20 woorden goed per minuut. Drie maanden later, aan het einde van de training-periode, lezen de beste E-leerlingen ongeveer 20 woorden goed per minuut. Op grond hiervan is de grens van 3000 ms gehandhaafd als de maximale tijd waarbinnen een woord gelezen moest zijn.

inhoud

Bij wijze van voortraining kregen alle leerlingen eerst 12 sessies met uitsluitend bestaande klankzuivere MKM-woorden. Daarna volgden 12 (trainingdagen) * 2 sets (trainingconditie en generalisatieconditie) van 12 woorden. De trainingset T bestond uit drie subsets van vier rijmwoorden, eindigend op dezelfde -KM combinatie of woordkern. De generalisatieset bestond uit twee subsets K en B van pseudowoorden, zie tabel 19 als voorbeeld.

Tabel 19: Voorbeeld van drie sets trainingwoorden uit deelonderzoek 2.

set 1	set 2	set 3
T	K	B
paar		
jaar	taar	daap
haar	laar	zaat
waar		
liep		
miep	wiep	biek
riep	ziep	hiem
diep		
bon		
kon	mon	jot
ton	hon	wop
zon		

De woorden uit de trainingset waren bestaande klankzuivere woorden, dus geen woorden met een eind-d, eind-b of een combinatie -eer, -oor of -eur. De serie bestond steeds uit een subset met een lange klinker, een subset met een korte klinker en een subset met een twee-tekenklank. De woordkernen eindigden op verschillende medeklinkers en de woorden begonnen met twaalf verschillende medeklinkers, die bovendien verschilden van de eindmedeklinkers. In de trainingset kwamen dus steeds drie klinkers en vijftien medeklinkers aan de orde. De generalisatieset bestond uit twee subsets van elk zes pseudowoorden, gevormd met dezelfde 15+3 grafemen. Subset-K had dezelfde woordkern, twee van elk. De beginmedeklinkers waren de beginmedeklinkers van de andere trainingssubsets dan die waarmee de woordkern overeen kwam, zoveel mogelijk een van elk van de twee andere. Subset-B was gevormd door een getrainde klinker te combineren met twee medeklinkers, zoveel mogelijk een van elk van de andere trainingssubsets. Deze woorden hadden dus niet dezelfde woordkern, maar bestonden wel uitsluitend uit in de voorgaande sessie getrainde letters. Voor de 12 trainingssessies waren 144 woorden nodig die aan al deze eisen voldeden en waarmee ook nog pseudowoorden te maken waren die klankzuiver waren en voldoende uitspreekbaar, dus niet eindigend op -h, -j, -v, -w, -z. Niet alle woorden waren dientengevolge even gewoon voor zes- en zevenjarigen.

trainingcondities

De trainingsets zijn in twee vormen aangeboden, als wisselrij - W - of een voor een in een willekeurige of *random* volgorde - R - op identieke wijze als in onderzoek 1. De wisselrijtjes waren uitgelijnd op de woordkern. Om te beginnen werden de twaalf woorden van de sessie eenmaal geoefend. De leerling moest proberen de woorden te lezen. De proefleider beoordeelde of het woord juist gelezen was. Na 3000 ms zei de computer het woord, hetzij ter bevestiging als het goed was, dan wel ter correctie. Ten slotte moest de leerling het woord (nog) eenmaal lezen (herhalen).

Het programma is voor deelonderzoek 2 zo aangepast dat ook de respons op de eerste oefenaanbieding van de woorden van een trainingset als goed of fout is geregistreerd. De wijze van aanbieden bij wijze van oefening kwam overeen met die van de training, wisselrij of *random*. De score op deze eerste aanbieding werd in het rapport geregistreerd bij wijze van voortoets. Ook de twaalf woorden van de twee generalisatiesets zijn in een *random* volgorde aangeboden. De subsets K en B werden wel afzonderlijk geregistreerd en ook het halen van een trial werd per subset geregistreerd. Een trial werd gehaald als minimaal vijf woorden goed en binnen de tijd gelezen waren. Als het criterium voor een subset gehaald was, werden die woorden niet meer aangeboden, maar werd alleen de andere subset afgemaakt.

volgorde

De totale verzameling bestond uit 144 trainingwoorden. Hiermee zijn 12 vaste sets gemaakt volgens de hiervoor beschreven uitgangspunten. 35 verschillende aanbiedingsvolgordes zijn willekeurig de proefpersonen toegewezen. Vervolgens is aan elke

leerling een conditievolvergader toegevoerd: W-R-W-R-enz. of R-W-R-W-enz. De generalisatieconditie werd gecodeerd op grond van de conditie die eraan vooraf ging, dus: GWK, GWB, GRK en GRB.

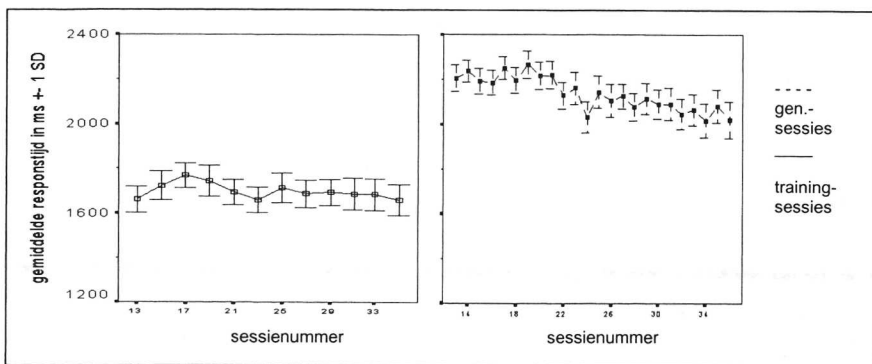
bepaling van trainingseffecten

Zowel de vlotheid van lezen (Tggd) als de accuratesse op de voortoetsen (VT) zijn in de analyse opgenomen. Voor het bepalen van verschillen in trainingseffecten werden de resultaten van de leerlingen in de trainingconditie vergeleken met de resultaten op de woorden van set 2 (de generalisatieconditie K) en set 3 (generalisatieconditie B).

Een specifieke set van woorden is door de verschillende randomisering even vaak op de verschillende momenten in de training aan de orde gekomen (het sessienummer) en tevens even vaak na de verschillende trainingcondities: W of R.

resultaten

Als afhankelijke variabele zijn, op identieke wijze als bij onderzoek 1, de gemiddelde responstijden per sessie berekend (Tggd) van alle items die zowel geldig als goed waren, maar nu van de eerste zes trials. Als accuratessemaat is in de analyse het aantal goed gelezen woorden op de voortoets (VT) per sessie opgenomen. Gemiddelde en spreiding van de responstijden zijn voor de training- en de generalisatieconditie apart in figuur 5 weergegeven. Op deze gemiddelde responstijden is een multilevel analyse uitgevoerd met herhaalde meting. Dit leverde een regressievergelijking op die is weergegeven in tabel 20; deze is als volgt te lezen. De trainingssessie met *random* aanbieding is gedefinieerd als referentieconditie. Andere condities worden daarmee vervolgens vergeleken. De middelste sessie van het geanalyseerde traject is ingesteld als sessienummer nul. De gemiddelde voorspelde responstijd - de intercept - in de referentieconditie is halverwege de training 1695 ms. Woorden die in de trainingssessies als wisselrij werden aangeboden, werden gemiddeld 248 ms sneller gelezen.



Figuur 5: Deelonderzoek 2, gemiddelde en spreiding van de gemiddelde responstijden per sessie van training- en generalisatiesessies afzonderlijk.

De gemiddelde responstijd verandert niet gedurende de training in algemene zin, maar in de generalisatieconditie worden de woorden elke volgende sessie gemiddeld 14 ms sneller gelezen. In de generalisatiesessies is gewerkt met pseudowoorden. Deze zijn voor alle lezers aanzienlijk moeilijker te lezen, maar er is een relatief generalisatie-effect. De generalisatiewoorden die in woordkern overlapt met de trainingwoorden blijken gemiddeld 282 ms langzamer gelezen te worden dan de trainingwoorden, de woorden die niet in woordkern overlapt met de trainingwoorden gemiddeld nog eens 46 ms langzamer. Ook het verschil tussen beide van 46 ms is significant.

Tabel 20: Deelonderzoek 2, multilevel analyse van gemiddelde responstijden (standaard meetfout) in de trainingssessies en aantal goed gelezen woorden op de voortoetsen (VT), met de indeling van proefpersonen als zwak (E-niveau) of niet-zwak (D-niveau of hoger) aan het einde van de trainingperiode.

	responstijd (SE)	aantal goed VT (SE)
Fixed effecten		
$\beta_{\text{intercept}}$	1.694,73 (62,16) ***	10,071 (0,282) ***
$\beta_{\text{sessienummer}}$	-4,37 (5,22) ^{ns}	0,012 (0,035) ^{ns}
$\beta_{\text{generalisatieconditie}}$	282,03 (28,38) ***	-0,100 (0,194) ^{ns}
$\beta_{\text{generalisatie beginletter}}$	46,30 (13,83) ***	-3,024 (0,144) ***
$\beta_{\text{generalisatieconditie} \times \text{sessienummer}}$	-13,83 (3,47) ***	0,110 (0,036) **
$\beta_{\text{training wisselij}}$	-248,43 (19,59) ***	
β_{zwak}	242,57 (77,69) **	-0,964 (0,392)*
$\beta_{\text{zwak} \times \text{generalisatieconditie}}$		-1,367 (0,405) ***
Random (niveau 2)		
$\text{var}[r_{\text{intercept}}]$	72.592 (18.156)	0,994 (0,293)
$\text{var}[r_{\text{sessienummer}}]$	672 (183)	0,013 (0,005)
$\text{var}[r_{\text{generalisatieconditie}}]$	18.058 (5.519)	
$\text{cov}[r_{\text{intercept}}, \text{sessienummer}]$	3.572 (1.426)	
$\text{var}[r_{\text{zwak} \times \text{generalisatieconditie}}]$		1,836 (0,794)
Residu (niveau 1)		
$\text{var}[r_{\text{intercept}}]$	40.033 (1.668)	4,370 (0,180)

*** $p < .001$ ** $p < .01$ * $p < .05$

Analoog aan deelonderzoek 1 is binnen de proefpersonen onderscheid gemaakt op basis van hun DMT1-score na afloop van de training. 16 leerlingen scoorden aan het einde van de trainingperiode nog steeds op E-niveau, zij zijn aangeduid als 'zwak'. De 19 relatief minder zwakke leerlingen die aan het einde van de trainingperiode op D-niveau scoorden zijn aangeduid als 'niet-zwak'. De zwakste lezers hadden gemiddeld 243 ms langer nodig om de woorden te lezen.

De multilevel analyse is herhaald met de voortoetsscores als afhankelijke variabele als alternatief voor het aantal goed gelezen woorden in de trainingssessies voor het aantonen van responsgeneralisatie. De trainingssessies betroffen steeds twaalf woorden, de generalisatiesessies kenden echter twee subsets van zes woorden die apart geregistreerd zijn. Om de scores te kunnen vergelijken zijn van de generalisatiesubsets de ruwe scores verdubbeld.

Het resultaat, weergegeven in de tweede kolom van tabel 20, levert enkele verschillen op in vergelijking met de analyse van de responstijden. De gemiddelde voorspelde voortoetsscore van de relatief minder zwakke lezers is in de trainingconditie 10,1. Er is op de voortoetsscore geen verschil tussen *random*- of wisselrijaanbieding. Op de voortoetsen van de generalisatiesessies was de gemiddelde voorspelde score voor alle leerlingen 0,9 (SE 0,2) lager ($p < .001$) en de zwakste lezers lazen daarenboven op de voortoetsen gemiddeld 1,9 (SE 0,5) woord minder goed ($p < .001$). Na introductie van de interactie van 'zwak' met 'generalisatie' verdween het generalisatie-effect en verminderte het hoofdeffect 'zwak' tot -0,96.

Tabel 21: Deelonderzoek 2, multilevel analyse van gemiddelde responstijden (standaard meetfout) in de trainingssessies.

	responstijd (SE)
Fixed effecten	
$\beta_{\text{intercept}}$	1.819,53 (51,85) ***
$\beta_{\text{sessienummer}}$	-4,37 (5,22) ns
$\beta_{\text{generalisatieconditie}}$	281,97 (28,38) ***
$\beta_{\text{generalisatie beginletter}}$	46,03 (13,83) ***
$\beta_{\text{training wisselrij}}$	-248,53 (19,58) ***
$\beta_{\text{generalisatieconditie} \times \text{sessienummer}}$	-13,83 (3,47) ***
Random (niveau 2)	
$\text{var}[r_{\text{intercept}}]$	87.322 (21.678)
$\text{var}[r_{\text{sessienummer}}]$	672 (183)
$\text{var}[r_{\text{generalisatieconditie}}]$	18.058 (5.519)
$\text{cov}[r_{\text{intercept}}, \text{sessienummer}]$	4.083 (1.572)
Residu (niveau 1)	
$\text{var}[r_{\text{intercept}}]$	40.034 (1.668)

*** $p < .001$ ** $p < .01$ * $p < .05$

De gemiddelde score op de voortoetsen verandert in algemene zin niet gedurende de training, maar in de generalisatieconditie wordt elke volgende trainingdag gemiddeld 0,11 woord meer goed gelezen. De pseudowoorden van de generalisatiesets die wel de letters, maar niet de woordkern gemeen hadden met woorden uit de daarvoor geoefende trainingset, blijken het moeilijkst. De zwakste lezers lazen daarop gemiddeld 5,4 woorden minder goed, de relatief betere lezers gemiddeld 3,0 woorden minder. Een differentieel effect bestond voor pseudowoorden die de woordkern wél overeenkwamen met de woorden uit de trainingset, alleen voor de zwakste lezers. Zij lazen gemiddeld 1,4 pseudowoorden minder goed. Voor de relatief betere lezers was er geen verschil tussen de trainingwoorden en de pseudowoorden van de generalisatiesets als die dezelfde woordkern hadden als de trainingwoorden. Een interactie van dit effect met de trainingconditie die eraan voorafging werd ook met de voortoetsscores niet aangetoond.

