

Verdiepen van woordkennis: een empirische studie naar de effecten van een trainingsprogramma

M. Verhallen, R. Schoonen en R. Appel

Samenvatting

In deze studie wordt gerapporteerd over een empirisch onderzoek naar de effectiviteit van een experimenteel lesprogramma ter verbetering van de zogenaamde diepe woordkennis van zwakke leerlingen in het basisonderwijs. Uit het onderzoek blijkt dat de diepe woordkennis van de getrainde leerlingen aantoonbaar verbetert ten opzichte van die van een niet-getrainde controlegroep. Een generalisatie naar niet-geoefende domeinen in het lexicon kon niet worden aangetoond, evenmin als een effect op de leesvaardigheid.

1 Inleiding

1.1 Woordkennis in het onderwijs

Voorbeeld 1

De kinderen in groep 5 werken aan de volgende opgave uit hun rekenboek:

Peter en Mark eten boterhammen. Peter eet er 4 op, dan lust hij niet meer. Mark eet 6 boterhammen. Hij schept op: "Ik kan 2 keer zoveel eten als jij". Peter zegt: "Dan moet je nog meer boterhammen eten." Hoeveel boterhammen moet Mark nog eten?

Mahmut steekt zijn vinger op en zegt:

Juf, ik snap het niet! Boterhammen kun je toch niet opscheppen?

Het bovenstaande voorbeeld (Voorbeeld 1) laat zien dat zelfs in de rekenles woordkennis een cruciale rol speelt. Onvoldoende woordkennis kan het onderwijsleerproces sterk frustreren.

Woordkennis is een voorwaarde voor schoolsucces: de leerkrachten en de leerboeken gebruiken woorden om kennis over te dragen, ze verwoorden denkstappen, ze geven woord voor woord uitleg en brengen redeneringen en conclusies onder woorden. In de klas komen woorden steeds meer

centraal te staan als verschijnselen en gebeurtenissen beschreven en verklaard worden die zich elders in tijd en ruimte bevinden. Leerlingen moeten zelf het woord voeren om antwoorden te formuleren en om deel te nemen aan de verbale interactie in de klas. Bijna alle leerstof is verpakt in woorden en leerlingen moeten over voldoende woordkennis beschikken om te kunnen participeren in het onderwijs. Zowel bij het kringgesprek in de onderbouw van het basisonderwijs, als bij het lezen van een zaakvaktekst in de bovenbouw en het voortgezet onderwijs is woordkennis een vereiste: leerlingen die de gebruikte woorden niet voldoende kunnen begrijpen, kunnen de les niet goed volgen en de tekst van de leerboeken niet volledig begrijpen.

Het onderwijs stelt hoge eisen aan woordkennis, maar tegelijkertijd is datzelfde onderwijs een belangrijke bron voor de opbouw van woordkennis. De verbale interactie in de klas en het lezen van teksten over allerlei onderwerpen geven dag in dag uit mogelijkheden voor uitbreiding van woordkennis. Woorden worden niet alleen expliciet geleerd (door uitleg van de leerkracht of door opzoeken in een woordenboek), maar worden voor een groot deel ook impliciet verworven uit de context (Nagy, Herman & Anderson, 1985, Nagy, Anderson & Herman, 1987, zie ook Coady, 1997 voor een toepassing op tweedetaalverwerving). Op school doen kinderen in elke les, bij elke activiteit ervaringen op met nieuwe woorden, nieuwe betekenisaspecten en nieuwe betekenisrelaties. Bijvoorbeeld in één enkele les over magneten komen leerlingen niet alleen nieuwe woorden tegen zoals *magnetisme*, maar ook nieuwe betekenisaspecten van bijvoorbeeld het woord *kracht*. Dat is in deze context niet in alledaagse zin aan de orde - in de betekenis van spierkracht -, maar het moet hier in "academische" zin worden begrepen als natuurkundige kracht. Verder komen nieuwe betekenisrelaties tussen woorden aan bod als kinderen ontdekken

dat ijzer, nikkel en tin alle drie *metalen* zijn, die door een magneet worden aangetrokken, terwijl *aantrekken* hier tegengesteld is aan *afstoten* en niet aan *uittrekken*.

Kinderen kunnen in verschillende onderwijsleersituaties hun woordkennis in allerlei richtingen uitbreiden: zodra er een nieuw onderwerp aan de orde komt, leren zij de woorden, betekenissen en betekenisrelaties van het betreffende semantisch domein, mits ze de taal begrijpen en de les kunnen volgen. Als de taal te veel over hun hoofden heen gaat, blijven deze gelegenheden tot woordenschatuitbreiding onbenut. Het bekende *Mattheus-effect* (Stanovich, 1986) treedt dan op. Kinderen die veel woorden en betekenissen kennen, zullen eerder nieuwe woorden en betekenissen bijleren en inpassen in wat ze al weten. Leerlingen die over minder woordkennis beschikken, zullen minder snel nieuwe woorden en betekenissen oppikken; de achterstand wordt groter.

De instroom van allochtone kinderen in het Nederlandse onderwijs heeft de aandacht voor lexicale ontwikkeling sterk doen toenemen en tot nieuwe impulsen geleid in taalonderwijs en taalonderzoek. Het accent ligt daarbij op het onderwijzen en leren van nieuwe woorden (verbreding van woordkennis). *Verdieping* van de woordkennis krijgt in het onderwijs nog weinig aandacht. Onder gunstige omstandigheden ontwikkelt diepe woordkennis zich geleidelijk “vanzelf” in schoolse situaties, maar onderzoek heeft uitgewezen dat met name allochtone kinderen ook hier achterblijven bij hun klasgenoten (zie par. 1.2). Die achterstanden kunnen belangrijke consequenties hebben voor het schoolsucces.

De vraag is nu of diepe woordkennis door gerichte systematische interventie onderwezen kan worden. Daarom is nieuw experimenteel lesmateriaal ontwikkeld waarin het onderwijzen van diepe woordkennis is geoperationaliseerd. In een empirisch onderzoek zijn de effecten van het remediërend onderwijsmateriaal geëvalueerd.

In dit artikel beschrijven we de operationalisatie van diepe woordkennis in deeltaken en de opzet van een lesprogramma (par. 1.2) en vervolgens rapporteren we de opzet en

resultaten van het effectonderzoek (par. 2, respectievelijk 3).

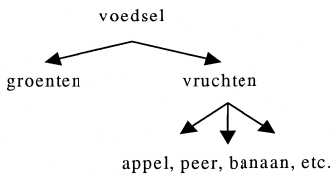
1.2 Ontwikkeling van diepe woordkennis in het basisonderwijs

In het onderwijs en onderzoek gelden autochtone, ééntalige kinderen vaak als referentiepunt. In het algemeen is het curriculum afgestemd op het niveau van een doorsnee Nederlandstalige leerling. Het is daarom zinvol om eerst een beeld te schetsen van de lexicale ontwikkeling van Nederlandse kinderen tijdens de basisschooljaren, zodat duidelijk wordt wat de uitgangspunten van het onderwijs zijn als het gaat om woordkennis.

Kleuters komen op school met een receptieve woordenschat van ruim 3000 woorden (Verhoeven & Vermeer, 1989). Ze breiden geleidelijk hun woordkennis uit. Belangrijk is dat woordkennis niet alleen kwantitatief toeneemt, in die zin dat er steeds meer woorden worden bijgeleerd en de omvang van de woordenschat groeit, maar dat ook de kwaliteit van woordkennis verandert. In de loop der jaren verdiept de woordkennis zich: de betekenisstoekenning en de opbouw van het lexicale netwerk worden verrijkt: kinderen leren steeds meer betekenisaspecten en er komen allerlei nieuwe verbindingen tussen woorden tot stand. In de kleutergroep gaat het nog om de appel op het plaatje in het prentenboek waar *Rupsje Nooitgenoeg* zich dwars doorheen eet, maar in de loop der jaren moeten kinderen leren dat een appel een *klokhuis* met *pitjes* heeft en dat die pitjes *zaadjes* zijn waaruit een nieuwe appelboom kan groeien.

In allerlei leersituaties verbreedt én verdiept de betekenisstoekenning zich. Naast alledaagse syntagmatische betekenisrelaties als *appel* → *eten* moeten kinderen steeds vaker andersoortige, abstractere relaties zoals *appel* → *vrucht* hanteren. Het gaat hier om paradigmatische relaties, die in de hogere groepen van het basisonderwijs de achterliggende structuur vormen bij de opbouw van veel schoolse kennis (zie voor een uitgebreide uitleg ook Verhallen, 1994; Schoonen & Verhallen, 1998a). In de lexicale en cognitieve ontwikkeling is de verwerving van paradigmatische relaties een belangrijk gegeven (zie bijvoorbeeld Snow, 1990). Met paradig-

matische relaties kunnen hiërarchische structuren binnen het semantisch netwerk worden opgebouwd. In een hiërarchische structuur worden woorden door coördinatie, subordinatie en superordinatie in een formele structuur geplaatst: principes van logische ordening en categorisering zijn hierbij aan de orde. Generalisatie en abstractie komen binnen bereik. De betekenisaspecten van hoger geordende woorden gelden voor alle onderliggende woorden, en overeenkomsten en verschillen krijgen een plek in een logisch verband. De betekenisaspecten van voedsel (=eetbaar) zijn per definitie overdraagbaar op alle onderliggende woorden, maar appels verschillen op bepaalde essentiële kenmerken van andere soorten voedsel. Elk woord krijgt een plaats binnen een uitgewerkte logische structuur.