Ook in deelonderzoek 2 werd gekeken naar de voorspellende waarde van individuele leercurves. Daartoe werd de factor 'zwak' uit de regressievergelijking verwijderd.

Behalve de intercept waren ook de factoren 'sessienummer' en 'generalisatieconditie' zogenaamde *random* factoren; naast een systematisch groepseffect, de *fixed factor*, van de generalisatieconditie is tevens sprake van betekenisvolle, significante verschillen zijn tussen individuele proefpersonen op deze factor (tabel 21). Er is geen significant *fixed* effect van het sessienummer, maar wel betekenisvolle verschillen tussen leerlingen op deze factor. De afwijking van individuele leerlingen van het groepsgemiddelde is per factor berekend. De onderlinge samenhang en de samenhang met de leesvaardigheid op didactische leeftijd 9 maanden is weergegeven in tabel 22.

Tabel 22: Deelonderzoek 2, correlatietabel van DMT-score aan het einde van de trainingperiode en de niveau-2 residuwaarden op de randomfactoren

Correlatiematrix	1	2	3	4
1. DMT 1 op dl 9	--	-.51***	-.34*	-.32*
2. intercept		--	.57***	-.10
3. lineaire component			--	-.13
4. generalisatiecomponent				--

*** $p < .001$ * $p < .05$

De gemiddelde DMT-1 score aan het eind van de trainingperiode was 20,5 (SD 6,2). Er is een sterke negatieve samenhang tussen de gemiddelde responstijd van de leerlingen, de intercept, en deze DMT-eindscore. De leerlingen met de laagste gemiddelde responstijden zijn de leerlingen met de hoogste DMT-score aan het einde van de trainingperiode. Er is een matige negatieve samenhang tussen DMT-score en zowel de lineaire als de generalisatiecomponent. De leerlingen met de hoogste DMT-score aan het einde van de trainingperiode zijn ook gemiddeld de leerlingen die in de loop van de training een snellere daling van responstijden laten zien en de leerlingen die in de generalisatiesessies minder extra tijd nodig hadden. De intercept en de lineaire component hangen onderling sterk samen, met andere woorden de leerlingen met de laagste gemiddelde responstijd zijn ook de leerlingen die in de loop van de training de snelste daling in responstijd lieten zien. De individuele waarden van deze variabelen zijn als predictor-variabelen opgenomen in een multiple regressieanalyse ($\beta_{\text{intercept}} = -.48$, $p < .01$; $\beta_{\text{lin}} = -.12$, n.s.; $\beta_{\text{gen}} = -.38$, $p < .01$), gezamenlijk wordt 41% van de variantie van de afhankelijke variabele verklaard ($r = .64$, $F(3,31) = 7,19$, $p < .001$). Zowel de intercept als de generalisatiecomponent leveren een significante bijdrage aan de voorspelling van de afhankelijke variabele, de DMT-score aan het einde van de training.

conclusie

In de training met woorden van niveau 1 die overlappen in woordkern (klinker + eindmedeklinker), is het gerealiseerde relatieve generalisatie-effect groter dan in het vorige deelonderzoek en het faciliterend effect van aanbieding van trainingwoorden in de vorm van een wisselrij is aanzienlijk. De differentiële instructiegevoeligheid op de training is gerelateerd aan de uitkomstmaat door een combinatie van de gemiddelde responstijd en het verschil in leesprestatie tussen bestaande en pseudowoorden.

4.3.7 Deelonderzoek 3

specificatie van de onderzoeksvraag

Evenals in de eerdere deelonderzoeken en op vergelijkbare wijze is de differentiële instructiegevoeligheid onderzocht in relatie tot de leesvaardigheid van de leerlingen aan het einde van de trainingperiode. De leerlingen van deelonderzoek 3 ontvingen gedurende een half schooljaar wekelijks viermaal een half uur individuele instructie op het gebied van technisch lezen. Het onderwijs in technisch lezen mag voor deze leerlingen daarom als optimaal worden getypeerd. Tegen deze achtergrond is tevens onderzocht of de differentiële instructiegevoeligheid zoals in voorgaande paragrafen geoperationaliseerd ook op individueel niveau voldoende betrouwbaar de leerlingen zou identificeren die aan het einde van de trainingperiode nog wel of juist niet meer tot de uitvallers gerekend zouden moeten worden. In het programma is, gebruikmakend van de resultaten van deelonderzoek 1 en 2, de mate van overlap tussen trainingwoorden gelijkmatige opgebouwd, in de verwachting dat dit zou leiden tot grotere verschillen tussen leerlingen.

proefpersonen

Deelonderzoek 3 is uitgevoerd met de 39 leerlingen die als groep T11 in dezelfde periode deelnamen aan het rt-onderzoek dat in hoofdstuk 3 is beschreven, zie tabel 23 voor beschrijvende informatie. Het aantal leerlingen aan het eind van de onderzoeksperiode was 38.

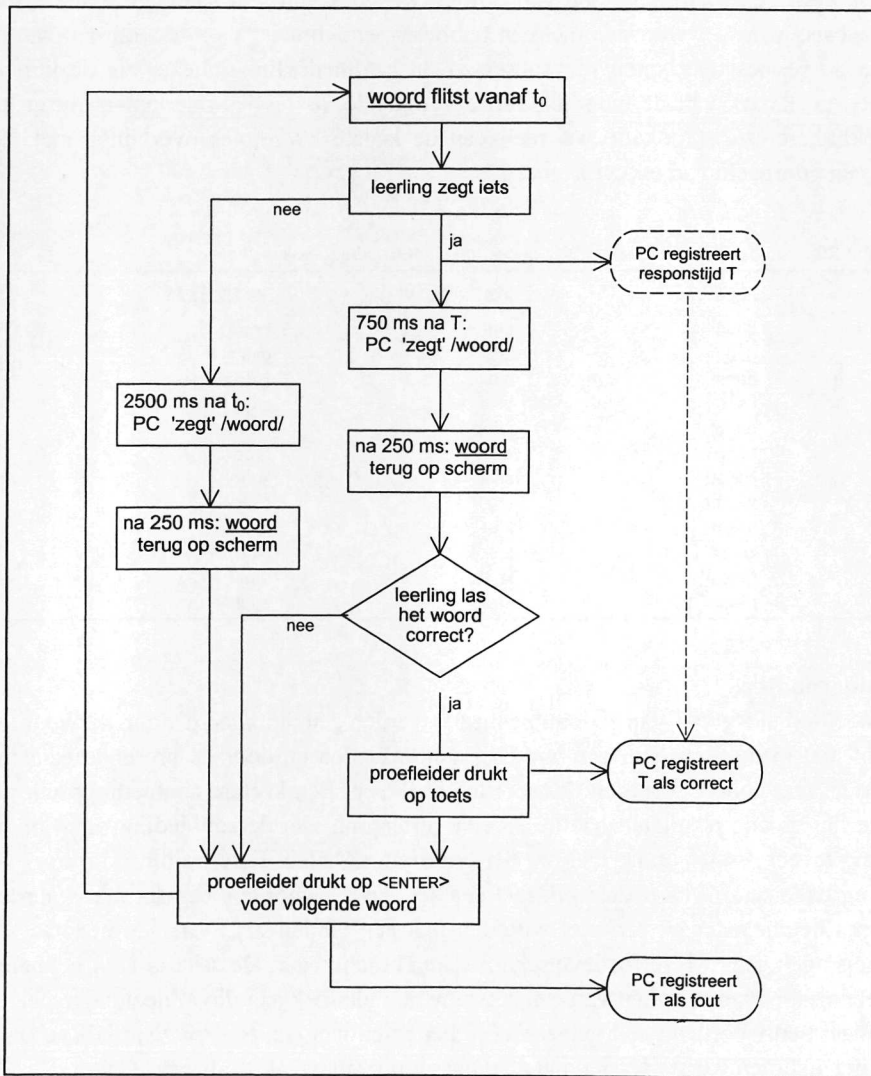
Tabel 23: Beschrijvende informatie over de proefpersonen (N=39) van deelonderzoek 3.

Sekse totaal	Jongens 26	Meisjes 13
SE-score totaal	1 21	1,25 7
	Min.	Max.
DMT-2, ruwe score	0	12
		M (SD)
		7,0 (3,0)

procedure

Dit deelonderzoek liep van september tot februari groep 4, vanaf didactische leeftijd 11 maanden tot didactische leeftijd van 16 maanden. De trainingprocedure is schematisch weergegeven in figuur 6. Daarbij waren in het programma acht moeilijkheidsniveaus ingebouwd (tabel 24). Bij deelonderzoek 1 was gebleken dat de ingestelde maximale responstijd van 3000 ms voor de meeste leerlingen ruimer was dan nodig, maar voor enkelen te kort. Om die reden is de maximale responstijd ingesteld op 2500ms, maar is er een extra 0-niveau toegevoegd met aanbieding in de vorm van een wisselrij en een maximale responstijd van 3500 ms. De training startte voor iedere leerling op niveau 2.

De resultaten werden per sessie bijgehouden. In het programma zaten twee tellers. De omhoogteller werd met één opgehoogd als de leerling de sessie in zes, zeven of acht trials haalde. Haalde hij de sessie in negen of tien sessies, dan werd dat niet geteld. Haalde hij een sessie in elf of twaalf sessies of haalde hij een sessie niet, dan werd de omlaagteller verhoogd met een. Als de omhoogteller het aantal van vier bereikt had, ging de leerling naar een hoger niveau. Beide tellers werden dan weer op nul gezet. Als de omlaagteller het aantal van zes bereikt had, ging de leerling naar een lager niveau en werden beiden tellers weer op nul gezet.



Figuur 6: Stroomschema van de woordtraining in deelonderzoek 3.

Tabel 24: Opklimmende moeilijkheidsniveaus van de woordtraining in deelonderzoek 3.

niveau 0:	aanbieding als wisselrij,	aanbiedingstijd	3500 ms
niveau 1:	aanbieding als wisselrij,	aanbiedingstijd	2500 ms
niveau 2:	aanbieding in random volgorde,	aanbiedingstijd	2500 ms
niveau 3:	aanbieding in random volgorde,	aanbiedingstijd	1000 ms
niveau 4:	aanbieding in random volgorde,	aanbiedingstijd	800 ms
niveau 5:	aanbieding in random volgorde,	aanbiedingstijd	600 ms
niveau 6:	aanbieding in random volgorde,	aanbiedingstijd	400 ms
niveau 7:	aanbieding in random volgorde,	aanbiedingstijd	200 ms

inhoud

Er zijn voor een trainingperiode van vijftien weken zestig sets samengesteld van elk drie subsets van vier woorden met een dubbele medeklinker. Van de woorden van de eerste 20 sessies overlaptten per subset zowel het medeklinkercluster als de klinker (MMK- en -KMM), van de woorden van de volgende 20 sessies overlaptten alleen het medeklinkercluster (MM), de woorden van de laatste 20 sessies overlaptten niet, zie voor een voorbeeld van elk type tabel 25.

Tabel 25: Voorbeeld van drie sets trainingwoorden uit deelonderzoek 3.

<u>sessie 1-20</u>	<u>sessie 21-40</u>	<u>sessie 41-60</u>
dans	blaf	bruut
gans	blik	glom
hans	blok	kroon
kans	bluf	sluis
doopt	snor	brem
hoopt	snuut	groot
koopt	snik	knoet
loopt	snel	sluip
hoest	krul	grijp
koest	kreuk	plus
moest	krijt	druil
roest	kraak	snot

trainingcondities

De moeilijkheidsgraad van de aanbieding is verder gemanipuleerd door de woorden wel of niet in de vorm van een wisselrij aan te bieden en door de presentatieduur te bekorten, met andere woorden de woorden te flitsen. Een kortere aanbiedingsduur als de leerling goede resultaten haalde en een verlenging van de aanbiedingsduur als de leerling teveel sessies nodig had om het criterium te halen. Elke trainingdag kreeg de leerling twee sessies na elkaar, elk met een set van 12 woorden, een set met woorden met een begincluster en een met woorden met een eindcluster. Voor de in totaal 30 trainingdagen waren 17 effectieve schoolweken beschikbaar. De criteria voor het halen van een trial en het afsluiten van een sessie waren identiek aan die in deelonderzoek 2: minimaal tien woorden goed gelezen voor het halen van een trial en zes trials gehaald voor het afsluiten van de sessie, met een maximum van twaalf trials per sessie.

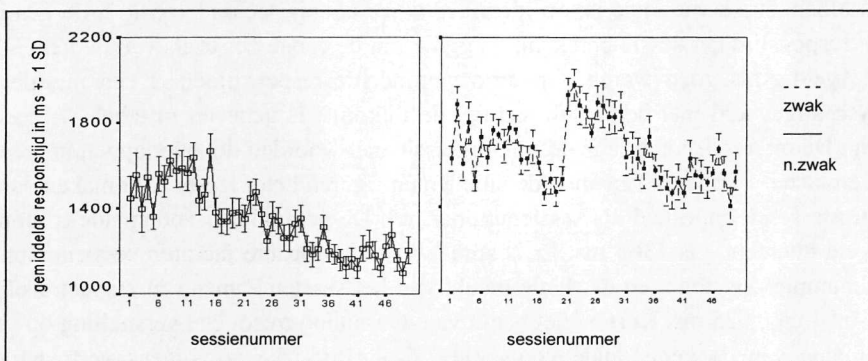
bepaling van trainingseffecten

Als afhankelijke variabele is, op identieke wijze als bij deelonderzoek 2, de gemiddelde responstijd per sessie berekend - Tggd - van alle items die zowel geldig als goed waren van de eerste zes trials als maat voor de vlotheid van lezen. De accuratessesmaat was het aantal goed gelezen woorden op de voortoets (VT) per sessie.

resultaten

Tussen de proefpersonen is op dezelfde manier als in deelonderzoek 1 onderscheid gemaakt op basis van hun DMT2-score na afloop van de training. 20 leerlingen scoorden aan het einde van de trainingperiode nog steeds op E-niveau scoorden, zij zijn aangeduid als 'zwak'. De 18 relatief minder zwakke leerlingen die aan het einde van de trainingperiode op D-niveau scoorden zijn aangeduid als 'niet-zwak'. Gemiddelde en spreiding van de responstijden zijn in figuur 7 voor deze twee onderscheiden groepen apart weergegeven.