Figuur 1. Diagram van voedsel.

Bij de ordening van schoolse kennis spelen paradigmatische relaties een belangrijke rol. Als alles goed gaat, leren kinderen dergelijke betekenisrelaties steeds bewuster te hanteren. In de loop der jaren worden dergelijke structuren steeds meer ingevuld en verfijnd. Uiteindelijk moeten kinderen opdrachten als de volgende uit een methode voor wereldoriëntatie kunnen maken.

Hoe hoger je op een berg komt, hoe minder loofbomen je ziet groeien. Er is op de berg een gebied dat zo koud is, dat alleen naaldbomen daar kunnen groeien. Welke bomen zouden daar wel kunnen groeien:

- esdoorn kastanje spar beuk

Hier is duidelijk de inzet van diepe woordkennis vereist: kinderen moeten weten dat de spar onder de categorie *naaldbomen* hoort. Dit voorbeeld laat overigens ook zien dat er een nauwe relatie is tussen de verwerving van conceptuele kennis en de verwerving van diepe woordkennis. Anders geformuleerd: het betreft conceptuele kennis die in woorden

wordt uitgedrukt en die ook deel uitmaakt van de betekenis van woorden. Het ligt voor de hand dat leerlingen vooral van dit soort kennis profiteren bij leesactiviteiten. Met name bij begrijpend lezen gaat het immers vaak om onderlinge relaties in een tekst. Onderzoek heeft ook laten zien dat diepe woordkennis een specifieke bijdrage levert aan de voorspelling van leesprestaties (Bekius & Simis, 1997; Qian, 1999, Schoonen & Verhallen, 1998b).

Opvallend is dat er in de meeste taalmethodes nauwelijks oefeningen te vinden zijn die toegesneden zijn op verdieping van woordkennis. Integendeel: woorden komen op z'n best als losse eenheden, hapsnap aan de orde en staan meestal zonder uitleg onder elkaar. Vaak vinden we in handleidingen zinnen als "Behandel de moeilijke woorden", waarmee de kous ook af is. De leerkracht krijgt zelden handreikingen voor de uitleg van woorden, of tips voor rijke betekenisomschrijvingen. Het inoefenen van woorden in logische betekenisrelaties, systematische betekenisuitbreiding, controle op verwerving; er is weinig van dit alles terug te vinden in de lesjes (zie ook Appel & Verhallen, 1998).

Versillen tussen allochtone en autochtone leerlingen

Bij veel allochtone kinderen verloopt de lexicale ontwikkeling minder geleidelijk. Zij komen op school met een woordenschat in hun eigen taal, die vanaf de eerste dag in het Nederlandse onderwijs meestal onbenut blijft. Ze kennen (veel) minder Nederlandse woorden dan Nederlandse kinderen en moeten in de tweede taal als het ware een lexicale inhaalmanoeuvre uitvoeren. Onderzoek naar woordenschat heeft inderdaad uitgewezen dat tweede-taalleerders in dit opzicht systematisch achterblijven bij hun eentalige klasgenootjes (zie bijvoorbeeld Vermeer, 1997).

Uit onderzoek naar diepe woordkennis blijkt dat daarnaast de kwaliteit van woordkennis van tweetalige kinderen verschilt van die van eentalige kinderen. Onderzoek van Verhallen (1994) bracht aan het licht dat betekenisstoekenning van negen- en elfjarige Turkse kinderen op belangrijke punten afwijkt van die van Nederlandstalige klasge-

nootjes. In een uitgebreid interview moesten kinderen al hun kennis van betekenisaspecten van een woord expliciteren. Turkse kinderen gaven met betrekking tot zowel het aantal als het soort betekenisaspecten blijk van een achterstand: ze noemden opvallend minder betekenisaspecten en de betekenisstoekenning was minder paradigmatisch en meer syntagmatisch van aard. Terwijl er oppervlakkig gezien geen verschillen waren (alle kinderen kenden de voor het onderzoek geselecteerde woorden), kwamen er op dieper liggend niveau belangrijke achterstanden aan het licht. De interviews werden ook in het Turks afgenomen om na te gaan of deze achterstanden in de tweede taal door lexicale kennis in de eerste taal werden gecompenseerd. Dit bleek niet het geval. Integendeel: in het Turks gaven kinderen nog minder betekenisaspecten bij de Turkse equivalente woorden en de betekenisomschrijvingen waren nóg minder paradigmatisch dan in het Nederlands (Verhallen, 1994; Verhallen & Schoonen, 1998).

Een studie van Vermeer (1997) wees eveneens uit dat tweetalige vijf- en zesjarige kinderen op een definitietaak en een associatietask minder goed scoren. Zie verder voor een overzicht van verschillen tussen allochtone en autochtone leerlingen Appel (1999).

De vergelijkingen laten zien dat anderstalige kinderen woorden minder diep kennen dan hun Nederlandse klasgenootjes; het onderwijs zou daarom extra aandacht moeten geven aan betekenisstoekenning en verdieping van woordkennis in het kader van onderwijskansen en verbetering van schoolsucces van anderstaligen. Een intensief programma is ontwikkeld om de woordkennis van kinderen die achterblijven, in korte tijd te verdiepen.

Van Woorden Weten: een trainingsprogramma voor verdieping van woordkennis

Het programma Van Woorden Weten¹ is ontwikkeld om woordkennis systematisch te verdiepen (zie Verhallen & Walst, 1997, 1999). Stap voor stap leren kinderen bewust om te gaan met woordbetekenissen en wordt de overgang van algemeen dagelijkse naar meer schoolse betekenisstoekenning gemaakt. Ordenen van begrippen, hanteren van betekeniskenmerken en definiëren van woorden staan centraal. Het programma is remedië-

rend van aard, dat wil zeggen, bestemd voor aanvullend onderwijs voor leerlingen van wie de diepe woordkennis tekortschiet. We bespreken hieronder eerst de voornaamste uitgangspunten van het programma. Daarna presenteren we het stramien voor verdieping van woordkennis en de operationalisatie van het proces in verschillende verwervingsstappen.

Het programma is bestemd voor leerlingen van 8 tot 10 jaar. Het materiaal is niet bedoeld voor de uitbreiding van het aantal woordvormen, maar voor de verbetering van de kwaliteit van de onderliggende of diepere kennis van woorden. In het programma wordt gewerkt met drie thema's: dieren, voertuigen en beroepen. De algemeen didactische uitgangspunten zijn:

- *Stapsgewijze opbouw van eenvoudig naar complex*

Het materiaal is opgebouwd in kleine stappen van eenvoudig naar moeilijk. In de eerste stap zijn de taaltaken volledig ingebed in non-verbale contexten, door middel van het gebruik van foto's en tekeningen. Gaandeweg valt deze inbedding weg, waardoor de activiteiten steeds meer worden gedecontextualiseerd. De tweede lijn loopt van receptief naar productief taalgebruik. Eerst ordenen kinderen foto's en kaartjes met woorden, pas later beargumenteren zij hun ordeningen en voeren ze schrijftaken uit.

Niet alleen binnen de lescyclus is er sprake van een oplopende moeilijkheidsgraad, ook de thema's gaan van eenvoudig naar complex. Het thema *dieren* is het meest concreet en bekend, het tweede thema (*voertuigen*) is al wat moeilijker doordat dit semantische domein meer hiërarchische indelingen en meer verschillende betekenisrelaties kent terwijl het derde thema (*beroepen*) het meest abstract is.

- *Intensieve begeleiding en directe feedback*

Bij het programma spelen de leerkrachten een belangrijke rol. Ze oriënteren kinderen op de inhoud van de taak, geven duidelijke instructies, stimuleren de kinderen tot reflectie op hun handelen en geven directe feedback.

- *Coöperatief leren en zelfstandig werken*
Alle opdrachten worden zelfstandig en in tweetallen uitgevoerd. Het werken in tweetallen bevordert onderling overleg. De opdrachten lokken discussie uit over woordbetekenissen.

Ten behoeve van een systematische aanpak volgt het programma een didactisch model met drie fasen: ordenen van begrippen, hanteren van kenmerken en definiëren. Zo brengt het programma de leerlingen geleidelijk dichterbij bij bewuste, abstracte, formele betekenisgeving. Binnen de drie fasen wordt een aantal stappen onderscheiden. Elke stap kent een vast patroon: introductie door de leerkracht, gevolgd door instructie, zelf werken - in groepjes van twee - en nabespreken van de resultaten. De leerkracht stimuleert de kinderen zo veel mogelijk om op elkaars werk te reflecteren. Ter illustratie volgt hieronder een korte beschrijving van de eerste fase en de daarin gevolgde stappen (zie voor meer informatie Verhallen & Walst, 1999).