Na de overgang van sessie 20 naar 21, van overlap in medeklinkercluster + klinker naar alleen overlap in medeklinkercluster, was de training volgens de trainers voor een aantal leerlingen erg moeilijk geworden, het criterium werd naar het oordeel van de trainers te vaak niet meer gehaald, terwijl er ook geen herstel optrad met het verloop van de sessies. Na sessie 30 is daarom besloten om de training met elf leerlingen voort te zetten met de gemakkelijker woordensets uit onderzoek 2, dat wil zeggen met MKM-woorden. De responstijden van de groep 'zwak' vanaf sessie 31 betreffen derhalve slechts 10 leerlingen.



Figuur 7: Deelonderzoek 3, gemiddelde en spreiding van de gemiddelde responstijden per sessie met de indeling van proefpersonen als zwak (E-niveau) of niet-zwak (D-niveau of hoger) aan het einde van de trainingperiode.

Het stoppen met de oorspronkelijke training met elf van de 39 leerlingen was gebaseerd op het klinisch oordeel van de trainers dat de training met medeklinkerclusters voor deze leerlingen te moeilijk was. Het blijkt achteraf dat dit oordeel over geen enkele leerling gegeven is die aan het einde van het onderzoek, $dl=16$, op de DMT-2

boven het E-niveau is gekomen ($\chi^2(1) = 12,21$; $p < .001$; tabel 26). Bij deze elf leerlingen zat ook de leerling die kort daarna verhuisde, deze is daarom bij de groep 'zwak' gerekend.

Tabel 26: Het klinisch oordeel van de trainers tegenover het leesniveau aan het einde van de trainingperiode.

oordeel trainers :	DMT-2 niveau op dl=16		
	E	D+	Totaal
teruggezet naar niveau MKM	10	-	10
doorgegaan op niveau MMKM/MKMM	10	18	28
Totaal	20	18	38

Binnen de overgebleven groep van 28 leerlingen met wie de training is voortgezet, is het percentage - uiteindelijk - zwakke lezers ten opzichte van de oorspronkelijke experimentele groep van 39 leerlingen verminderd van ruim de helft naar ruim een derde. Deze 28 leerlingen hebben minimaal 52 sessies gedaan. De multilevel analyse zoals gegeven in tabel 27 betreft de eerste 50 sessies, het maximaal mogelijke aantal herhaalde metingen in het gebruikte analyseprogramma. Een belangrijk voordeel van deze techniek is, dat missing data niet snel een probleem vormen. Dit maakte het mogelijk toch alle 39 leerlingen in één analyse te betrekken. Behalve de overgang van sessie 20 naar sessie 21, van overlap in medeklinkercluster + klinker naar alleen medeklinkercluster, is ook de overgang van sessie 40 naar sessie 41 als *dummy*-variabele geïntroduceerd, dat wil zeggen de overgang van overlap in medeklinkercluster naar het werken met woordensets die niet meer overlappen.

Als afhankelijke variabele is, op identieke wijze als bij deelonderzoek 2, de gemiddelde responstijd per sessie berekend - Tggd - van de eerste zes trials van de items die zowel geldig als goed waren. Op deze gemiddelde responstijden is een multilevel analyse uitgevoerd met herhaalde meting, de uitkomst is gegeven in tabel 27, eerste kolom. De referentieconditie is de trainingssessie met woorden die overlappen in medeklinkercluster + klinker. Teneinde de tabellen en figuren beter leesbaar te maken, is de twintigste sessie ingesteld als sessienummer nul. De gemiddelde voorspelde responstijd - de intercept - is 1366 ms. Er is sprake van significante factoren sessienummer, sessienummer-kwadraat en de derde macht van het sessienummer van respectievelijk -33, -0,84 en 0,025 ms. Kort na het begin van de training treedt een versnelling op van de afname van de gemiddelde voorspelde responstijd, die vervolgens vertraagt en uiteindelijk afvlakt.

Voor het vaststellen van verschillen in instructiegevoeligheid zijn de leerlingen verdeeld in een groep leerlingen die aan het einde van de trainingperiode, dl=16, op de DMT nog steeds op E-niveau scoorden, versus de leerlingen die aan het einde van de training op D-niveau of hoger scoorde. De zwakste lezers hadden over alle condities heen gemiddeld 139 ms langer nodig.

Het lezen van de woorden van woordensets die alleen in cluster overlappen, bleek gemiddeld moeilijker dan bij overlap van klinker plus medeklinkercluster. Deze factor

is aangeduid als 'KMM'⁶. Vanaf sessie 21 bleken de leerlingen gemiddeld 270 ms langer nodig te hebben, maar dit effect verdween na introductie van de interactie met de factor 'zwak'. Ook de overgang van overlap in medeklinkercluster naar trainingsets waarin de woorden geen overlap meer vertoonden, aangeduid als de factor 'MM', bleek voor alle leerlingen gemiddeld moeilijker, de woorden werden gemiddeld 69 ms langzamer gelezen.

Tabel 27: Deelonderzoek 3, multilevel analyse van gemiddelde responstijden in de trainingssessies (standaard meetfout) en aantal goed gelezen woorden op de voortoetsen (VT), met de indeling van proefpersonen als zwak (E-niveau) of niet-zwak (D-niveau of hoger) aan het einde van de trainingperiode.

	responstijd (SE) N=39	responstijd (SE) N=28	aant. goed voortoets (SE)
Fixed effecten			
$\beta_{\text{intercept}}$	1.366,14 (52,40) ***	1.385,15 (51,93) ***	9,479(0,283) ***
$\beta_{\text{sessennummer}}$	-32,86 (4,16) ***	-31,21 (4,62) ***	0,057(0,019) **
$\beta_{\text{sessie-kwadraat}}$	-0,84 (0,16) ***	-0,93 (0,18) ***	0,0028(0,001) ***
$\beta_{\text{sessie-derde macht}}$	0,025 (0,003) ***	0,026 (0,003) ***	
$\beta_{\text{sessie-vierde macht}}$			-0,000007(0,000002) ***
β_{KMM}	80,94 (60,88)	57,07 (37,33)	-1,657(0,369) ***
β_{MM}	68,65 (17,64) ***	67,19 (17,36) ***	-1,475(0,273) ***
$\beta_{\text{sessenr. x KMM}}$	26,43 (6,12) ***	25,95 (6,75) ***	
β_{zwak}	139,08 (66,93) *	206,34 (78,49) **	-1,644(0,346) ***
$\beta_{\text{zwak x KMM}}$	351,78 (75,00) ***	162,60 (53,62) **	-1,936(0,385) ***
$\beta_{\text{eindcluster}}$	-55,28 (7,15) ***	-56,80 (7,40) ***	
Random (niveau 2)			
$\text{var}[r_{\text{intercept}}]$	45,208 (11,409)	43,447 (12,816)	0,965(0,346)
$\text{var}[r_{\text{sessennummer}}]$	81,85 (24,49)	77,49 (26,63)	0,00203(0,00106)
$\text{var}[r_{\text{sessie-kwadraat}}]$	0,10 (0,03)	0,09 (0,03)	0,000002(0,000001)
$\text{var}[r_{\text{KMM}}]$	66,405 (17,732)	16,437 (7,160)	1,755(0,704)
$\text{cov}[r_{\text{sessenr.}, \text{sessie-kwadraat}}]$	-2,51 (0,78)	-2,28 (0,83)	
$\text{cov}[r_{\text{sessie-kwadraat}, \text{KMM}}]$	44,60 (18,28)		
Residu (niveau 1)			
$\text{var}[r_{\text{intercept}}]$	29,971 (1,271)	26,309 (1,300)	3,540(0,179)
$\text{var}[r_{\text{sessennummer}}]$	21,50 (4,88)	23,02 (5,01)	0,0018(0,0008)
$\text{cov}[r_{\text{intercept}, \text{sessennummer}}]$	-643,82 (63,04)	-587,30 (65,50)	

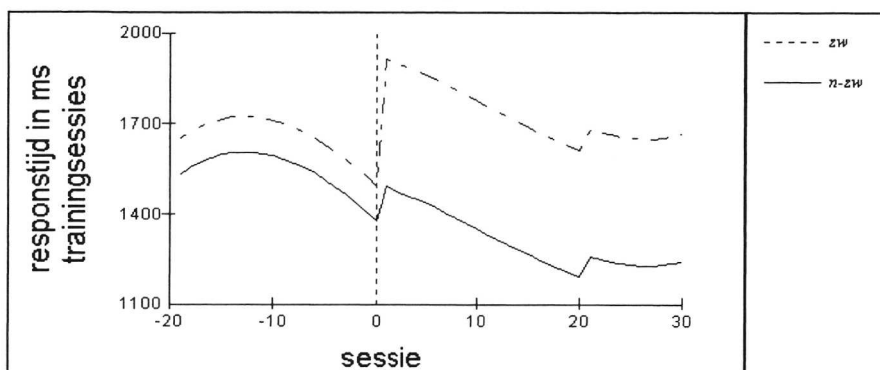
*** p < .001 ** p < .01 * p < .05

Evenals in deelonderzoek 1 blijken woorden met een eindcluster gemakkelijker gelezen te worden dan woorden met een begincluster, de gemiddelde responstijd is 55 ms korter. Ten slotte is er nog een extra positieve lineaire component vanaf de overgang van KMM naar MM van 26 ms.

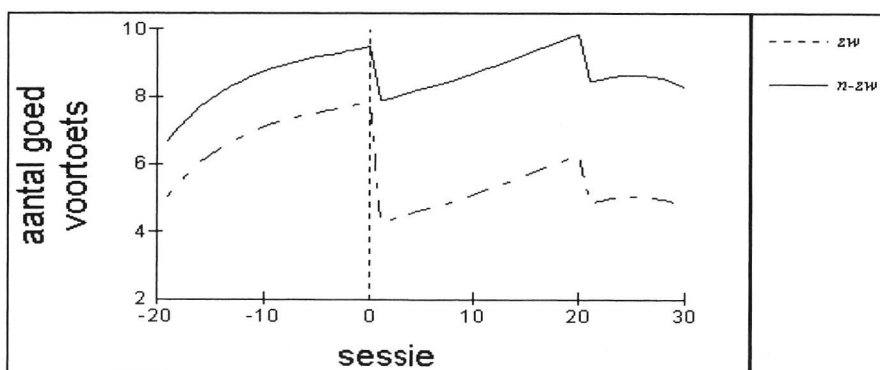
Het meest in het oog springende effect is de interactie tussen 'zwak' en 'KMM'. De zwakste lezers hadden voor de sets vanaf sessie 21 (de overgang van KMM naar MM)

⁶ Met de factor KMM worden behalve de woorden van het type MKMM steeds ook MMKM-woorden bedoeld; ook KMM moet daarom worden gelezen als KMM of MMK.

gemiddeld nog eens 352 ms langer nodig dan de minder zwakke lezers. Figuur 8 illustreert de het verloop van de gemiddelde responstijden van de leerlingen die aan het eind van de training nog steeds op E-niveau scoorden ten opzichte van de leerlingen die aan het eind van de training de betere lezers waren.



Figuur 8: Deelonderzoek 3, gemiddelde responstijd op de trainingssessies met indeling van proefpersonen in E-niveau (zw) en D-niveau of hoger (n.zw) aan het einde van de trainingperiode.



Figuur 9: Deelonderzoek 3, gemiddeld aantal goed op de voortoetsen met de indeling van proefpersonen in E-niveau (zw) en D-niveau of hoger (n.zw) aan het einde van de trainingperiode.

Inspectie van het verloop van de ruwe responstijden in figuur 7 suggereert dat het juist de leerlingen zijn geweest die de sterkste reactie vertoonden op de overgang na sessie 20 met wie de training op MKMM-niveau na sessie 30 is gestopt. Om deze veronderstelling te toetsen is de analyse op de responstijden herhaald met uitsluiten de 28 leerlingen die alle 50 sessies gedaan hebben. Deze leverde dezelfde effecten en interacties op, met twee uitzonderingen (tabel 27, tweede kolom). Het effect van de factor 'zwak' is met 206 ms (SE 78) groter en de interactie van 'zwak' met 'KMM' is met 163 ms (SE 54) kleiner dan in de eerste analyse. Dit bevestigt het vermoeden.

De analyse is herhaald met de voortoetsscores als afhankelijke variabele (tabel 27, derde kolom). Het resultaat is vergelijkbaar met de analyse van de responstijden. De gemiddelde voorspelde voortoetsscore in sessie 20 is 9,5 voor de leerlingen die aan het eind van de trainingperiode de betere lezers waren; de zwakste lezers lezen gemiddeld 1,6 woord minder goed. Er is sprake van een positieve lineaire en kwadratische component en van een negatieve vierdemachts component. Het samengestelde resultaat daarvan is dat de aanvankelijk snelle toename vermindert en uiteindelijk afvlakt. Na de overgang van overlap in KMM naar MM (sessie 21) daalt de score met 1,6 woord voor de betere lezers en 3,8 woord voor de zwakste lezers. Ook de overgang van overlap in medeklinkercluster naar trainingsets waarin de woorden geen overlap meer vertoonden (sessie 41), bleek voor alle leerlingen gemiddeld moeilijker; er wordt gemiddeld 1,5 woord minder goed gelezen. Figuur 9 illustreert het verloop van de accuratesse op de voortoetsen van de leerlingen die aan het eind van de training nog steeds op E-niveau scoorden ten opzichte van de leerlingen die aan het eind van de training betere lezers waren.