Deze eerste fase, het ordenen van begrippen, loopt van spontaan ordenen tot hiërarchisch ordenen. Om te beginnen: in stap 1 werken de kinderen met beeldmateriaal. Ze ordenen samen foto's bij het eerste thema van dieren, volgens hun eigen indeling, bijvoorbeeld op basis van informele criteria als: *lieve dieren, enge dieren, zachte dieren*. De manier waarop leerlingen deze opdracht uitvoeren biedt de leerkracht enig zicht op hun voorkennis. Bij stap 2 werken ze met andere, meer gerichte ordeningen. Zo bekijken ze samen met de leerkracht een tekening waarin ordeningsfouten verwerkt zijn: op een tekening van een dierentuin staan bijvoorbeeld een leeuw en een lam in dezelfde ruimte. Kinderen halen de fouten uit de tekening en geven aan waarom de tekening niet klopt. In de derde stap volgt de overgang naar het rubriceren van beeldmateriaal in formele categorieën. Kinderen groeperen foto's van dieren onder familienamen: vogels, vissen, insecten, roofdieren, knaagdieren.

In de volgende stap gebruiken ze in plaats van foto's woordkaartjes. Bij de stappen 5 en 6 zijn de taaltaken nog meer gedecontextualiseerd en wordt de productieve woordkennis aangesproken. Leerlingen vullen samen

werkbladen in. Eerst noteren ze woorden onder het kopje van een familienaam net zoals ze met de foto's en de woordkaartjes deden, en uiteindelijk komen ze in stap 6 tot het invullen van een boomdiagram, vergelijkbaar met het bovenstaande diagram van voedsel (Figuur 1).

Het programma is bedoeld als prototype en heeft in eerste instantie een remediërende doelstelling. Door gebruik van het materiaal raken leerkrachten vertrouwd met het stappenplan en de diepe woordkennisactiviteiten. De verwachting is dat zij de gevolgde werkwijze op den duur zelf ook toepassen bij de instructie van andere woorden of semantische domeinen, bijvoorbeeld *gebouwen*, en dat ze dat ook in gewone lessen gaan doen.

In het algemeen moet bij woordenschatuitbreiding altijd worden bedacht dat het om grote aantallen woorden gaat. Het onderwijs beschikt echter over te weinig tijd om al die woorden aan te bieden. Als verbreding van de woordenschat het doel is, dus het verwerven van nieuwe woorden en woordbetekenissen, wordt daarom ook vaak voorgesteld om leerlingen woordleerstrategieën te leren gebruiken (zie bijvoorbeeld Van Daalen-Kapteijns, Elshout-Mohr & De Gloppe, 2001; Verhallen en Verhallen, 1994). Met die strategieën kunnen ze zelf onbekende woorden verwerken en toevoegen aan hun woordenschat, zonder dat er specifieke instructie nodig is. Bij verdieping van woordenschat geldt ook dat het onderwijs onvoldoende tijd heeft om dit voor een groot aantal thema's of domeinen te bewerkstelligen. Het idee achter het hier onderzochte programma is dan ook dat leerlingen in bepaalde domeinen leren om formele, hiërarchische relaties te leggen en dat ze dit vervolgens ook gaan doen in andere, niet behandelde domeinen. Anders geformuleerd: de verwachting bestaat dat transfer gaat optreden van de in de lessen behandelde domeinen naar andere domeinen.

1.3 Onderzoeksvragen

De algemene vraagstelling van het onderzoek betreft het effect van het experimentele programma Van Woorden Weten op de diepe woordkennis van kinderen. De eerste en belangrijkste onderzoeksvraag gaat over het ef-

fect van het programma op de diepe woordkennis in de geofende domeinen. De vraag kan als volgt worden geformuleerd:

- 1 Heeft het experimentele trainingsprogramma een positief effect op de diepe woordkennis van de kinderen in de geofende domeinen?

Aan de hand van de tweede onderzoeksvraag wordt nagegaan of er inderdaad sprake is van de verwachte transfer van de diepe woordkennis. De onderzoeksvraag kan als volgt worden geformuleerd:

- 2 Heeft het experimentele trainingsprogramma een positief effect op de diepe woordkennis van de kinderen in de *niet*-geofende domeinen?

Tenslotte is er de veronderstelde relatie tussen diepe woordkennis en vaardigheid in begrijpend lezen. In verband met deze relatie zullen we de volgende onderzoeksvraag beantwoorden:

- 3 Heeft het experimentele trainingsprogramma een positief effect op leesvaardigheid van kinderen?

2 Methode van onderzoek

2.1 Proefpersonen

Het lesprogramma richt zich op kinderen uit de midden- en bovenbouw van het basisonderwijs (8-10 jaar). Het effectonderzoek is uitgevoerd bij kinderen uit groep 5 van basisscholen uit de regio Haarlem/Amsterdam.

Tijdens de werving van scholen bleek er een ongekend grote belangstelling voor deelname aan het onderzoek, zodat er meer scholen waren dan beschikbaar experimenteel lesmateriaal. Om scholen niet teleur te stellen zijn de scholen aselekt in twee groepen (A en B) verdeeld, die in twee opeenvolgende rondes met het programma mochten werken. Daarnaast is er een derde groep (C) geworven; op deze laatste groep scholen werden alleen de toetsen afgenomen, zonder dat er met het programma gewerkt is. Hoe de verschillende groepen vergeleken kunnen worden, wordt toegelicht in de sectie Onderzoeksontwerp.

In totaal zijn ruim 1500 leerlingen uit 67 groepen-5 van 55 scholen bij het onderzoek betrokken geweest. Binnen elke klas zijn op basis van een voormeting en deels in overleg

met de leerkracht ca. vijf leerlingen geïdentificeerd als relatief zwakke leerlingen als het gaat om diepe woordkennis. Deze leerlingen vormen de doelgroep van het lesprogramma. Naar deze kinderen zullen we in het vervolg verwijzen als *zwakke leerlingen*. De zwakke leerlingen hebben met het lesprogramma gewerkt (groepen A en B) of fungeerden als controlegroep (C). De overige leerlingen uit een klas zullen we in het vervolg aanduiden met *klasgenoten*. De klasgenoten hebben dus geen van allen met het lesprogramma gewerkt, maar zij hebben om praktische redenen wel meegedaan met de toetsafnamen.

De evaluatie van het lesprogramma richt zich dus op de prestaties van de zwakke leerlingen. Hun klasgenoten zullen alleen bij de analyses betrokken worden als het gaat om een referentiepunt of controles op aanvaardbare (groeps)verschillen.

Niet alle leerlingen hebben alle toetsen op de verschillende tijdstippen kunnen maken (zie ook Procedure). De groepsgrootte voor de zwakke leerlingen varieerde voor de verschillende groepen (A, B of C) en toetsen tussen de 52 en 92; voor de klasgenoten varieerde die tussen de 216 en 364. Per analyse zal de exacte steekproefgrootte gerapporteerd worden, maar in alle gevallen kan aangenomen worden dat de omvang voldoende was om een acceptabel onderscheidingsvermogen te hebben (>.80; zie Cohen, 1988).

2.2 Instrumenten

In het onderzoek zijn feitelijk twee constructen geoperationaliseerd, namelijk *diepe woordkennis* (onderzoeksvraag 1 en 2) en *leesvaardigheid* (onderzoeksvraag 3). Bij diepe woordkennis is een onderscheid gemaakt naar diepe woordkennis in domeinen uit het lesprogramma (onderzoeksvraag 1) en diepe woordkennis in het algemeen; d.i. in niet-geofende domeinen (onderzoeksvraag 2). Een effect van het programma zou zich in de eerste plaats moeten openbaren in een verbetering van de diepe woordkennis op de geofende domeinen. Vervolgens willen we kunnen vaststellen of de verdieping van de woordkennis generaliseert naar andere domeinen. Diepere woordkennis zou uiteindelijk het lezen van (zaakvak)teksten moeten faciliteren.

Voor de operationalisatie van diepe woordkennis is gebruikgemaakt van de Woord Associatie Toets (WAT, zie Schoonen & Verhallen, 1998a). In de WAT moeten kinderen bij een stimuluswoord drie geassocieerde woorden kiezen uit een set van zes woorden. Daarbij gaat het erom dat zij onderscheid kunnen maken tussen toevallige en gedecontextualiseerde, systematische associaties.

Voor de algemene diepe woordkennis is een WAT gebruikt zoals beschreven in Schoonen en Verhallen (1998a); voor de diepe woordkennis over de geoefende domeinen van het vocabulaire is volgens hetzelfde stramen een toets geconstrueerd waarbij stimulus- en associatiewoorden gebruikt zijn uit de domeinen *dieren*, *voertuigen* en *beroepen*. De toets is overigens niet beperkt tot de woorden die expliciet in het programma aan de orde kwamen. Deze variant noemen we voortaan C-WAT (curriculumgerelateerd). De WAT bestaat uit 30 items, de C-WAT uit 36.

Voor de operationalisatie van de leesvaardigheid is gebruikgemaakt van clozetoetsen. Voor de voormeting is een clozetoets Leesvaardigheid groep 5 van het Cito gebruikt: de toets *Wat ben jij?* (Cito, 1990). In deze toets is na de beginzin elk zevende woord weggelaten en moet het kind op basis van zijn tekstbegrip dit woord *raden*. Deze test bleek voor de kinderen relatief moeilijk en enigszins demotiverend te zijn. Daarom is voor de volgende meetmomenten een zelf ontwikkelde clozetoets (BOS) gebruikt, waarin inhoudswoorden zijn weggelaten. De tekst handelt over bossen en is een tekst zoals die in zaakvakmethodes voorkomen. De Citocloze bestaat uit 41 items, de BOScloze uit 28.

2.3 Onderzoeksonderwerp

Omdat het lesprogramma in twee onderzoeksrondes is aangeboden, is het onderzoek in zekere zin tweevoudig van aard. Elke ronde is geëvalueerd. In de eerste ronde was er sprake van scholen die zich allemaal vrijwillig hadden aangemeld voor deelname mét het lesprogramma; deze scholen zijn aselekt toegewezen aan groep A of B. De zwakke kinderen uit groep A kregen het lesprogramma, de zwakke kinderen uit groep B niet. In

de eerste ronde is er dus sprake van een “echt” experimentele opzet (Kerlinger, 1973). Groep B werd experimentele groep in de tweede ronde en de apart geworven groep C fungeerde nu als controlegroep. Groep A was niet meer bij de tweede ronde betrokken (zie Figuur 2). Omdat er in de tweede ronde geen aselekte toewijzing was, is er hier sprake van een quasi-experimenteel onderzoeksontwerp. Uit Figuur 2 blijkt ook dat groep B mogelijk een licht voordeel heeft ten opzichte van groep C, omdat zij de WAT al een keer eerder gemaakt hebben.

groep	mm1	ronde 1	mm2	ronde 2	mm3
			C-WAT		
A	WAT	<i>programma*</i>	WAT		
	Citocloze		BOScloze		
			C-WAT		C-WAT
B	WAT	<i>geen programma</i>	WAT	<i>programma*</i>	WAT
	Citocloze		BOScloze		BOScloze
					C-WAT
C			WAT	<i>geen programma</i>	WAT
			BOScloze		BOScloze

* Alleen de zwakke leerlingen hebben met het lesprogramma gewerkt, hun klasgenoten niet.

Figuur 2. Onderzoeksonderwerp voor het effectonderzoek met twee rondes en drie meetmomenten (mm). Per meetmoment zijn de afgenomen toetsen weergegeven.

2.4 Procedure

De WAT en clozetoetsen zijn (onder schooltijd) afgenomen door geïnstrueerde proefleiders volgens een nauwkeurig vastgelegd protocol. De C-WAT is met een instructie achtergelaten bij de leerkracht. Na afname van de C-WAT stuurde de leerkracht de toetsen in. Niet alle leerkrachten zijn er in geslaagd om op tijd de C-WAT af te nemen / in te sturen, zodat er voor de C-WAT iets meer uitval is.

Het effectonderzoek vond plaats in het voorjaar 1998. De tweede ronde zat relatief dicht tegen de zomervakantie aan waardoor sommige leerkrachten (uit groep B) niet het volledige lesprogramma konden afronden

en/of niet de C-WAT konden afnemen, hetgeen voor meer uitval in de tweede ronde zorgde dan in de eerste.

De leerkrachten die met het lesprogramma gingen werken, kregen een uitgebreide instructie tijdens een scholingsdag (voor elke ronde apart). Tijdens het eerste dagdeel werden ze geïnformeerd over het begrip diepe woordkennis, de opbouw van diepe woordkennis en de mogelijke achterstanden die kinderen kunnen oplopen in diepe woordkennis. Het tweede dagdeel was gereserveerd voor instructie over het programma aan de hand van een demonstratievideo, de handleiding en het lesmateriaal. De leerkrachten werd ook verzocht een logboekje bij te houden waarin ze de feitelijke uitvoering van het lesprogramma konden rapporteren.

2.5 Scoring

De (C-)WAT kan op twee manieren gescoord worden: één punt voor elke goed gekozen associatie, waardoor per item 0-3 punten behaald kan worden, of een goed/fout-scoring per item waarbij een item pas goed gerekend wordt als alle drie de beoogde associaties aangestreept zijn. In de praktijk blijken de beide scoringswijzen onderling hoog te correleren (Schoonen & Verhallen, 1998a). Omdat de eerste scoringswijze per definitie sensitiever is dan de tweede, is in dit effectonderzoek voor de eerste scoringswijze gekozen.

De clozetoetsen zijn streng gescoord in de zin dat alleen antwoorden die overeenkomen met het weggelaten woord goed gerekend zijn. Andere, mogelijk plausibele antwoorden zijn fout gerekend. Hoewel dit oneerlijk lijkt tegenover de kinderen, blijkt veelal dat beide scoringswijzen hoog correleren. De coulantere scoringswijze leidt uiteraard tot hogere scores voor de kinderen, maar introduceert tevens beoordelingsproblemen. In de scoring van de clozetoetsen zijn spelfouten genegeerd.

2.6 Analyses

Per meetmoment wordt eerst per toets de betrouwbaarheid gerapporteerd.

De effectiviteit van het lesprogramma is in twee analyses geëvalueerd: een per onderzoeksrunde (zie Figuur 2). Binnen zo'n

analyse is gebruik gemaakt van covariantieanalyse. Covariantieanalyse biedt de mogelijkheid om (statistisch) voor beginverschillen te controleren, hetgeen met name in de tweede, quasi-experimentele ronde van belang is. Bovendien wordt in een covariantieanalyse de foutenvariantie verkleind, waardoor er "scherper" getoetst kan worden (Stevens, 1992)².

De analyses zullen worden uitgevoerd op de scores van de zwakke leerlingen uit de experimentele groep en uit de controlegroep. Als extra controle en vergelijking zullen ook de scores van de betreffende klasgenoten worden geanalyseerd. Mochten er tussen deze klasgenoten uit de experimentele en controlegroep ook verschillen optreden dan kan dat niet aan het materiaal toegeschreven worden, omdat de klasgenoten niet met het materiaal gewerkt hebben. We moeten er dan rekening mee houden dat er ook andere verschillen tussen de condities zijn opgetreden: bijvoorbeeld ontwikkelingen op de scholen of bij de leerkrachten die niet alleen de onderzoekskinderen betreffen, maar de gehele klas of school.

3 Resultaten

3.1 Betrouwbaarheid

Allereerst is nagegaan of de toetsen betrouwbare metingen opleverd hebben. De geschatte betrouwbaarheden voor de toetsen zijn steeds gebaseerd op de kinderen die alle items gemaakt hebben. Voor de WAT 30, voor de curriculumafhankelijke C-WAT 36, voor de Citoclozetoets (op meetmoment 1) 41 en voor de BOSclozetoets 28. Zoals uit Tabel 1 blijkt, zijn alle betrouwbaarheden .82 of hoger en alleszins acceptabel. De toetsen kunnen als voldoende betrouwbaar beschouwd worden om eventuele effecten van het programma aan te tonen³.

3.2 Resultaten eerste onderzoeksrunde: groep A versus groep B

Zoals uit Figuur 2 blijkt, hebben we drie afhankelijke variabelen; het gaat om de verschillen op de C-WAT, de WAT en de BOSclozetoets op meetmoment 2. Voor de C-WAT en de WAT geldt de voormeting van

diepe woordkennis meetmoment 1 met de WAT als covariaat, voor de BOSclozetoets geldt de Citoclozetoets op meetmoment 1 als covariaat.

Tabel 1
Betrouwbaarheidsschattingen volgens Cronbachs α , per toets per meetmoment (mm)

	mm 1	mm 2	mm 3
WAT ($k=30$)	.83	.83	.82
	$n=942$	$n=1314$	$n=796$
	A & B	A, B & C	B & C
C-WAT ($k=36$)		.86	.86
		$n=766$	$n=716$
		A & B	B & C
Citocloze ($k=41$)	.86		
	$n=966$		
	A & B		
BOScloze ($k=28$)		.83	.88
		$n=1325$	$n=809$
		A, B & C	B & C

In deze vergelijking zijn de zwakke leerlingen uit groep A de experimentele groep en de zwakke leerlingen uit groep B de controle-groep.

We verwachten vooral een effect op de C-WAT; deze toets be vraagt de drie domeinen die in het lesmateriaal aan de orde kwamen. In tweede instantie verwachten we een effect op de WAT; het gaat dan om de generalisatie naar diepe woordkennis in het algemeen (andere domeinen). Ten slotte hopen we effect te zien op de BOScloze; het gaat dan om de toepassing in een algemene leestaak.

Per afhankelijke variabele wordt een analyse voor de zwakke leerlingen gerapporteerd als toetsing van de effectiviteit van het programma en een analyse voor de klasgenoten als controle.

Effecten op de geoefende domeinen (C-WAT)

In Tabel 2 worden de gemiddelde scores voor de voormeting (WATmm1) gerapporteerd, evenals de gemiddelde scores op de nameting voor de geoefende domeinen (C-WATmm2). Bij aanvang van het onderzoek zijn de groepen nagenoeg gelijk; zowel de zwakke leerlingen die in aanmerking komen voor het

programma, als hun klasgenoten. Dat de zwakke leerlingen daadwerkelijk relatief zwak zijn, blijkt uit de vergelijking met de scores van hun klasgenoten; die liggen gemiddeld zo'n acht punten hoger.

Op de nameting (C-WATmm2) zijn de scores voor de zwakke leerlingen die met het programma gewerkt hebben duidelijk hoger dan voor hun controlegroep van zwakke leerlingen (groep B). Dit effect is niet alleen statistisch significant ($F(1,162)=41.68, p=.000$), maar ook aanzienlijk ($\eta^2=.205$)⁴. De zwakke leerlingen die met het programma gewerkt hebben, bereiken haast het niveau van hun klasgenoten. De klasgenoten uit de twee groepen verschillen ook bij de nameting nauwelijks van elkaar, zodat we mogen aannemen dat tussen de voor- en nameting de klassen als geheel niet bijzonder veranderd zijn. De scores van meetmoment 1 (WATmm1) en meetmoment 2 (C-WATmm2) kunnen niet rechtstreeks vergeleken worden omdat beide toetsen een andere schaal hebben: 0-90 respectievelijk 0-108.