Tabel 28: Deelonderzoek 3, multilevel analyse van gemiddelde responstijden in de trainingssessies (standaard meetfout).

	responstijd (SE)
Fixed effecten	
$\beta_{\text{intercept}}$	1.441,11 (39,46) ***
$\beta_{\text{sessienummer}}$	-32,81 (4,15) ***
$\beta_{\text{sessie-kwadraat}}$	-0,84 (0,16) ***
$\beta_{\text{sessie-derde macht}}$	0,025 (0,003) ***
β_{KMM}	270,36 (55,24) ***
β_{MM}	68,90 (17,63) ***
$\beta_{\text{sessienummer} \times \text{KMM}}$	26,32 (6,12) ***
$\beta_{\text{eindcluster}}$	-55,27 (7,15) ***
Random (niveau 2)	
$\text{var}[r_{\text{intercept}}]$	49.202 (12.314)
$\text{var}[r_{\text{sessienummer}}]$	81,20 (24,32)
$\text{var}[r_{\text{sessie-kwadraat}}]$	0,10 (0,03)
$\text{var}[r_{\text{KMM}}]$	104.442 (26.318)
$\text{cov}[r_{\text{sessienr.}}, r_{\text{sessie-kwadraat}}]$	-2,52 (0,78)
$\text{cov}[r_{\text{sessie-kwadraat}}, r_{\text{KMM}}]$	57,53 (22,55)
Residu (niveau 1)	
$\text{var}[r_{\text{intercept}}]$	29.955 (1.270)
$\text{var}[r_{\text{sessienummer}}]$	21,54 (4,88)
$\text{cov}[r_{\text{intercept}}, r_{\text{sessienummer}}]$	-644,03 (63,02)

*** $p < .001$

De overgang van overlap in KMM naar MM heeft effect op alle leerlingen, maar in het bijzonder op de zwakste lezers. Het effect van de overgang van overlap in medeklinkercluster MM naar training met woorden die niet meer overlaptten, lijkt veel minder uitgesproken. Mogelijk wordt het laatste effect in de analyses overschaduwed door drie

omstandigheden: de invloed van de derde- respectievelijk vierdemachts sessiefactor krijgt vooral bij de hoge sessienummers invloed, na de laatste overgang van MM naar niet-MM bij sessie 41 konden nog slechts tien sessies in de analyse betrokken worden en het feit dat een aantal zwakke lezers in de tweede helft van de training gehalveerd was. Om die reden is de analyse herhaald over de resultaten van sessie 21 tot 60. Voor de vergelijkbaarheid is het nulpunt gesteld op sessie 21. De gemiddelde voorspelde responstijd in sessie 21 is voor de relatief betere lezers in deze selectie van 28 leerlingen 1477 ms. Er is een significante lineaire component 'sessienummer' van -21 ms en een kwadratische component van 0,3 ms. De gemiddelde responstijd neemt af, maar naarmate de leerlingen verder komen in de training wordt de afname minder groot. Er is sprake van een vergelijkbaar eindclustereffect van -26 ms. De leerlingen - die aan het eind van de training nog steeds op E-niveau scoorden hebben gemiddeld 359 ms langer nodig dan de leerlingen die aan het eind van de training betere lezers waren. Zoals verwacht is er in deze analyse geen sprake van een significante derdemachtscomponent van het sessienummer, maar wel een effect van de overgang van sessie 40 naar sessie 41 van 62 ms. Er zijn echter geen significante interacties van de factor 'zwak' met enige andere factor; het effect van wel naar geen overlap in medeklinkercluster blijft voor alle leerlingen even groot en aanzienlijk geringer dan de overgang van KMM naar MM.

Op groepsniveau lijkt vooral het effect van de overgang van KMM naar MM het onderscheid te maken tussen de leerlingen die aan het einde van de training op of boven E-niveau zullen scoren. Om dit te onderzoeken werden uit de regressievergelijking met de responstijd als afhankelijke variabele de factor 'zwak' en de interactie van de factor 'zwak' met de overgang KMM naar MM verwijderd (tabel 28). De factoren intercept, sessienummer, sessiekwadraat en overgang van KMM naar MM zijn zogenaamde *random* factoren, dat wil zeggen dat er naast een systematisch groepseffect ook betekenisvolle, significante verschillen zijn tussen individuele proefpersonen op deze factoren.

Tabel 29: Deelonderzoek 3, correlatietabel van DMT-score aan het einde van de trainingperiode en de niveau-2 residuwaarden op de randomfactoren.

Correlatiematrix	1	2	3	4	5
1. DMT 2 op dl 16	--	-.33*	.03	-.10	-.65***
2. intercept		--	-.26'	.04	-.10
3. lineaire component			--	-.88***	-.37**
4. kwadratische component				--	.56***
5. factor overgang MM - KMM					--

*** p < .001

** p < .05

* p < .05

' p < .10

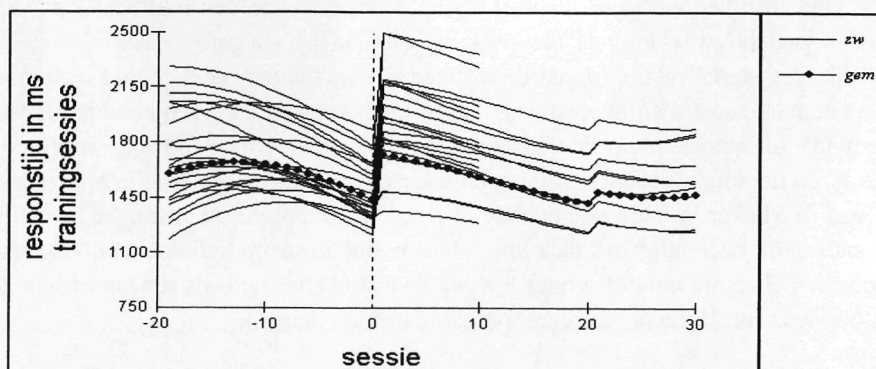
De gemiddelde DMT-2 score aan het eind van de trainingperiode was 23,2 (SD 11,1). Er is een matige negatieve samenhang tussen de gemiddelde responstijd van de leerlingen, de intercept, en deze DMT-eindscore. De leerlingen met lagere responstijden in sessie 20 van de training zijn de leerlingen met gemiddeld hogere DMT-scores aan het einde van de trainingperiode. Er is een sterk negatieve samenhang tussen de mate waarin de

overgang van KMM naar MM effect heeft op de responstijd en de DMT-score aan het eind van de training. Ook de lineaire en de kwadratische component hangen negatief samen. De leerlingen bij wie de responstijden op de trainingssessies het snelst dalen, zijn ook de leerlingen bij wie de afname van die daling het geringst is (tabel 29). De individuele waarden van deze variabelen zijn als predictorvariabelen opgenomen in een multiple regressieanalyse ($\beta_{\text{intercept}} = -.32, p < .05$; $\beta_{\text{lin}} = 0,22, \text{n.s.}$; $\beta_{\text{kwad}} = .23 \text{ n.s.}$; $\beta_{\text{KMM-MM}} = -.89, p < .001$), gezamenlijk wordt 61% van de variantie van de afhankelijke variabele verklaard ($r = .78$; $F(4,33) = 12,68$; $p < .001$). Naast de intercept levert vooral de factor 'overgang van KMM naar MM' een significante bijdrage aan de voorspelling van de afhankelijke variabele, de DMT2-score aan het einde van de training

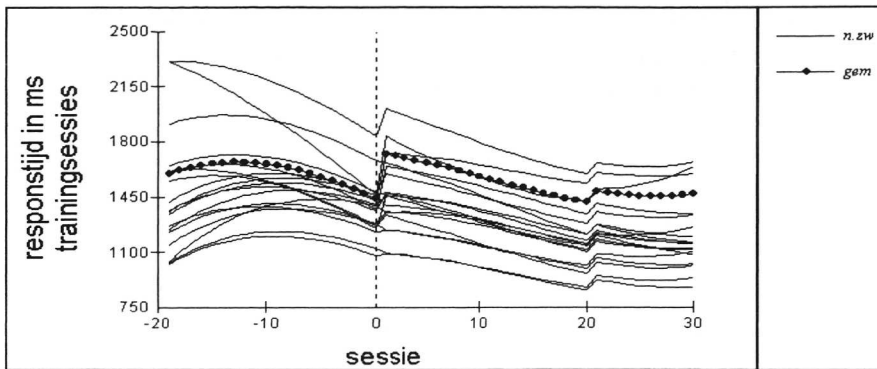
Tabel 30: Voorspelling van het trainingresultaat op basis van de gebleken instructiegevoeligheid.

instructiegevoeligheid op de training	DMT niveau op dl=16		
	E	D+	Totaal
laag	16	2	18
hoog	4	16	20
Totaal	20	18	38

Maar een significante maat op basis van groepsgemiddelden betekent nog niet dat op basis van de differentiële instructiegevoeligheid de leesscore na afloop van de training ook op individueel niveau betrouwbaar is te voorspellen, zie hoofdstuk 2. Daarom is gekeken naar de voorspellende waarde van de individuele leercurves. De groep leerlingen is in tweeën gedeeld op basis van de DMT-2 score aan het eind van de trainingperiode. Een ruwe score < 24 staat voor het E-niveau, scores ≥ 24 voor D-niveau of hoger. De individuele berekende leercurves van leerlingen die aan het einde van de trainingperiode nog steeds op E-niveau scoorden, zijn weergegeven in figuur 10. De individuele berekende leercurves van de leerlingen met een score op D-niveau of hoger in figuur 11. De gestippelde lijn geeft het gemiddelde van de totale groep weer.



Figuur 10: Deelonderzoek 3, berekende individuele leercurves van responstijden van de leerlingen die aan het einde van de trainingperiode lazen op E-niveau (zw) en het gemiddelde van de totale groep (gem).



Figuur 11: Deelonderzoek 3, berekende individuele leercurves van responstijden van de leerlingen die aan het einde van de trainingperiode lazen op D-niveau of hoger (n.zw) en het gemiddelde van de totale groep (gem).

Leerlingen met een voorspelde dmt-score < 24 met behulp van de multiple regressie-vergelijking op basis van hun leercurve op de training, zijn benoemd als de leerlingen met een relatief lage instructiegevoeligheid. Twee van deze leerlingen blijken op het einde van de trainingperiode op de DMT-2 toch op D-niveau te scoren; de vals positieven. Vier leerlingen met een zo berekende relatief hoge instructiegevoeligheid scoorden op het einde van de trainingperiode op de DMT-2 toch op E-niveau; de vals negatieven. De voorspellende waarde op basis van de berekende individuele leercurve (tabel 30) is significant ($\chi^2(1) = 18,03; p < .001$).

conclusie

De woordtraining in deelonderzoek 3 was grotendeels gelijk aan die in deelonderzoek 1, maar de mate van overlap tussen trainingwoorden werd sterker gevarieerd. De verwachting was dat hierdoor de verschillen tussen leerlingen zouden toenemen en dat deze toename groter zou worden gedurende het verloop van de training. Het loslaten van de klinker in het overlappende deel tussen de trainingwoorden lijkt echter de verschillen tussen de 'zwakke' en de 'niet-zwakke' lezers te hebben geforceerd.

Er blijkt een sterke relatie tussen de differentiële instructiegevoeligheid op de training en de DMT-score als uitkomstmaat. Deze wordt bepaald door een combinatie van de gemiddelde responstijd op de training en de snelheid waarmee de responstijd in de loop van de training afneemt, maar vooral door de grootte van de reactie op het loslaten van de klinker in het overlappende deel van de woorden. De instructiegevoeligheid zoals in dit deelonderzoek naar voren kwam, blijkt ook op individueel niveau een goede voorspeller. Vooral het verwerken van medeklinkerclusters als subwoordeenheid lijkt voor de zwakste lezers een zeer moeilijk te nemen drempel.

4.3.8 Deelonderzoek 4

specificatie van de onderzoeksvraag

Voor deelonderzoek 4 is gezocht naar een trainingvorm waarmee de moeilijkheidsgraad van het werken met dubbele medeklinkers verder gereduceerd kon worden teneinde de allerzwaksten toch met woorden van niveau-2 te kunnen laten oefenen. In plaats van actief lezen is gekozen voor een matchingtaak. De leerling wordt gevraagd een vergelijking te maken tussen een aangeboden klankvorm van een woord of woorddeel en de geschreven vorm van een woord. Matching is gemakkelijker dan hardop lezen. De interpretatie daarvan is, dat bij hardop lezen de fonologische code van het woord tweemaal moet worden geactiveerd, tegen eenmaal in een matchingtaak (Hogaboam & Perfetti 1978). Deze wijze van werken is ontleend aan Frederiksen e.a. (1985) en in Nederland toegepast door Das-Smaal e.a. (1996) en Yap (1993). Zij werkten echter met subwoordeenheden waarin steeds ook een klinker was opgenomen en niet met geïsoleerde medeklinkerclusters. De training weerspiegelde het voorkomen van hoog- en laagfrequente clusters. Om responsgeneralisatie vast te kunnen stellen is parallel aan de training viermaal een experimentele woordleestoets afgenomen waarin de in de training gerealiseerde verdeling van hoog- en laagfrequente medeklinkerclusters terugkwam.

proefpersonen

Met de scholen is, per school, bekeken welke leerlingen na afronding van het rt-onderzoek/deelonderzoek 3 de zwakste lezers waren. Daaruit zijn de leerlingen gekozen als proefpersonen voor het laatste onderzoek. De 38 leerlingen die voor onderzoek 4 werden geselecteerd scoorden bij herhaalde meting, op $dl=10$, $dl=11$ en $dl=16$, op de DMT-2 op E-niveau. Dit waren 28 leerlingen uit groep T16, 5 leerlingen uit groep C3, de zittenblijvers die aanvankelijk door hun school van deelname waren uitgesloten, en 5 leerlingen uit groep T11 volgens de subgroepaanduiding zoals beschreven in hoofdstuk 3. Per school werden deze leerlingen ingedeeld in gematchte paren op basis van hun ruwe score op de DMT-2 in februari. In totaal waren er zo 19 paren leerlingen. Vervolgens werden per school twee vergelijkbare traininggroepen samengesteld, T1 en T2, door de leerlingen per paar aan een van beide toe te wijzen. Bij 5 koppels bleken de leerlingen dezelfde leerkracht te hebben en van hetzelfde geslacht te zijn. Zij werden door loting aan een groep toebedeeld. In de overige gevallen werden met de toewijzing de groepen op het punt van de leerkracht en op het punt van de sekse van de leerlingen zo vergelijkbaar mogelijk gemaakt. Uiteindelijk werd door loting bepaald welke groep in de periode maart-juni welke variant van de training zou krijgen. Helaas is kort na de start van onderzoek 4 een leerling verhuisd, zodat ten slotte met 37 leerlingen gewerkt is. Beschrijvende informatie is weergegeven in tabel 31, geen van de verschillen tussen de beide traininggroepen is significant.