Tabel 2
Gemiddelde scores op de covariaat van meetmoment 1 (WATmm1) en de met covariaat gecorrigeerde afhankelijke variabele op meetmoment 2 (C-WATmm2), met de toetsing van de significantie ($F(df1, df2)$ en p) en proportie verklaarde variantie door conditie (η^2)

	WATmm1 (covariaat, max. 90)	C-WATmm2 (gecorrigeerd, max. 108)	Toetsing
<i>Zwakke leerlingen</i>			
Groep A ($n=87$)	71.36	94.61	$F(1,162)=41.68$
Groep B ($n=78$)	70.67	87.82	$p=.000, \eta^2=.205$
<i>Klasgenoten</i>			
Groep A ($n=238$)	79.01	95.56	$F(1,546)=0.071$
Groep B ($n=311$)	79.16	95.40	$p=.790, \eta^2=.000$

Ter nadere exploratie is nagegaan of er verschillen zijn tussen de drie thema's. De curriculumafhankelijke C-WAT is opgesplitst in drie deelscores: één voor elk thema. Op elk van die themascores is eveneens een covariantieanalyse uitgevoerd.

Het blijkt dat het hierboven gerapporteerde effect op de curriculumafhankelijke toets bij de zwakke leerlingen niet toe te schrijven

is aan één thema. Op alle drie de thema's kunnen significante effecten gerapporteerd worden ($F=19.33$, $F=26.18$ resp. $F=36.82$, alle $p=.000$). Het lijkt er zelfs op dat er een toenemend effect is van het programma, $\eta^2=.107$, $.139$ respectievelijk $.185$. In hoeverre er sprake is van een cumulatief effect van de training of van een groter effect bij een moeilijker thema is met deze gegevens niet vast te stellen (zie Discussie).

Effecten op de niet-geoefende domeinen (WAT)

Op de scores voor de niet-domeingebonden diepe woordkennistoets (WAT) is een zelfde analyse uitgevoerd als hiervoor gerapporteerd voor de C-WAT. In Tabel 3 staan de resultaten. De aantallen wijken iets af van die in Tabel 2, omdat niet altijd alle leerlingen aanwezig waren bij de afnames (zie Procedure).

Tabel 3

Gemiddelde scores op de covariaat van meetmoment 1 (WATmm1) en de covariaat gecorrigeerde afhankelijke variabele op meetmoment 2 (WATmm2), met de toetsing van de significantie ($F(df1,df2)$) en proportie verklaarde variantie door conditie (η^2)

	WATmm1 (covariaat, max. 90)	WATmm2 (gecorrigeerd, max. 90)	Toetsing
<i>Zwakke leerlingen</i>			
Groep A ($n=91$)	70.58	75.30	$F(1,160)=1.14$
Groep B ($n=72$)	70.93	74.33	$p=.288$, $\eta^2=.007$
<i>Klasgenoten</i>			
Groep A ($n=275$)	78.64	80.42	$F(1,608)=1.30$
Groep B ($n=336$)	79.06	80.01	$p=.255$, $\eta^2=.002$

Ook nu blijkt dat de leerlingen uit groep A en B op de voormeting (WATmm1) gemiddeld even hoog scoren. Op de nameting scoren de getrainde zwakke leerlingen (groep A) weliswaar iets hoger dan de controlegroep van zwakke leerlingen, maar dit verschil is niet statistisch significant ($F(1,160) = 1.14$, $p=.288$) en zeer klein (zie noot 4). Het verschil is wel iets groter dan het verschil dat er tussen de klasgenoten geconstateerd kan worden.

De zwakke leerlingen (uit beide groepen) blijven duidelijk achter op hun klasgenoten,

hoewel ze meer vooruitgang lijken te boeken. Deze (grotere) vooruitgang kan echter een artefact van de onderzoeksofzet zijn. Het gaat hier namelijk om leerlingen die geselecteerd zijn op hun zwakke prestaties tijdens de voormeting; door regressie naar het gemiddelde is het waarschijnlijk dat zij bij een volgende meting hoger zullen scoren.

Effecten op de leesvaardigheid (BOScloze)

Als de effecten van het programma (nog) niet of nauwelijks naar andere domeinen van het mentale lexicon gegeneraliseerd zijn, mogen we zeker ook nog geen effecten verwachten op de leesvaardigheid, temeer daar de clozeleestekst over bossen andere domeinen van het lexicon aanspreekt dan die geoefend zijn.

In Tabel 4 staan de resultaten van de covariantieanalyse van de scores op de clozeleestoeetsen. De scores op de voor- en nameting zijn niet rechtstreeks te vergelijken door de verschillende schalen van de toetsen.

Op de voormeting scoren de zwakke leerlingen uit de beide groepen nagenoeg gelijk; op de nameting is een klein, statistisch niet-significant verschil in het voordeel van de getrainde kinderen. Het is zeer de vraag of dit verschil verband houdt met de diepe woordkennislessen, omdat ook de klasgenoten van de getrainde kinderen uit groep A enige vooruitgang boeken ten opzichte van hun vergelijkingsgroep, de klasgenoten uit groep B. De klasgenoten hebben geen van allen met het lesprogramma gewerkt, dus de verschillen hebben een andere oorzaak, of moeten toegeschreven worden aan toeval.

3.3 Resultaten tweede onderzoeksrunde: groep B versus groep C

In de tweede ronde, waarin we de resultaten van de eerste ronde proberen te repliceren, is groep B de experimentele groep en groep C controlegroep. Alleen de zwakke leerlingen uit groep B krijgen les met het experimentele lesprogramma. Deze leerlingen zullen vergeleken worden met de relatief zwakke leerlingen uit groep C. De deelnemende klassen uit groep C zijn geworven voor alleen de toetsafnames op meetmoment 2 en meetmoment 3; groep B en C zijn dus niet aselekt samengesteld en het belang van correctie voor eventuele beginverschillen middels een cova-

riaat neemt dus toe. Voor het overige zijn de analyses voor deze ronde gelijk aan die uit de vorige en gaat het om dezelfde drie afhankelijke variabelen: C-WAT, WAT en BOScloze op meetmoment 3 (zie Figuur 2).

De voormeting wordt weer gebruikt als covariaat. Voor groep B is echter de vraag welke meting als voormeting beschouwd moet worden (zie Figuur 2): die van meetmoment 2 of die van meetmoment 1? Omdat de leerlingen uit groep B geselecteerd zijn op hun scores op meetmoment 1, lijken die scores de *zuiverste* voormeting⁵. De WATscores op meetmoment 1 zullen voor groep B als voormeting gelden in de analyse van de diepe woordkennis.

Voor leesvaardigheid geldt dat de cloze-toetsen op meetmoment 1 en 2 verschillende toetsen met verschillende schalen zijn (Cito- en BOScloze), zodat voor de analyse van de leesprestaties ook voor groep B de scores op meetmoment 2 als voormeting gelden (en niet die op meetmoment 1, zoals wenselijker zou zijn).

Effecten op de geoefende domeinen (C-WAT)

Tabel 5 laat zien dat de gemiddelde scores op de covariaat voor de zwakke leerlingen uit de beide groepen nagenoeg gelijk zijn. Bij de nameting scoren de getrainde kinderen duidelijk hoger dan de niet-getrainde kinderen uit groep C. Het verschil is statistisch significant en substantieel. De scores van de getrainde, zwakke leerlingen zijn weliswaar nog niet zo hoog als die van hun klasgenoten, maar ze hebben een flinke inhaalslag gemaakt. Ook bij de klasgenoten is er een statistisch significant verschil opgetreden in het voordeel van de groep B-leerlingen. Dit verschil is echter aanzienlijk kleiner dan het verschil bij de zwakke leerlingen, zodat een algemeen verschil tussen groep B en groep C niet de (volledige) verklaring kan zijn voor het geconstateerde verschil bij de zwakke leerlingen.

Evenals bij Ronde 1 is nagegaan of er verschillen zijn tussen de drie thema's. De C-WAT is weer opgesplitst in drie deelscores, één voor elk thema. Het blijkt ook nu dat het hierboven gerapporteerde effect op de C-WAT bij de zwakke leerlingen niet toe te

Tabel 4

Gemiddelde scores op de covariaat van meetmoment 1 (Citocloze1) en de met covariaat gecorrigeerde afhankelijke variabele op meetmoment 2 (BOScloze2), met de toetsing van de significantie ($F(df_1, df_2)$ en p) en proportie verklaarde variantie door conditie (η^2)

	Citocloze mm1 (covariaat, max. 41)	BOScloze mm2 (gecorrigeerd, max. 28)	Toetsing
<i>Zwakke leerlingen</i>			
Groep A ($n=92$)	10.72	6.54	$F(1,162)=1.53$
Groep B ($n=73$)	10.71	5.99	$p=.219, \eta^2=.009$
<i>Klasgenoten</i>			
Groep A ($n=273$)	14.71	9.95	$F(1,606)=5.65$
Groep B ($n=336$)	15.42	9.27	$p=.018, \eta^2=.009$

Tabel 5

Gemiddelde scores op de covariaat (Groep B: WATmm1; Groep C: WATmm2) en met covariaat gecorrigeerde afhankelijke variabele op meetmoment 3 (C-WATmm3), met de toetsing van de significantie ($F(df_1, df_2)$ en p) en proportie verklaarde variantie door conditie (η^2)

	WATmm1/2 (covariaat, max. 90)	C-WATmm3 (gecorrigeerd, max. 108)	Toetsing
<i>Zwakke leerlingen</i>			
Groep B ($n=52$)	71.71	95.59	$F(1,124)=16.60$
Groep C ($n=75$)	71.63	90.46	$p=.000, \eta^2=.118$
<i>Klasgenoten</i>			
Groep B ($n=216$)	79.54	99.19	$F(1,563)=15.43$
Groep C ($n=350$)	80.97	97.13	$p=.000, \eta^2=.027$

schrijven is aan één thema. Op alle drie de thema's kunnen significante effecten gerapporteerd worden ($F=7.56, F=8.12$ resp. $F=20.11$, alle $p \leq .007$). En ook hier lijkt het erop dat er een toenemend effect is van het programma, $\eta^2=.057, .061$ respectievelijk .140.