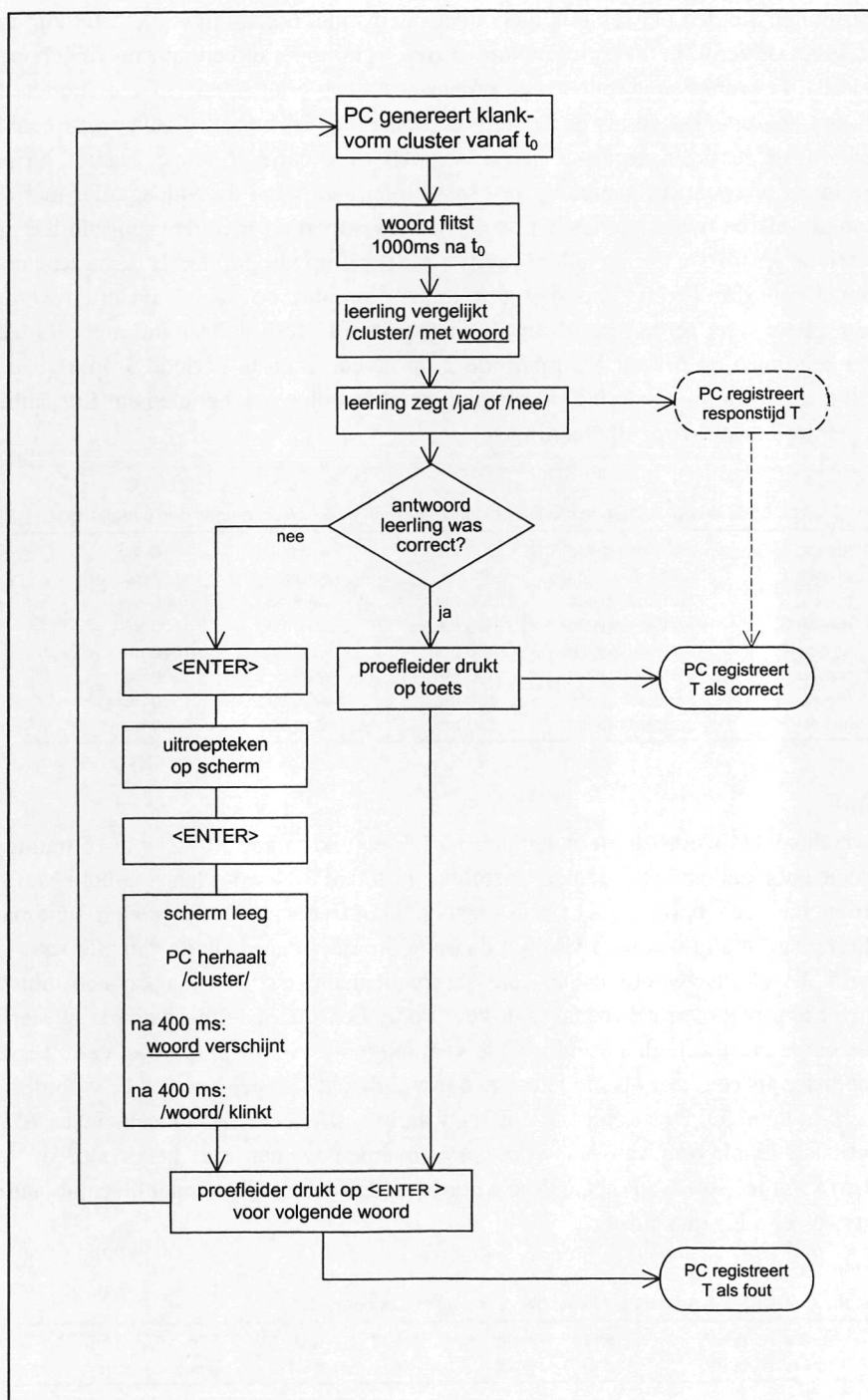
Tabel 31: Beschrijvende informatie over de proefpersonen van deelonderzoek 4.

Sekse	N	Jongens		Meisjes
traininggroep 1	19	12		7
traininggroep 2	18	12		6
SE-score	N	1	1,25	1,9
traininggroep 1	19	11	5	3
traininggroep 2	18	12	4	4
DMT-2, ruwe score	N	Min.	Max.	M (SD)
traininggroep 1	19	6	21	13,7 (4,8)
traininggroep 2	18	6	20	14,4 (4,1)

procedure

De training in deelonderzoek 4 vraagt de leerling een vergelijking te maken tussen een aangeboden klankvorm en de geschreven vorm van een woord. Het computerscherm is leeg, de PC genereert een klankvorm van bijvoorbeeld het foneemcluster /st/. Het scherm blijft leeg gedurende 1000 ms. Dan verschijnt een woord op het scherm, bijvoorbeeld staaf, gedurende de aanbiedingstijd van bijvoorbeeld 800ms. Na de aanbiedingstijd wordt het scherm weer leeg. De leerling heeft daarna maximaal 1500 ms de tijd om te beslissen of het gehoorde cluster in het woord voorkomt of niet. In 50% van de items staat het cluster inderdaad in het woord, in 50% van de items niet, bijvoorbeeld de combinatie van het klankcluster /gr/ met staaf. De PC registreert het moment waarop de leerling begint te praten, *speech onset*. De proefleider beoordeelt of de leerling de vergelijking juist maakte. Zo ja, dan volgde een registratie identiek aan die in de eerdere deelonderzoeken. Zo niet, dan verscheen op het scherm een groot uitroep-teken. Nadat de proefleider heeft vastgesteld dat de leerling zijn aandacht bij de taak heeft, drukt deze op een willekeurig toets, waarna de computer het gesproken klankcluster herhaalt, na 400 ms verschijnt het woord weer op het scherm en na nog eens 400 ms zegt de computer het hele woord. De trainingprocedure is schematisch weergegeven in figuur 12. Het programma voorzag ook in de mogelijkheid dat de leerling niets zou zeggen. Op vergelijkbare wijze als bij de eerdere deelonderzoeken werd dit gescoord als T-fout. Dit deed zich echter vrijwel niet voor, om het schema overzichtelijk te houden is deze mogelijkheid hierin niet opgenomen.

Bij elke sessie zijn de trainingwoorden bij wijze van voortoets in *random* volgorde aangeboden, aanbiedingstijd maximaal 3000ms, met de instructie niet te spellen, maar direct het hele woord te zeggen. De PC registreerde ook de reactietijd; 700 ms na *speech onset* verdween het woord van het scherm. De proefleider registreerde of het woord wel of niet goed gelezen was. Na beëindiging van de sessie, met het behalen van het criterium of omdat het maximaal aantal trials bereikt was, werd deze procedure herhaald bij wijze van natoets.



Figuur 12: Stroomschema van de woordtraining in deelonderzoek 4.

De resultaten werden per sessie bijgehouden. Items met begin- en eindcluster zijn niet apart geregistreerd. In het programma waren acht moeilijkheidsniveaus ingebouwd (tabel 32). Er waren twee tellers, een omhoog- en een omlaagteller. De omhoogteller werd met een opgehoogd als de leerling de sessie in zes, zeven of acht *trials* haalde. Haalde hij de sessie in negen of tien *trials*, dan werd dat niet geteld. Haalde hij een sessie in elf of twaalf of haalde hij een sessie niet, dan werd de omlaagteller met een verhoogd. Als de omhoogteller het aantal van twee bereikt had, dan ging de leerling naar een hoger niveau. Beide tellers werden dan weer op nul gezet. Als de omlaagteller het aantal van drie bereikt had, dan ging de leerling naar een lager niveau en werden beiden tellers weer op nul gezet. In trainingperiode 1 werd de training met alle leerlingen begonnen op niveau 1, in periode 2 op niveau 3 en in periode 3 op niveau 5. Behalve ander trainingmateriaal werd vanaf periode 2 dus ook het element flitsaanbieding geïntroduceerd voor alle leerlingen.

Tabel 32: Opbouw en opklimmende moeilijkheidsniveaus van de woordtraining in deelonderzoek 4.

niveau 0:	aanbiedingstijd	3500 ms,	wachttijd	0 ms
niveau 1:	aanbiedingstijd	2500 ms,	wachttijd	0 ms
niveau 2:	aanbiedingstijd	1500 ms,	wachttijd	1000 ms
niveau 3:	aanbiedingstijd	1000 ms,	wachttijd	1500 ms
niveau 4:	aanbiedingstijd	800 ms,	wachttijd	1500 ms
niveau 5:	aanbiedingstijd	600 ms,	wachttijd	1500 ms
niveau 6:	aanbiedingstijd	400 ms,	wachttijd	1500 ms
niveau 7:	aanbiedingstijd	200 ms,	wachttijd	1500 ms

inhoud

De tweemaal 324 woorden uit onderzoek 1 zijn gebruikt om de 24 sets van 16 trainingwoorden voor onderzoek 3 samen te stellen, in totaal 384 woorden. De helft van de woorden had een hoogfrequent cluster. Een hoogfrequent cluster kwam minimaal achttien maal in alle woorden voor en daarmee in meer dan de helft van alle sets. De andere helft van de woorden had een laagfrequent cluster. Een laagfrequent cluster kwam in alle woorden maximaal acht keer voor. Een uitzondering was het cluster st dat als enige cluster in elke trainingset zoveel mogelijk voorkwam, zowel een keer als begincluster als een keer als eindcluster. Een voorbeeld van een set van 16 woorden is gegeven in tabel 33. De afleiders (de items waarin het foneem, het foneemcluster of de aangeboden klankvorm van het woord *niet* overeenkwamen met het woord op het scherm) kwamen steeds uit het andere woord van de woordparen, zodat de combinaties slechts verschilden qua cluster.

Tabel 33: Voorbeeld van een set trainingwoorden uit deelonderzoek 4.

graaf	staaf	krijt	smijt	lukt	lust	fors	fort
grap	stap	krul	smul	mikt	mist	kaars	kaart

trainingcondities

Voor de training waren twaalf effectieve schoolweken beschikbaar, verdeeld in drie perioden van vier weken. Er is gewerkt met twee varianten; de volgorde foneem-cluster-woord is trainingvariant 1 genoemd (*tr1*), de volgorde woord-cluster-foneem trainingvariant 2 (*tr2*). In de tweede periode ontvingen alle leerlingen dezelfde training volgens hiervoor beschreven basisschema van de training met medeklinkerclusters. De volgorde binnen trainingvariant 1 komt overeen met die van een traditionele opbouw-methodiek zoals die in het aanvankelijk lezen gebruikelijk is; trainingvariant 2 is daarvoor een controleconditie.

In trainingperiode 1 kreeg de ene helft van de leerlingen, traininggroep T1, dezelfde trainingopzet, maar werd auditief niet een medeklinkercluster, in het voorbeeld /st/ of /gr/, aangeboden, maar een enkel foneem, /s/, /t/, /g/ of /r/. Vanzelfsprekend kwamen zowel het eerste als het tweede foneem van het cluster even vaak voor en in willekeurige volgorde. De andere helft van de leerlingen, traininggroep T2, kreeg in de eerste periode auditief het hele woord aangeboden, dus /staaf/ of /graaf/, en moest zeggen of dit het woord was dat daarna op het scherm verscheen. In trainingperiode 3 kreeg groep T1 die in trainingperiode 1 de enkele fonemen gekregen had nu het hele woord en andersom kreeg groep T2 die in trainingperiode 1 het hele woord kreeg, in trainingperiode 3 de enkele fonemen (tabel 34).

Tabel 34: Opbouw en opklimmende moeilijkheidsniveaus van de woordtraining in deelonderzoek 4.

trainingperiode 1	trainingperiode 1	trainingperiode 1
traininggroep T1: foneem, start op niveau 1	cluster, start op niveau 3	woord, start op niveau 5
traininggroep T2: woord, start op niveau 1	cluster, start op niveau 3	foneem, start op niveau 5

bepaling van trainingeffecten

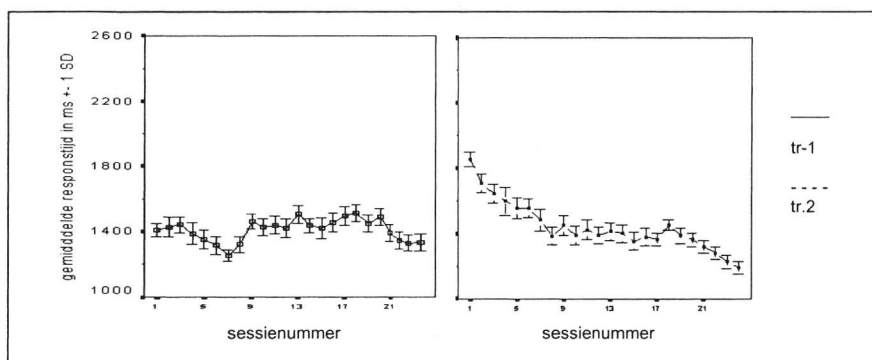
De maat voor de vlotheid van lezen in deelonderzoek 4 was om te beginnen de gemiddelde responstijd per trainingssessie. Maar omdat de training per periode zowel in taak als in aanbiedingsconditie verschilde en binnen periode 1 en 3 tevens voor de twee trainingsgroepen, maakt dat vergelijking tussen de perioden lastig. Van elke sessie waren echter ook twee metingen beschikbaar die wel over de hele onderzoeksperiode heen vergelijkbaar bleven, namelijk de voor- en natoets, waarvan in dit deelonderzoek naast de accuratesse ook de responstijd gemeten is, in totaal 48 herhaalde metingen per leerling. De natoets was een directe maat voor stimulusgeneralisatie als resultaat van de oefening in een sessie, de voortoets een maat voor responsgeneralisatie.

Naast de voor- en natoets per sessie is voorafgaand en na iedere trainingperiode een experimentele woordleestoets afgenomen. De toets bestond uit 56 woorden, de helft van de toetswoorden had het hoogfrequente cluster st of kt, dat ook in elke trainingset meestal tweemaal voorkwam. De andere helft van de woorden had een van de laagfrequente clusters lf, rt, pl of dr, die in de gehele trainingperiode slechts tweemaal voorkwamen. De toets bestond uit 24 trainingwoorden, acht uit elke periode. Daarnaast waren er 16 niet getrainde woorden en 16 nonsense woorden, elk met dezelfde clusters

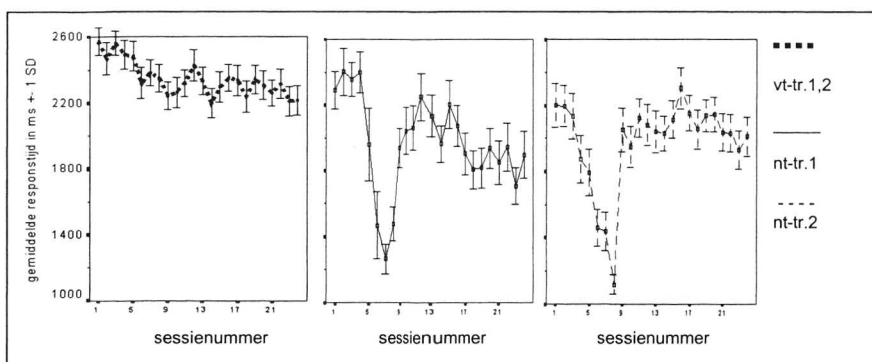
in dezelfde frequentie. De toetswoorden verschenen in een *random* volgorde een voor een op het computerscherm. 700 ms na de *speech onset* verdween het woord van het scherm. De computer registreerde de responstijd. De proefleider beoordeelde of de respons van de leerling juist was. Na maximaal 4000 ms verdween het woord van het scherm. De leerling kreeg op de toets geen feedback.

resultaten

Gemiddelde en spreiding van de responstijden op de trainingssessies zijn voor de twee trainingvarianten apart weergegeven in figuur 13. Gemiddelde en spreiding van de responstijden op de voor- en natoetsen zijn voor de twee trainingvarianten apart weergegeven in figuur 14.



Figuur 13: Deelonderzoek 4, gemiddelde en spreiding van de gemiddelde responstijden op de trainingssessies, voor trainingvariant 1 (tr-1) en trainingvariant 2 (tr-2) afzonderlijk.



Figuur 14: Deelonderzoek 4, gemiddelde en spreiding van de gemiddelde responstijden per sessie, op de voortoetsen voor beide trainingvarianten gezamenlijk (vt-tr.1,2) en op de natoetsen voor trainingvariant 1 (nt-tr.1) en trainingvariant 2 (nt-tr.2) afzonderlijk.

Allereerst is een multilevel analyse gemaakt van de responstijden binnen de training (tabel 35). Het verloop van de gemiddelde voorspelde responstijden is weergegeven in figuur 15. Trainingvariant 2, die begint met het vergelijken van hele woorden, is in de analyses de referentieconditie. De ontwikkeling van de responstijden wordt beschreven met een derdegraads functie met als fixed effect uitsluitend een significante negatieve derdemachtscomponent. Het resultaat hiervan is een daling van de responstijd in het begin van de training in periode 1, een afvlakking van de afname in periode 2 en ten slotte weer een versnelling van de afname in periode 3.

Tabel 35: Deelonderzoek 4, multilevel analyse van gemiddelde responstijden in de trainingssessies en op de voor- en natoetsen en aantal goed gelezen woorden op de voor- en natoetsen (standaard meetfout).

	responstijd (SE) trainingssessies	responstijd (SE) voor- en natoets	aantal goed (SE) voor- en natoets
Fixed effecten			
$\beta_{\text{intercept}}$	1.425,14 (39,77) ***	2.308,23 (108,62) ***	10,994 (0,653) ***
$\beta_{\text{sessienummer}}$		11,03 (5,41) *	-0,090 (0,039) *
$\beta_{\text{sessie-kwadraat}}$		1,31 (0,32) ***	
$\beta_{\text{sessie-derde macht}}$	-0,14 (0,03) ***	-0,17 (0,03) ***	0,00077 (0,00025) **
β_{natoets}		-229,22 (36,20) ***	1,420 (0,286) ***
$\beta_{\text{natoets} \times \text{sessie-kwadraat}}$			-0,011 (0,004) **
$\beta_{\text{natoets} \times \text{sessie-derde macht}}$		-0,14 (0,04) ***	0,00090 (0,00027) ***
$\beta_{\text{periode-1}}$			
$\beta_{\text{periode-3}}$			
$\beta_{\text{natoets} \times \text{periode-1}}$		-456,88 (60,30) ***	2,860 (0,458) ***
$\beta_{\text{natoets} \times \text{periode-3}}$			
$\beta_{\text{tr.variant-1} \times \text{periode-1}}$	-188,04 (37,28) ***		
$\beta_{\text{tr.variant-1} \times \text{periode-3}}$	64,09 (30,51) *		
$\beta_{\text{natoets} \times \text{tr.variant-1} \times \text{periode-1}}$		193,64 (74,94) **	
$\beta_{\text{natoets} \times \text{tr.variant-1} \times \text{periode-3}}$		-152,68 (60,52) **	1,107 (0,425) **
Random (niveau 2)			
$\text{var}[r_{\text{intercept}}]$	34.112 (8.223)	213.183 (50.740)	7,735 (1,866)
$\text{var}[r_{\text{sessienummer}}]$	581,15 (151,89)	216,18 (59,06)	0,00554 (0,00176)
$\text{var}[r_{\text{sessie-kwadraat}}]$	1,93 (0,52)	1,66 (0,60)	0,00007 (0,00003)
$\text{var}[r_{\text{sessie-derde macht}}]$	0,03 (0,01)		
$\text{var}[r_{\text{natoets}}]$		5.635 (2.866)	0,645 (0,239)
$\text{cov}[r_{\text{int.}, \text{sessie-kwadraat}}]$	-182,68 (56,34)		
$\text{cov}[r_{\text{sessienr.}, \text{sess-derd. macht}}]$	-3,51 (1,01)		
$\text{cov}[r_{\text{intercept}, \text{natoets}}]$			-1,494 (0,546)
Residu (niveau 1)			
$\text{var}[r_{\text{intercept}}]$	13.349 (992)	72.019 (3.914)	4,30 (0,155)
$\text{var}[r_{\text{sessienummer}}]$	9,08 (17,33)	131,79 (67,63)	
$\text{cov}[r_{\text{intercept}, \text{sessienummer}}]$	-365,34 (57,90)	-1.212 (221)	

*** $p < .001$ ** $p < .01$ * $p < .05$

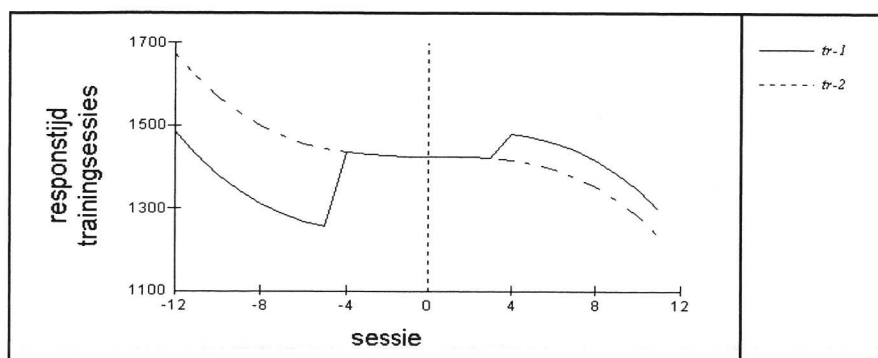
In periode 1 waarin het geschreven woord zichtbaar blijft, blijkt het vergelijken van een enkel foneem met een geschreven woord, trainingvariant 1, aanzienlijk gemakkelijker dan het vergelijken van de klankvorm van het hele woord, trainingvariant 2. In

periode 2 hield de training voor beide groepen leerlingen in dat de klankvorm van een medeklinkercluster vergeleken moest worden met een geschreven woord. De groepen verschilden dus slechts in de voorbereiding hierop uit periode 1; hiervan was geen onderscheidend effect aantoonbaar. Periode 3 verschilde van de eerste doordat de woorden geflitst werden. In periode 3 vraagt het vergelijken van hele woorden, nu in trainingvariant 1, gemiddeld 64 ms meer. Ook in periode 3 is het vergelijken van de klankvorm van hele woorden moeilijker dan het vergelijken van een enkel foneem, maar onder flitsaanbieding is het effect minder sterk.

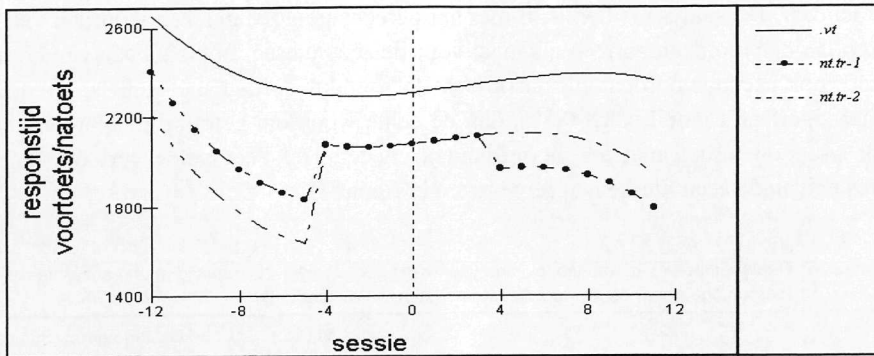
In de analyse van de responstijden op voor- en natoets is de voortoets in trainingperiode 2 bij trainingvariant 2 gedefinieerd als de referentieconditie. De gemiddelde responstijd in de referentieconditie was 2308 ms. De woorden van de natoets worden gemiddeld in alle condities 229 ms sneller gelezen. In trainingperiode 1 is er een extra natoetseffect van 263 ms voor de leerlingen van traininggroep T1 en van 457 ms voor de leerlingen van traininggroep T2. In periode 3 zien we hetzelfde effect, maar minder sterk. De leerlingen van traininggroep T1, die nu het hele woord krijgen aangeboden, maar met flitsaanbieding, lezen de woorden van de natoets gemiddeld 153 ms sneller. Traininggroep T2 laat ten opzichte van periode 2 geen extra leereffect zien.

De multilevel analyse is herhaald met het aantal goed gelezen woorden op de voor- en natoetsen als afhankelijke variabele. Het resultaat, weergegeven in de laatste kolom van tabel 35, geeft eenzelfde natoetseffect, maar in periode 1 geen specifiek effect van een trainingvariant. Het verloop van responstijden en toetsscore op voor- en natoetsen is weergegeven in figuren 16 en 17.

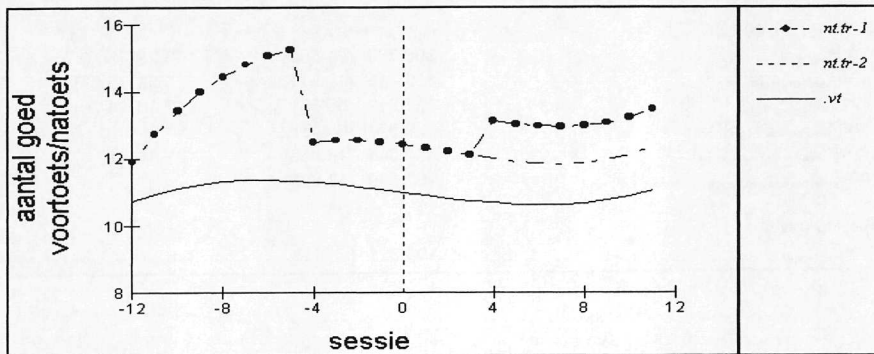
Er is aan het einde van de trainingperiode geen verschil tussen beide traininggroepen in leesvaardigheid. De gemiddelde DMT-2 score was 22,4 (SD 11,4) voor de leerlingen van traininggroep 1 en 22,9 (SD 7,7) voor de leerlingen uit traininggroep 2.



Figuur 15: Deelonderzoek 4, gemiddelde responstijd per sessie op de trainingssesies voor trainingvariant 1 (tr-1) en trainingvariant 2 (tr-2) afzonderlijk.



Figuur 16: Deelonderzoek 4, gemiddelde responstijden per sessie, op de voortoetsen voor beide trainingvarianten gezamenlijk (vt) en op de natoetsen voor trainingvariant 1 (nt-tr.1) en trainingvariant 2 (nt-tr.2) afzonderlijk.



Figuur 16: Deelonderzoek 4, gemiddeld aantal goed per sessie, op de voortoetsen voor beide trainingvarianten gezamenlijk (vt) en op de natoetsen voor trainingvariant 1 (nt-tr.1) en trainingvariant 2 (nt-tr.2) afzonderlijk.

Woordleestoets

De afhankelijke variabelen zijn de responstijd van de items die zowel geldig waren als correct gelezen en, vanwege de verschillende aantallen, het percentage goed gelezen woorden per subcategorie. Om de invloed van de aspecten clusterfrequentie, wel of niet getraind en woorden of pseudowoorden te kunnen vaststellen, zijn deze tien subsets afzonderlijk gescoord en leverden zo veertig herhaalde metingen per leerling op, die per tien eenzelfde sessienummer hebben gekregen.