Effecten op de niet-geoefende domeinen (WAT)

Op de niet-geoefende domeinen van diepe woordkennis zijn er geen verschillen tussen de getrainde en de niet-getrainde zwakke leerlingen op de nameting. Dit betekent dat de resultaten van de eerste ronde gerepliceerd worden: er treedt wel een trainingseffect op binnen de geoefende domeinen, maar er is geen sprake van transfer naar andere domeinen of diepe woordkennis in het algemeen, zoals gemeten met de WAT.

Ook de klasgenoten van de zwakke leerlingen uit groep B en C verschillen niet of nauwelijks van elkaar.

Tabel 6

Gemiddelde scores op de covariaat (Groep B: WATmm1; Groep C: WATmm2) en de met covariaat gecorrigeerde afhankelijke variabele op meetmoment 3 (WATmm3), met de toetsing van de significantie ($F(df1, df2)$ en p) en proportie verklaarde variantie door conditie (η^2)

	WAT1/2 (covariaat, max. 90)	WAT3 (gecorrigeerd, max. 90)	Toetsing
<i>Zwakke leerlingen</i>			
Groep B (n=61)	71.39	75.90	$F(1,140)=0.05$
Groep C (n=82)	71.02	76.10	$p=.823, \eta^2=.000$
<i>Klasgenoten</i>			
Groep B (n=253)	79.51	82.01	$F(1,612)=0.004$
Groep C (n=362)	80.87	81.98	$p=.947, \eta^2=.000$

Tabel 7

Gemiddelde scores op de covariaat (BOScloze-mm2) en de met covariaat gecorrigeerde afhankelijke variabele op meetmoment 3 (BOScloze-mm3), met de toetsing van de significantie ($F(df1, df2)$ en p) en proportie verklaarde variantie door conditie (η^2)

	BOScloze mm2 (covariaat, max. 28)	BOScloze mm3 (gecorrigeerd, max. 28)	Toetsing
<i>Zwakke leerlingen</i>			
Groep B (n=63)	6.05	8.74	$F(1,142)=0.273$
Groep C (n=82)	8.12	8.50	$p=.602, \eta^2=.002$
<i>Klasgenoten</i>			
Groep B (n=278)	9.45	12.37	$F(1,639)=6.087$
Groep C (n=364)	12.50	11.83	$p=.014, \eta^2=.009$

Effecten op de leesvaardigheid (BOScloze)

In deze analyse gelden de prestaties op de BOSclozetoets op meetmoment 2 voor alle leerlingen als voormeting. In Tabel 7 blijkt dat de zwakke leerlingen bij de nameting nagenoeg gelijk presteren als men rekening houdt met de beginverschillen. Opvallend is wel dat de zwakke leerlingen uit de controle-groep (groep C), evenals hun klasgenoten bij de voormeting relatief hoog scores op de clozetoets: 8.12 versus 6.05, respectievelijk 12.50 versus 9.45. Hiervoor hebben we nog geen verklaring; de toets kan niet bekend ge-

weest zijn bij leerlingen uit groep C, omdat hij voor dit onderzoek ontwikkeld is. Hoewel de klassen niet aselekt aan de condities zijn toegewezen, lijken ze toch niet erg verschillend als we afgaan op de gemiddelde WAT-scores (zie Tabel 5 en 6). Bovendien is het goed om te bedenken dat zowel groep B als C uit leerlingen van zo'n 18 klassen bestaat, zodat het ook onwaarschijnlijk is dat karakteristieken van individuele leerkrachten en hun lessen de verschillen kunnen verklaren.

Bij de klasgenoten treedt een (klein) significant verschil op tussen de twee groepen, omdat de leerlingen uit groep B ten opzichte van de leerlingen uit groep C iets meer vooruit zijn gegaan. Overigens zijn de leerlingen uit groep C niet achteruitgegaan, zoals Tabel 7 suggereert. De gemiddelden op de nameting zijn de gecorrigeerde uit de covariantie-analyse, d.w.z. dat de gemiddelden zijn bepaald als hadden de beide groepen een gelijk beginniveau, namelijk dat tussen de huidige beginniveaus (9.45 en 12.50: 11.18). Groep B is ten opzichte van dit referentiepunt dus meer vooruitgegaan dan groep C.

4 Conclusie en discussie

Het lesprogramma lijkt in beide onderzoeksrondes effectief geweest te zijn: de getrainde, zwakke kinderen vertonen vooruitgang ten opzichte van niet-getrainde, zwakke kinderen, terwijl hun klasgenoten gelijk blijven ten opzichte van de controleklasgenoten (onderzoeksvraag 1). Met deze laatste constatering wordt min of meer uitgesloten dat de leerkrachten die met het programma werkten, zich in het algemeen anders zijn gaan gedragen. Het geconstateerde effect kan evenmin het gevolg zijn van *training-to-the-test*, omdat het format en de inhoud van de curriculumgerelateerde test (C-WAT) niet als zodanig in het lesprogramma aan de orde kwam. Niettemin kan er ook een enkele kanttekening geplaatst worden bij de resultaten. Zo hebben de kinderen in de controleconditie alleen regulier onderwijs gekregen. Dit betekent dat in de drie geoefende domeinen het lesprogramma effectiever is dan *gewoon* onderwijs. In hoeverre de *specifieke aanpak* van het lesprogramma hiervan de oorzaak is, zou

men verder moeten onderzoeken door het programma af te zetten tegen een *traditioneel* woordenschatprogramma, waarin dezelfde drie domeinen behandeld worden, maar dan zonder specifieke aandacht voor de paradigmatische relaties tussen woorden in een domein en zonder geleidelijke decontextualisatie van woordkennis.

In ons onderzoek hebben we geen retentiemetingen uitgevoerd, zodat we niet weten wat de effecten van het lesprogramma op de langere termijn zijn. Het is uiteraard niet uitgesloten dat dergelijke positieve effecten langzaam wegebben, zeker omdat het hier om een betrekkelijk korte interventie ging, maar het is ook voorstelbaar dat de trainingseffecten meer tijd nodig hebben om ook effect te sorteren in andere domeinen van het mentale lexicon en om de leesvaardigheid van de kinderen te beïnvloeden (zie ook hierna).

Het positieve onderzoeksresultaat wordt verder ondersteund door leerkracherveraringen zoals beschreven in de logboeken. Leerkrachten rapporteren dat hun inzicht in de lexicale ontwikkeling van de kinderen toegenomen is. De gestructureerde opbouw van het programma laat precies zien hoe de verdieping van woordkennis zich stap voor stap bij de kinderen voltrekt. In elke lesfase is reflectie op de activiteiten een belangrijk onderdeel. Het praten met de leerlingen over achterliggende betekenistoekenning en woordrelaties wordt als zeer verhelderend ervaren. Een aantal uitspraken van de leerkrachten zelf:

- “Het levert zoveel op, zowel voor de leerlingen als voor mezelf”;
- “De bespreking is erg zinvol; je ziet precies waar de leerlingen zitten, wat ze al weten en wat ze nog moeten leren”;
- “Steeds weer was de tijd eigenlijk te kort: ik wou dat er gelegenheid was om langer door te praten, dat laat zoveel zien”.

Wat betreft de generalisatie of transfer van het effect hebben we niet kunnen aantonen dat het generaliseert naar andere domeinen van het mentale lexicon (WAT) (onderzoeksvraag 2); de leesvaardigheid van de kinderen is evenmin aantoonbaar verbeterd (onderzoeksvraag 3). In dat opzicht is er geen sprake van transfer. Er zijn (minstens) drie plausible verklaringen waartussen we op basis

van ons onderzoek niet kunnen beslissen. Deze verklaringen sluiten elkaar overigens niet uit en zijn mogelijk ook in combinatie relevant.

De eerste verklaring betreft de duur en de omvang van het programma. Deze zijn mogelijk te kort en te beperkt geweest om het mentale lexicon als geheel te beïnvloeden. Dit is zeer goed denkbaar omdat er slechts drie domeinen aan de orde zijn geweest in relatief korte tijd: (maximaal) 4 uur per week per domein, dus voor het totale lesprogramma ca. 12 uur. Het mentale lexicon en de structuur ervan is in de loop der jaren opgebouwd en het is misschien te optimistisch te verwachten dat in een korte periode te kunnen veranderen. Onderzoek met een uitgebreider en/of intensiever programma zou uitsluitel kunnen geven over deze mogelijke verklaring. In een uitgebreider programma is het wenselijk dat hetgeen de kleine groepjes zwakke leerlingen in de remediërende lessen geleerd hebben, terugkeert in de reguliere, klassikale lessen. Dit is ook een nadrukkelijke wens van de leerkrachten die met het materiaal gewerkt hebben. Van der Nulft en Verhallen (2001) ontwikkelen momenteel een aanpak voor leerkrachten om diepe woordkennis consequent te integreren in hun onderwijs, zowel het taalonderwijs als ook het zaakvakonderwijs.