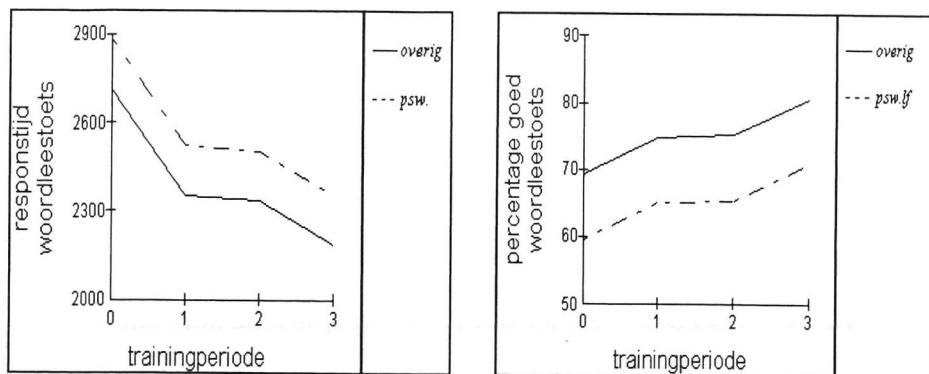
De uitkomst van de multilevel analyse is weergegeven in tabel 36. De getrainde woorden zijn gedefinieerd als referentieconditie. De gemiddelde responstijd was 2331 ms. Ten opzichte van de responstijd bij woorden, zowel getrainde als niet-getrainde, was het lezen van - niet getrainde - pseudowoorden moeilijker, er was gemiddeld 168 ms langer voor nodig. Er waren geen andere significante hoofdeffecten of interacties

aantoonbaar. De analyse is herhaald met het percentage goed gelezen woorden van de verschillende woordcategorieën als maat voor de accuratesse. In plaats van een hoofdeffect voor pseudowoorden, bleken alleen de pseudowoorden met een laagfrequent cluster significant moeilijker, -9,8%, dan de echte woorden, getraind of niet-getraind, of de pseudowoorden met een hoogfrequent cluster. Het voorspelde verloop van de scores en van de accuratesse is weergegeven in figuur 18.

Tabel 36: Deelonderzoek 4, multilevel analyse van gemiddelde responstijden en accuratesse (percentage goed) op de woordleestoets (standaard meetfout).

	responstijd (SE)	percentage goed (SE)
Fixed effecten		
$\beta_{\text{intercept}}$	2.330,74 (105,01) ***	74,96 (2,48) ***
$\beta_{\text{sessienummer}}$		
$\beta_{\text{sessie-kwadraat}}$	52,47 (25,93) *	
$\beta_{\text{sessie-derde macht}}$	-77,88 (31,47) *	1,70 (0,71) *
$\beta_{\text{pseudowoorden}}$	167,68 (21,20) ***	
$\beta_{\text{laagfrequent cluster x pseudowoorden}}$		-9,78 (2,37) ***
Random (niveau 2)		
$\text{var}[r_{\text{intercept}}]$	398.390 (94.650)	192,28 (50,05)
$\text{var}[r_{\text{sessienummer}}]$	153.411 (42.511)	15,08 (5,05)
$\text{var}[r_{\text{sessie-kwadraat}}]$	21.819 (5.758)	12,00 (4,78)
$\text{var}[r_{\text{sessie-derde macht}}]$	29.927 (8.445)	
$\text{cov}[r_{\text{intercept}}, \text{sessiekwadraat}]$	-50.044 (18.760)	-25,54 (12,47)
$\text{cov}[r_{\text{sessiekwadraat}}, \text{sessie-derde macht}]$	-62.699 (18.303)	
Residu (niveau 1)		
$\text{var}[r_{\text{intercept}}]$	104.821 (4.111)	326,77 (12,64)

*** $p < .001$ ** $p < .01$ * $p < .05$



Figuur 18: Deelonderzoek 4, responstijd en percentage goed gelezen woorden op de woordleestest; de afwijkende resultaten voor pseudowoorden (psw) respectievelijk laagfrequente pseudowoorden (psw.lf) zijn apart weergegeven.

conclusie

In de gerealiseerde training lijken zowel een focus op grafeem/foneemcombinaties als op het hele woord tot meer stimulusgeneralisatie te leiden dan een focus op het medeklinkercluster. Responsgeneralisatie is in geen van de varianten vastgesteld. De poging in deelonderzoek 4 een werkvorm te vinden waarmee ook de zwakste lezers zouden worden getraind in het verwerken van de dubbele medeklinkercluster lijkt niet gelukt en daarmee wordt de conclusie uit de eerdere deelonderzoeken versterkt, dat in het bijzonder de dubbele medeklinker voor de zwakste lezers de grootste moeilijkheid lijkt te vormen.

4.4 Discussie

De eerste drie deelonderzoeken waren in opzet vergelijkbaar, maar verschilden wat betreft inhoud- en trainingcondities. In alle gevallen werd leerlingen per trainingssessie een gesloten set van 12 woorden aangeboden die zij moesten oefenen, en in elke volgende sessie een nieuwe set woorden. De wijze van aanbidding was zodanig dat in principe reeds vanaf het begin een goede mate van accuratesse werd bereikt. In de training werd met de woorden verder geoefend teneinde een kortere responssnelheid te bereiken. De afhankelijke variabele voor de vlotheid van lezen was per trainingssessie de gemiddelde responssnelheid van de goed gelezen woorden.

In deelonderzoek 1 is gewerkt met woordensets van niveau 2, dat wil zeggen MKMM- en MMKM-woorden, die slechts in beperkte mate *friends* waren, ze overlappen alleen in het medeklinkercluster. De gemiddelde individuele leercurve, zowel op vlotheid als op accuratesse, kon worden beschreven als een tweedegraads functie; een geleidelijke afname van de trainingwinst met het voortduren van de training. Dit patroon is niet alleen kenmerkend voor trainingsonderzoek op het gebied van lezen, maar ook het standaardpatroon in leerexperimenten van uiteenlopende aard (Peterson 1965, Judd & Glaser 1969). Dat de leerlingen op een volgende nieuwe set woorden gemiddeld een lagere responstijd lieten zien, met behoud van accuratesse, kan worden geïnterpreteerd als indicatie voor de beoogde responsgeneralisatie. De uitkomst was voorts naar verwachting, in die zin dat leerlingen onderlinge verschillen lieten zien in de snelheid waarmee ze de trainingwoorden lazen én in de mate waarin ze hierin vooruit gingen. De verschillen tussen leerlingen namen in de loop van de training bovendien toe en deze differentiële instructiegevoeligheid was gerelateerd aan het bereikte niveau van leesvaardigheid aan het eind van de onderzoeksperiode. Leerlingen die op de training gemiddeld kortere responstijden hadden, waren de leerlingen die na afloop van de training ook de betere lezers bleken te zijn.

Het effect in deelonderzoek 1 van twee onafhankelijke variabelen in de vorm van systematisch gevarieerde randcondities, zijnde overlap in medeklinkercluster tussen twee trainingsets en aanbidding van de trainingwoorden in de vorm van een wisselrij, was, hoewel statistisch significant, marginaal. Zowel in responstijden als in de accura-

tesse van het lezen vinden we een - bescheiden - additioneel generalisatie-effect als training- en generalisatieset de medeklinkerclusters gemeen hadden.

De opzet is met jongere leerlingen in deelonderzoek 2 herhaald met woorden van niveau 1, dat wil zeggen MKM-woorden, die maximaal *friends* waren, het waren rijmwoorden die slechts één letter van elkaar verschilden. Een vergelijkbare leereffect tussen sessies was alleen aantoonbaar in de generalisatiesessies; hierin overlaptten de te lezen pseudowoorden met de daarvoor getrainde bestaande woorden. De leercurve nam over de sessies constant af en had niet de gebruikelijk kwadratische component waardoor de afname in de loop van de training afvlakt, zoals in deelonderzoek 1, maar het aantal trainingssessies in dit deelonderzoek was gering.

Het effect op de gemiddelde responstijd van het aanbieden van de trainingwoorden in de vorm van een wisselrij was in vergelijking met deelonderzoek 1 aanzienlijk. Woorden van niveau 1 lijken gemakkelijker gelezen te kunnen worden als ze worden aangeboden in de orthografisch redundante omgeving van een wisselrij. Tegen het aanbieden van trainingmateriaal in deze vorm is wel het bezwaar gemaakt, dat door de voorspelbaarheid het overlappende gedeelte minder aandacht van de leerling zou krijgen en hij dit gedeelte daardoor juist minder goed zou leren. Dit kon op MKM-niveau worden getoetst door de woorden in de generalisatiesessies zo samen te stellen dat de helft *friends* waren en de andere helft geen overlap vertoonden, maar wel uitsluitend uit getrainde letters bestond. De *friends* werden in de generalisatiesessies zowel sneller als accurater gelezen. Maar ook al op de voorttoetsen van de generalisatiesessies werden de woorden die de woordkern overeenkwamen met de juist daarvoor getrainde woorden-set, beter gelezen dan de woorden die wel de letters overeenkwamen, maar niet de woordkern. Daarmee lijkt het genoemde bezwaar bij training met woorden op MKM-niveau weerlegd.

Een positief extra generalisatie-effect kon met woorden op niveau-2 echter niet aangetoond worden, evenmin overigens als een negatief effect van deze aanbiedingsvorm. Levy (2001) vond wel een extra generalisatie-effect, zowel op accuratesse als op leessnelheid van, in het Engels, *blocked* aanbieding, maar minder retentie van de zo geoefende woorden. Het aanbieden van trainingwoorden in de vorm van een wisselrij lijkt minimaal een manier om de moeilijkheidsgraad van trainingmateriaal te reduceren en is als zodanig toepasbaar als orthodidactische maatregel.

Naast het gemiddelde niveau van de leerlingen op de training, de intercept, en de daling van de responstijden, hangt in deelonderzoek 2 ook het pseudowoordeffect samen met de leesvaardigheid van de leerlingen aan het einde van de trainingperiode. In hoofdstuk 4.2 veronderstelden we dat responsgeneralisatie vrijwel alleen aangetoond is binnen taken die noodzakelijk alleen via decoding op subwoordniveau kunnen worden uitgevoerd. Aangezien de generalisatiesets uit pseudowoorden bestonden, is daarvan per definitie sprake. Deze differentiële instructiegevoeligheid kan daarom worden geïnterpreteerd als een verschil in responsgeneralisatie en als een indicatie voor verwerking van het trainingmateriaal op subwoordniveau.

Het relatieve generalisatie-effect tussen trainingsets van woorden die op subwoord-niveau overlaptten, was in deelonderzoek 2 iets groter dan in deelonderzoek 1. Het verschil in stimulusmateriaal in de beide deelonderzoeken kan op twee manieren worden benoemd. De overlap was relatief groter, de woorden verschilden slechts een tegenover twee letters, en de overlap betrof tevens de klinker. Aangehaalde onderzoeken waarin via training op subwoordniveau ook met woorden boven MKM-niveau generalisatie kon worden bereikt (Frederiksen e.a. 1985, Yap 1993, Van Daal 1993, Das-Smaal e.a. 1996), blijken ook steeds de klinker in de subwoordeenheid betrokken te hebben, evenals in deelonderzoek 2. Hiervan is in deelonderzoek 3 gebruik gemaakt door introductie van een tussenniveau van trainingwoorden van niveau-2 die maximaal *friends* waren door overlap van medeklinkercluster + klinker (sessie 1-20). Na een fase die overeenkwam met de opzet van deelonderzoek 1 waarin alleen het medeklinkercluster overlapte (sessie 21-40), werd de training afgerond met trainingwoorden die in het geheel niet meer overlaptten (sessie 41-60). Met andere woorden, er werden drie moeilijkheidsniveaus in de training geïntroduceerd, waarbij met alle leerlingen in vergelijking met deelonderzoek 1 op een gemakkelijker niveau met de training begonnen. De verwachting was dat het patroon van deelonderzoek 1, namelijk verschillen tussen leerlingen in de leessnelheid, in de mate van vooruitgang en een toename van de verschillen tussen leerlingen in de loop van de training, door deze differentiatie in moeilijkheidsgraad en een groter aantal sessies sterker naar voren zouden komen.

Het patroon op het traject van sessie 21-60 is redelijk vergelijkbaar met de uitkomst van deelonderzoek 1. Het meest opvallende en niet verwachte resultaat lijkt echter de sterke invloed te zijn van de eerste overgang in het stimulusmateriaal van sessie 20 naar sessie 21. De relatie tussen de training en de leesvaardigheid aan het einde van de trainingperiode lijkt door twee aspecten bepaald te worden: de *response to treatment* zoals die ook in deelonderzoek 1 naar voren kwam, waarbij de drie afzonderlijke maten intercept, lineaire en kwadratische component in dit deelonderzoek sterk blijken samen te hangen, en de reactie op het loslaten van de overlap in medeklinkercluster + klinker. Het eerste aspect komt neer op de getoonde leesvaardigheid in de specifieke trainingssituatie. Maar vooral het effect van de overgang van trainingwoorden die overlaptten in medeklinkercluster + klinker (KMM) naar overlap in alleen medeklinkercluster (MM) op de prestaties van de leerlingen is groot en lijkt op twee manieren het verschil te markeren tussen de leerlingen die aan het einde van de trainingperiode niet langer als uitvallers worden gezien en zij die dan nog steeds uitvallen op technische leesvaardigheid. De zwakste lezers reageren veel sterker, in negatieve zin, op deze overgang en lijken zich, althans binnen de context van de training, ook niet te herstellen.

In termen van de dynamische systeemtheorie lijken de leerlingen aanvankelijk relatief goed op de training te reageren. De overgang van KMM naar MM betekent dan een plotselinge verhoging van de taakeisen (perturbatie). Een deel van de leerlingen past zich daar vrij snel op aan en zet de ontwikkeling door van korter wordende responstijden, een ander deel vindt een nieuw evenwicht op een prestatieniveau met blijvend significant hogere responstijden (bifurcatie).