Een tweede mogelijke verklaring voor het uitblijven van transfer is dat diepe woordkennis niet *spontaan* transfereert naar andere domeinen. Dit zou impliceren dat voor elk domein diepe woordkennis aangeleerd moet worden. Deze wat sombere conclusie laat overigens onverlet dat diepe woordkennis belangrijk is voor een succesvol onderwijsleerproces. Relevant in dit verband is dat leerkrachten rapporteerden dat uitleg voor volgende domeinen steeds makkelijker werd en door kinderen sneller opgepikt werd. Uitspraken als “kinderen zijn enthousiast omdat ze de oefeningen bij het tweede en derde thema herkennen” en “ze gaan er makkelijker doorheen” wijzen op facilitering van het leerproces en daarmee op een vorm van transfer van het eerder geleerde. In het verlengde hiervan tekenen we ook aan dat er voor de opeenvolgende domeinen een steeds groter effect

van het programma geconstateerd werd - in beide onderzoeksrondes - wat eveneens op een vorm van transfer kan wijzen.

De derde verklaring is dat het programma niet of te weinig gericht is geweest op transfer. Kinderen leren weliswaar de woordkennis in bepaalde domeinen verdiepen, maar er werden hun niet systematisch bepaalde strategieën of procedures geleerd. Het aanleren van adequate strategieën en ook metacognitieve kennis van het gebruik en het nut van die strategieën zijn twee condities (naast andere) die transfer kunnen bevorderen (Simons, 1990). In het experimentele programma is bijvoorbeeld geen bewuste aandacht voor de toepassing van strategieën in domein B, die al eerder gebruikt zijn in domein A. Mogelijk wordt er in het programma te veel uitgegaan van de verwachting dat kinderen min of meer automatisch of via analogievorming ook in een nieuw domein systematische en hiërarchische relaties tussen woorden gaan leggen, waar ze dat eerder ook in een geoefend domein hebben gedaan. Bovendien moet hierbij aangetekend worden dat het om zwak presterende kinderen gaat bij wie geen groot transfervermogen verwacht mag worden (vgl. Simons, 1990; p.24).

Verder is de implementatie van het programma in het onderzoek niet altijd volledig geweest, met name in de tweede onderzoeksronde kwam het derde thema bij enkele docenten in het gedrang vanwege de naderende zomervakantie. Door een andere planning van (de implementatie van) het programma zou de uitvoering verder geoptimaliseerd kunnen worden. Overigens is de implementatie van het programma tijdens het onderzoek niet verder begeleid; ze bestond uit de eerder genoemde instructiedag met uitleg en demonstratievideo. Coaching van docenten, zeker als de aandacht voor diepe woordkennis ook in de reguliere lessen moet terugkeren, zal ook verder kunnen bijdragen aan een grotere effectiviteit van de aanpak.

Vanuit een theoretisch perspectief is het interessant te constateren dat diepe woordkennis *trainbaar* is. Zoals al opgemerkt, zijn er voor de *generalisatie* van het effect alleen kleine, indirecte aanwijzingen. Deze vraag verdient verdere aandacht, omdat zowel theoretisch als praktisch belangrijk is te weten of

de *herstructurering* van het mentaal lexicon gezien moet worden als het aanleren van nieuwe, gedecontextualiseerde woordrelaties die steeds apart aangeleerd moeten worden, of dat het gezien moet worden als metacognitieve of strategische kennis die ingezet kan worden bij alle domeinen van het lexicon. Als in het laatste geval deze kennis, dit inzicht, in voldoende mate verworven is, zou het mentale lexicon als geheel *geherstructureerd* worden. De verdieping van het mentale lexicon blijft dan niet beperkt tot de geoefende domeinen. Dit betekent waarschijnlijk wel dat in het programma bepaalde strategieën ook bewust geleerd moeten worden, met andere woorden dat de beoogde transfer ook daadwerkelijk in het programma wordt bewerkstelligd.

In de inleiding werd gewezen op de relatie tussen diepe woordkennis en leesvaardigheid (Bekius & Simis, 1997; Qian, 1999; Schoonen & Verhallen, 1998b). In hoeverre die relatie causaal is, is onzeker. In deze experimentele onderzoeksopzet was ook leesvaardigheid een afhankelijke variabele. Echter, omdat de uitbreiding van diepe woordkennis niet of onvoldoende generaliseerde naar diepe woordkennis in het algemeen, konden we geen effect meer verwachten op de algemene leesvaardigheid zoals gemeten met de clozetoets. Om vast te kunnen stellen of diepe woordkennis een (causaal) effect heeft op de leesvaardigheid zal er eerst een breder effect van de diepe woordkennistraining op moeten treden, of de operationalisatie van leesvaardigheid moet beperkt worden tot de domeinen die in het programma aan de orde komen: dus leesteksten waarin dieren, beroepen of voertuigen een belangrijke rol spelen.

De vraag dringt zich op in hoeverre het hier bestudeerde lesprogramma zich nog richt op de woordenschat van de kinderen en niet op hun conceptuele ontwikkeling. De lexicale en de conceptuele kennis ontwikkelen zich in een voortdurende wisselwerking (Waxman, 1991). De rol van de taal wordt daarin vaak als faciliterend gezien voor de conceptuele ontwikkeling (Keil, 1991; Waxman, 1994). Uit onderzoek bij jonge kinderen (voorschools) blijkt dat het woordgebruik soms vooruitloopt op een precieze conceptu-

ele invulling van de woordbetekenis, zoals blijkt uit overextensies; de conceptuele kennis van een kind kan daarentegen ook mede bepalen welke betekenis een kind aan nieuwe woorden toekent (b.v. *oom* als verwijzend naar bepaalde gedragskenmerken of als uitdrukking van bepaalde biologische familiere-laties; vgl. Keil 1991). Vooral als men zich richt op de zogenaamde diepe woordkennis, dan zijn conceptuele en lexicale ontwikkeling lastig te onderscheiden. In de inleiding werd al gewezen op een verschuiving in de lexicale ontwikkeling van syntagmatische naar meer paradigmatische betekenistoekenning. Ook als het gaat om de conceptuele ontwikkeling van kinderen is er een dergelijke verschuiving te constateren, van *association* naar meer *theory-based* kennis (Keil, 1991). Voor het lesprogramma, en voor het inlopen van de achterstanden van zwakke leerlingen, maakt het weinig uit of we spreken over lexicale of over conceptuele ontwikkeling. Concepten worden in woorden uitgedrukt en relaties tussen concepten komen tot uiting in relaties tussen woorden. Zoals hiervoor al opgemerkt werd, kan de taal faciliterend werken bij het construeren van conceptuele hiërarchieën door kinderen. In deze zin is er dus sprake van overlap tussen de conceptuele ontwikkeling en de ontwikkeling van diepe woordkennis. Van belang is dat kinderen over voldoende en adequate woordkennis beschikken, zodat hun kans op succesvol onderwijs vergroot wordt.

Zwakke leerlingen beginnen aan de les-reeks met sterk subjectieve betekenistoekenningen. Ze hanteren aanvankelijk ordeningen als *dieren uit warme landen* → *dieren uit koude landen* → *dieren uit Nederland*, komen gaandeweg tot oplossingen als: “we hebben de uil tussen roofdieren en vogels ingelegd, want ze eten altijd muizen” en komen ten slotte tot definities als “een mier is een soort spin, heeft 6 poten, is een insect en heeft grote ogen”. Dat leerlingen aan het einde van het programma in alle getrainde domeinen over redelijk wat diepe woordkennis beschikken, wordt niet alleen ondersteund door de rapportage van de leerkrachten in hun logboeken, maar ook door de definities die leerlingen geven van woorden in de betreffende thema's (zie Voorbeeld 2).

Voorbeeld 2

Definities van (taalzwakke/allochtone) kinderen in groep 5 na het volgen van de lessen

Na het thema *dieren*:

- Een libelle is een insect die alle kleuren kan hebben. Z'n vleugels zijn doorzichtig. Hij vliegt in het rond.
- Een tijger is een roofdier dat vlees eet en een bedreigde diersoort is.
- Een eend is een vogel. Hij leeft in sloten en meertjes, niet in zee. Die is te zout. Hij legt eieren, het is een watervogel met drie tenen en zwemvliezen.

Na het thema *voertuigen*:

- Een ambulance is een ziekenwagen die zieke mensen naar het ziekenhuis brengt. Het is een landvoertuig, groot, met een bestuurder. Hij heeft zwaailichten en er kunnen 5 mensen in.
- Een roeiboot is een voertuig zonder motor, je moet zelf vooruit zien te komen.
- Een auto heeft een stuur. Landvoertuig, er kunnen mensen in. Het heeft een merk, lichten, een toeter en er moet benzine in.