Bij woorden in de eerste 20 sessies die een maximale overlap vertonen, is de intrawoordredundantie groter dan vanaf sessie 21. In de eerste serie trainingssessies is derhalve sprake van een relatief sterke beperking van het aantal responsalternatieven en deze wordt daardoor gemakkelijker. De zwakste lezers blijven op dit traject wel achter bij de relatief betere lezers, maar komen niet steeds verder achterop, zoals wel gebeurde bij deelonderzoek 1 waarin de overlap beperkt was tot het medeklinkercluster. Het is echter de vraag of dit het gevonden effect voldoende kan verklaren. De overgang van sessie 40 naar sessie 41 is in dat opzicht immers veel groter. Om die reden zou men ook een interactie verwachten van deze overgang met de factor 'zwak'. Maar het effect van het verschil tussen een beetje *friends* (zelfde MM) en geen overlap (verschillende MM) op de gemiddelde responstijden is veel geringer dan dat van de overgang van sessie 20 naar sessie 21 en lijkt ook niet onderscheidend tussen de relatief betere en relatief zwakste lezers. Daarbij moet wel worden aangetekend dat in deze laatste analyse het aantal zwakste lezers gehalveerd was, zie figuur 10.

Mogelijk is het niet alleen het aantal overlappende grafemen, maar speelt de klinker in de subwoordeenheid een specifieke rol. Volgens de *perceptual centre* hypothese (Goswami e.a. 2002) correspondeert juist de klinker binnen de lettergreep met een snelle toename van energie in de middenfrequenties, waarvan accurate detectie om die reden gemakkelijker is dan van medeklinkers. Welke de functie is van de klinker binnen de lettergreep en hoe die mogelijk kan worden benut in het uitlokken van generalisatie, is een fascinerende vraag voor vervolgonderzoek.

In de deelonderzoeken 1 en 3 bleken woorden met een eindcluster gemakkelijker te lezen dan woorden met een begincluster. Hieraan zijn verschillende interpretaties gegeven. Een verklaring zou kunnen liggen in de intrawoordredundantie. De voorspelbaarheid van een grafeem neemt immers toe met de positie ervan binnen een woord. Bovendien zijn er minder MKMM- dan MMKM-woorden én hebben MKMM-woorden meer *friends* waarmee ze het medeklinkercluster delen. Van Daal (1993) vond eenzelfde eindclustereffect in een experiment waarin sprake was van auditieve *priming* van de te lezen woorden door voorafgaande aanbieding van de klankvorm van de KMM/MMK-woordkern. Assink e.a. (1998) vonden een *priming* effect van medeklinkerclusters als visuele herkenningseenheden. Beperking van het aantal responsalternatieven zou daarom ook een deel van de verklaring voor dit eindclustereffect kunnen zijn. Mogelijk speelt ook de in de vorige alinea aangegeven *perceptual centre* hypothese een rol, als bijvoorbeeld de energietoename die het begin van de klinker markeert, in kortere tijd plaatsvindt na een enkele medeklinker dan na een medeklinkercluster en daarom gemakkelijker is waar te nemen.

Uit de resultaten van de eerste drie deelonderzoeken lijkt het medeklinkercluster het moeilijkste niveau van verwerking te vertegenwoordigen. Trainen met woorden op MKM-niveau die de KM-woordkern overeenkomen blijkt gemakkelijker en er treedt generalisatie op, ook bij de zwakste lezers. Op MKMM/MMKM-niveau blijven de allerzwaksten aanzienlijk minder achter bij de relatief minder zwakke lezers als trainingwoorden niet alleen het MM-cluster gemeen hebben, maar ook de klinker. Het niet

kunnen maken van de stap van MKM-woorden naar MKMM- en MMKM-woorden lijkt een slechte prognose in te houden voor de mogelijkheid een grote leesachterstand in te lopen. Tevens lijken materiaalkenmerken voor de zwakste lezers erg bepalend voor de moeilijkheidsgraad van een leestaak.

De poging in deelonderzoek 4 om ook de zwakste lezers effectief met medeklinkerclusters als eenheid van verwerking te laten werken door middel van een matchingtaak, lijkt niet gelukt. In de analyses is geen onderscheid gemaakt tussen de leerlingen die aan het einde van de trainingperiode nog steeds op E-niveau scoorden en de relatief beter lezers, volgens de norm voor eind groep 4, omdat deelonderzoek 4 niet tot doel had door middel van de training onderscheid te maken tussen de leerlingen op basis van gebleken instructiegevoeligheid, maar om het effect van enkele aanbiedingscondities nader te exploreren. Introductie van een factor 'zwak' leverde in alle gevallen een significant zwakkere gemiddelde prestatie voor dit deel van de leerlingen. Voorts alleen nog een significante interactie bij het aantal goed gelezen woorden van 'zwak' met de natoetsscore. Het totaalbeeld van de resultaten van deelonderzoek 4 werd derhalve door het opnemen van een factor 'zwak' niet betekenisvol genuanceerd. De reden hiervoor is waarschijnlijk dat, waar deze onderverdeling in de andere deelonderzoeken leidde tot groepen van ongeveer gelijke grootte, in deelonderzoek 4 slechts een kwart van de leerlingen aan het eind van de onderzoeksperiode niet als uitvaller scoorde.

Het vergelijken van hele woorden is in vergelijking tot het vergelijken van een enkel foneem weliswaar moeilijker (analyse 1), maar leidt wel tot een groter leereffect op de getrainde woorden (analyse 2). Een differentieel effect ten gunste van trainingvariant 2 in periode 1, als er geen sprake is van flitsen, wordt wel gevonden op de responstijden, maar niet op het aantal goed gelezen woorden. Mogelijk is sprake van een plafond-effect op de natoets (analyse 3). Door de verschillende analyses en variatie in aanbiedingconditie heen lijkt één conclusie zich op te dringen: het werken met medeklinkerclusters is moeilijker en levert zelfs op de getrainde woorden minder leereffect op, dan elke andere conditie. Dit uit zich in significant geringere leereffecten in periode 2, zoals blijkt uit de verschillen tussen voor- en natoets in de verschillende periodes. Op de voortoetsen, bedoeld als maat voor responsgeneralisatie, is geen leereffect vastgesteld. Opvallend is dat van het vergelijken van een enkele medeklinker uit een medeklinkercluster onder de conditie van beperkte aanbiedingsduur geen leereffect op woordniveau lijkt uit te gaan, terwijl dit bij onbeperkte aanbiedingsduur wel zo is. Ook lijkt het leereffect bij onbeperkte aanbiedingsduur groter dan bij flitsaanbieding. Ten slotte lijkt training met hele woorden op die woorden een groter leereffect op te leveren dan training op subwoordniveau. Wellicht is decoderen op woord- of subwoordniveau allebei mogelijk, en meer afhankelijk van specifieke taakeisen dan dat er sprake is van een bepaalde kwaliteit van het leesproces een algemene zin. Zo is bijvoorbeeld een frequentie-effect alleen onder zeer specifieke condities aantoonbaar: alleen de niet-geoefende pseudowoorden met een laagfrequent medeklinkercluster werden op de woordleestoets die parallel aan het onderzoek werd afgenomen, minder accuraat gelezen.

In de vier deelonderzoeken lijkt het medeklinkercluster het moeilijkste niveau van verwerking te vertegenwoordigen, moeilijker dan verwerking van hele woorden. De opbouw die in alle moderne methoden voor aanvankelijk lezen gevolgd wordt, loopt van kleinste naar grootste eenheid, dus:

- van volledige verklanking van grafemen;
- naar grotere subwoordeenheden;
- naar uiteindelijk het hele woord.

De resultaten van de verschillende deelonderzoeken lijken daarentegen eerder te suggereren dat in de ontwikkeling van het technisch lezen de volgorde lijkt te zijn:

- van volledige verklanking van grafemen;
- naar directe woordherkenning;
- naar verwerking op subwoordniveau.

Responsgeneralisatie betekent in beide volgordes verwerking van subwoordeenheden binnen nieuwe woorden. Het verschil is de weg waarlangs representaties van eenheden groter dan het grafeem, maar kleiner dan het hele woord worden opgebouwd. Ehri (1998) spreekt over een *consolidated alphabetic phase* na de *full alphabetic phase* die staat voor directe woordherkenning, die gebaseerd is op een volledige verbinding tussen de grafemen in het geschreven woord en de fonemen in de klankvorm van het woord. In deze fase zouden ook bepaalde lettersequenties geconsolideerd raken, als resultaat van herhaalde ervaringen ermee in het lezen, maar gebaseerd op die volledige verbindingen uit de *full alphabetic phase*. Zij geeft als voorbeeld het herkennen van de subwoordeenheid est na voldoende ervaring met nest, pest, test, vest. Deze volgorde spoort met de opvatting dat responsgeneralisatie binnen het lezen loopt via verwerking van subwoordeenheden. Bij responsgeneralisatie zien we dat gedeeltelijk nieuw gedrag mogelijk wordt, als variatie op en aanpassing van gedrag dat reeds in het repertoire aanwezig is. Een gedeeltelijk nieuwe respons die gebaseerd is op een nieuwe combinatie van elementen van beschikbare responsen, zodanig dat de nieuwe respons is afgestemd op de eisen die de nieuwe situatie stelt. De zwaarte van een leesstoornis zou kunnen worden gezien als de mate waarin met name die laatste stap wel of niet mogelijk is of de mate waarin deze zich, onder instructiecondities, kan ontwikkelen. Zo wordt ook inzichtelijk waarom het leren lezen juist voor zwakke lezers meer *item-based* lijkt te zijn, wat welbeschouwd ook een andere manier is om te zeggen dat er geen (respons)generalisatie optreedt.

Het is een geruststellende gedachte dat de gevolgde didactiek voor het gros van de leerlingen niet veel uitmaakt voor het niveau dat zij uiteindelijk bereiken; bij leerlingen bij wie de technische leesontwikkeling zonder noemenswaardige problemen verloopt, treedt de gewenste generalisatie toch wel op. Maar de leesontwikkeling van zwakke lezers is, bijna per definitie, juist afhankelijk van een didactiek die de ontwikkeling leidt. Zij doen over alle stappen langer en generaliseren minder. Mogelijk zijn dit twee aspecten van hetzelfde mechanisme. Ook zwakke lezers kunnen leren om woorden te ontsleutelen via verklanking van grafemen en auditieve synthese. Ook kunnen zij woorden direct leren herkennen, al hebben ze daar meer trials voor nodig. Volgens

Adams (1990) is het kenmerk van goede lezers dat het decoderen van woorden volledig geautomatiseerd en gelijktijdig op alle mogelijke niveaus plaatsvindt. Dat goede lezers een woord gelijktijdig op verschillende niveaus verwerken, zou betekenen dat verschillende routes bij hen een continuüm vormen. Op afzonderlijke routes kan men wel zicht krijgen, maar alleen door stimuluscondities te creëren die verschillende routes afdwingen, bijvoorbeeld het werken met onbekende woorden, zoals pseudo-woorden. Als goede lezers in staat zijn bij voor hen nog onbekende woorden hun taakaanpak zo af te stemmen dat zij het woord op subwoordniveau verwerken, dan houdt dat de gewenste generalisatie in. Als zwakke lezers daar niet of veel minder toe in staat zijn, dan oefenen zij daardoor ook niet het vehikel voor de (respons)generalisatie en dat bij zwakke lezers vooral de kwaliteit van de fonologische representatie van woorden tekortschiet, is daarvan dan zowel oorzaak als gevolg.

De consequentie van het voorgaande zou zijn dat oefenen op subwoordniveau met eenheden groter dan het grafeem geen tussenstap is tussen grafeem en hele woord, maar volgt op het kunnen lezen van het hele woord. Oefeningen op subwoordniveau leiden niet tot directe herkenning van dat woord, maar zouden gebaseerd moeten zijn op woorden die reeds als eenheid direct herkent kunnen worden, leidend, als het goed gaat, tot snellere verwerving van directe herkenning van nieuwe woorden.

tot slot

Het paradigma van de ontwikkelingspsychopathologie legt nadruk op de complexe manier waarop aanleg en omgevingsfactoren op elkaar inspelen, leidend tot verschillende ontwikkelingstrajecten. De dynamische systeemtheorie schetst hoe het individu in dit samenspel voor tekorten oplossingen vindt. Het nemen van zijn toevlucht tot verwerking van woorden op het niveau van de kleinste eenheden, de grafemen, of op het niveau van het hele woord, terwijl de taak een andere aanpak vraagt, zijn voorbeelden van zulke oplossingen, zoals bij de bespreking van spellend en radend lezen is opgemerkt. De woordtraining zoals gerealiseerd in deelonderzoek 3 als onderdeel van intensieve remedial teaching biedt, tegen deze achtergrond, uitbreiding en intensivering van de leer- en instructietijd waarbinnen de beoogde generalisatie zal kunnen optreden bij die leerlingen die op een eerder moment in hun leesontwikkeling van het juiste spoor zijn afgeraakt zijn, terwijl hun mogelijkheden toch voldoende zijn om het te redden met wat het onderwijs te bieden heeft. Aan de andere kant voorziet het traject zo in het onderscheiden van de leerlingen bij wie dat desondanks niet lukt, wat een indicatie is voor dyslexie. Een van de problemen met onderzoek op het gebied van leerstoornissen is dat de leergeschiedenis van de proefpersonen vaak niet of onvoldoende bekend en mogelijk zeer uiteenlopend is. Leerlingen die uit dit traject naar voren komen, vormen in dat opzicht een relatief homogene groep, van wie de leergeschiedenis in detail bekend is.

Interventieonderzoek met leerlingen die van het nu gerealiseerde traject onvoldoende hebben geprofiteerd, kan gegevens opleveren omtrent de behandelmogelijkheden van dyslectische leerlingen. Behandeling zal erop gericht moeten zijn die omstandigheden

te creëren waarbinnen leerlingen uitgelokt worden juist die wijzen van verwerking van woorden te oefenen die zij onder normale condities vermijden. Het woord vermijden heeft wellicht de connotatie van een bewuste gehanteerde strategie; dat is niet bedoeld. Het is slechts - letterlijk - de weg van de minste weerstand, waarmee leerlingen precies niet dat gedrag vertonen, dat zij zouden moeten oefenen. Aanknopingspunten zullen moeten worden gevonden in, in gedragstherapeutische termen, een nauwgezette taak- en functieanalyse.