Na het thema *beroepen*:

- Een beeldhouwer is een kunstenaar en maakt beelden van hout of van steen.
- Een dierenarts: als een dier ziek is gaat ze het opereren. Het is een dokter of een zuster.
- Een piloot is een vliegtuigbestuurder. Er kunnen maar 2 mensen voor, want als er één gaat eten bijvoorbeeld heb je de andere nodig.

Noten

- 1 Het programma is ontwikkeld door het Expertisecentrum NT2 van de PABO Haarlem (Verhallen & Walst, 1997). Het programma zal in herziene versie in 2002 verkrijgbaar zijn, gekoppeld aan een scholingscursus.
- 2 Aannames voor een covariantieanalyse zijn de aannames voor een variantieanalyse: onafhankelijke waarnemingen, normaliteit in elke groep, gelijke populatievarianties en daarbij de extra aannames van een lineaire relatie tussen covariaat en afhankelijke variabele, gelijke hellingshoek voor de regressies in elke groep en betrouwbare covariaat (Stevens, 1992). Voor de eerste ronde zijn er alleen enige schendingen van de normaliteit van de scoreverdeling. Hoewel

het geen mooie normaalverdelingen zijn, is er geen sprake van bimodale verdelingen o.i.d., dus we mogen aannemen dat de schending van normaliteit weinig invloed heeft op onze statistische toetsingen (zie Stevens, 1992). Voor de tweede ronde zijn er ook kleine schendingen van de homogeniteit van varianties en de paralleliteit van de regressies, maar deze schendingen zijn ook betrekkelijk klein.

- 3 Bij de clozetoetsen zijn overgeslagen items fout gerekend. Een analyse van de betrouwbaarheid van de clozetoetsen met alleen de leerlingen die geen items hebben overgeslagen leidt tot vergelijkbare betrouwbaarheidsschattingen voor de achtereenvolgende meetmomenten: .86, .82 en .83 ($n=389$, 728 en 565 respectievelijk).
- 4 Cohen (1988) gebruikt f als maat voor effectgrootte en spreekt vanaf $f=.40$ van grote effecten; een f van .40 komt overeen met een $\eta^2=.1379$. Een middelmatig effect correspondeert omgerekend met $\eta^2=.0588$ en een klein effect met $\eta^2=.0099$.
- 5 Aanvankelijk waren we uitgegaan van een voormeting op hetzelfde tijdstip, dus voor zowel groep B als groep C de scores op meetmoment 2 als voormeting (zie Figuur 2). Echter, bij nadere beschouwing bleek het beginniveau op de WAT van zwakke leerlingen uit groep B duidelijk hoger dan van de zwakke leerlingen uit groep C, terwijl de klasgenoten uit beide groepen een vergelijkbaar gemiddelde behaalden. Bovendien maakten de zwakke leerlingen uit groep C een enorme progressie op de WAT, in tegenstelling tot hun klasgenoten. Dit laatste maakt een "gewoon" her-testeffect een onwaarschijnlijke verklaring. Immers, dan zouden zowel de zwakke leerlingen als de klasgenoten uit groep C vooruit zijn gegaan. De meest plausibele verklaring is dat we ook hier te maken hebben met regressie naar het gemiddelde. De zwakke leerlingen zijn geselecteerd op hun zwakke prestatie op de WAT. Deze selecte groep zal als gevolg van het genoemde verschijnsel van regressie naar het gemiddelde, een volgende meting altijd hoger scoren. De zwakke leerlingen van groep B zijn geselecteerd op hun WATscore op meetmoment 1, de zwakke leerlingen van groep C op hun WATscore op meetmoment 2. Om een zuivere vergelijking te krijgen voor deze selecte groepen dienen de scores waarop ze geselecteerd zijn als voormeting, dus WATmm1 voor groep B en WATmm2 voor groep C.

Literatuur

- Appel, R. (1999). De woordenschat van allochtone tweede-taalverwervers; werk in uitvoering. *Toegepaste Taalwetenschap in Artikelen* 61, nr. 1, 9-19.
- Appel, R. & Verhallen, M. (1998). Woordenschat en taalonderwijs op de basisschool. *Toegepaste Taalwetenschap in Artikelen* 58, nr. 1, 31-39.
- Bekius, A. & Simis, A. (1997). *Diepe woordkennis en tekstbegrip van Nederlandstalige en anderstalige kinderen in het basisonderwijs*. Amsterdam: Universiteit van Amsterdam, Algemene Taalwetenschap (Doctoraalscriptie).
- Cito (1990). *Clozetoetsen leesvaardigheid*. Arnhem: Cito.
- Coady, J. (1997) L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.) *Second language vocabulary acquisition* (pp. 225-237). Cambridge: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Daalen-Kapteijns, M. van, Elshout-Mohr, M. & Gloppe, K. de (2001). Deriving the meaning of unknown words from multiple contexts. *Language Learning*, 51, 1, 145-181.
- Keil, F. C. (1991). Theories, concepts, and the acquisition of word meaning. In A. S. Gelman & P. J. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 197-221). Cambridge: Cambridge University Press.
- Kerlinger, F.N. (1973). *Foundations of behavioral research* (2nd ed.). New York etc.: Holt, Rinehart and Winston, Inc.
- Nulft, D. van der & Verhallen, M. (2001). *Met woorden in de weer*. Bussum: Coutinho (te verschijnen september 2001).
- Qian, D.D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in ESL reading comprehension. *The Canadian Modern Language Review*, 56, 2, 282-308.
- Schoonen, R. & Verhallen, M. (1998a). Kennis van woorden: de toetsing van diepe woordkennis. *Pedagogische Studiën*, 75, 3, 153-168.
- Schoonen, R. & Verhallen, M. (1998b). *Aspects of vocabulary knowledge and reading performance*. Paper presented at the Annual Meeting of the American Educational Research Association. San Diego, April 13-17th, 1998.
- Simons, P. R. J. (1990). *Transfervermogen. Rede uit-*

gesproken bij de aanvaarding van het ambt van gewoon hoogleraar in de interdisciplinaire onderwijskunde, in het bijzonder de onderwijsverbetering, aan de Katholieke Universiteit Nijmegen op woensdag 12 december 1990. Nijmegen: Katholieke Universiteit Nijmegen.

- Snow, C.E. (1990). The development of definitional skills. *Journal of Child Language*, 17, 697-710.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 4, 360-407.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd edition). Hillsdale, NJ/Hove/London: Lawrence Erlbaum Associates.
- Verhoeven, L. & Vermeer, A. (1989). *Diagnose van kindertaal*. Tilburg: Zwijssen.
- Verhallen, M. (1994). *Lexicale vaardigheid van Turkse en Nederlandse kinderen. Een vergelijkend onderzoek naar betekenis-toekenning* (Academisch proefschrift). Amsterdam: Universiteit van Amsterdam/IFOTT.
- Verhallen, M. & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics*, 19, 4, 452-470.
- Verhallen, M. & Verhallen, S. (1994). *Woorden leren, woorden onderwijzen*. Hoevelaken: CPS.
- Verhallen, M. & Walst, R. (1997). "Van Woorden Weten"; *experimenteel lesprogramma voor verdieping van woordkennis*. Haarlem: Expertisecentrum NT2, PABO-Haarlem.
- Verhallen, M. & Walst, R. (1999). Een libelle kan alle kleuren hebben. Het lesprogramma "Van Woorden Weten". *TooN*, 2, 1 (januari).
- Vermeer, A. (1997). Breedte en diepte van woordenschat in relatie tot toenemende taalverwerving en frequentie van taalaanbod. *Gramma/TTT*, 6, 169-187.
- Waxman, S. R. (1991). Convergences between semantic and conceptual organization in the preschool years. In A.S. Gelman & P.J. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 107-145). Cambridge: Cambridge University Press.
- Waxman, S. R. (1994). The development of an appreciation of specific linkages between linguistic and conceptual organization. *Lingua*, 92 (1-4), 229-257.

Auteurs

Dr. Marianne Verhallen is werkzaam als coördinator van het Expertisecentrum NT2 aan de Pabo Hogeschool Haarlem.

Dr. Rob Schoonen is werkzaam als universitair docent bij de Leerstoelgroep Tweede-taalverwerving van de Faculteit der Geesteswetenschappen aan de Universiteit van Amsterdam.

Prof. Dr. René Appel is werkzaam als Bijzonder Hoogleraar Verwerving en Didactiek van het Nederlands als tweede taal bij de Leerstoelgroep Tweede-taalverwerving van de Faculteit der Geesteswetenschappen aan de Universiteit van Amsterdam.

Correspondentieadres: mw. dr. M. Verhallen, Expertisecentrum NT2, PABO, Hogeschool Haarlem, Postbus 614, 2003 RP Haarlem.
E-mail: m.verhallen@hshaarlem.nl

Abstract

Deep word knowledge: an empirical study on the effects of an experimental programme

The present article reports on an empirical study on the effects of an experimental programme for specific vocabulary skills of children in primary education, which aimed at improving their so-called deep word knowledge: the ability to understand words and their meanings in a logical, hierarchical structure. The programme was remedial in nature, i.e. it was meant for children with limited deep word knowledge. An experimental group received training in three domains: animals, vehicles and occupations. Evaluation of the effects of the programme pointed out that the deep word knowledge of the experimental group improved significantly (when compared with a control group that did not follow a training). This result only applied to the domains dealt with in the programme. There was no transfer to other domains and neither to reading skills.