

Voorlezen op school aan jonge kinderen: een review*

H. Blok

Samenvatting

In kindercentra en in kleutergroepen wordt praktisch dagelijks door leidsters en leerkrachten voorgelezen. Eén van de oogmerken is de bevordering van de taal- en leesontwikkeling van de kinderen. Dit artikel beschrijft een review naar de effecten die het voorlezen op school heeft op de taal- en leesontwikkeling van jonge kinderen (tot acht jaar). In totaal werden elf studies (met twaalf steekproeven) gevonden die het voorlezen op een kindercentrum of op school tot onderwerp hebben. De gecombineerde effectgrootte voor 'mondelinge taal' bedraagt .64, voor 'lezen' .22. Hoewel met name het eerste effect niet te verwaarlozen is, wordt gewaarschuwd voor een al te optimistische interpretatie. De empirische basis is erg smal. Niet alleen is het aantal studies beperkt, maar ook laat een kritische analyse van de studies ernstige tekorten zien, zowel in de opzet als in de rapportage.

1 Inleiding

Op basisscholen wordt veelvuldig aan jonge kinderen voorgelezen. Exacte cijfers zijn niet gemakkelijk te geven. Maar de indruk bestaat dat in de meeste kleuterklassen (groepen 1 en 2) tenminste één keer per dag door de leerkracht voorgelezen wordt (Veen, Overmars & De Glopper, 1995; p. 181). Waarschijnlijk neemt de frequentie van voorlezen af naarmate de leerlingen ouder worden. Hierbij speelt een rol dat kinderen in de middenbouw (groepen 3, 4 en 5) steeds beter zelf gaan lezen. In de bovenbouw is de leesvaardigheid van de mees-

te leerlingen voldoende hoog om zelfstandig een breed scala van teksten te kunnen lezen. Toch wordt ook in de groepen 7 en 8 nog veelvuldig aan kinderen voorgelezen. In een steekproef van 28 leerkrachten (deels groep 7, deels groep 8) uit Amsterdam en omgeving bleek dat het merendeel van hen meermalen per week naar een voorleesboek greep (Blok, 1992).

In dit artikel beperk ik me tot het voorlezen – in een geïnstitutionaliseerde context – aan jonge kinderen, geoperationaliseerd als kinderen tot ongeveer 8 jaar. Een strikte ondergrens is er niet, omdat deze afhangt van de leeftijd waarop kinderen naar een kindercentrum of naar school gaan. In ons land is dat als regel 4 jaar, maar soms ook eerder. In de Verenigde Staten, waar het merendeel van het wetenschappelijke onderzoek verricht wordt, gaan kinderen als regel vanaf hun vijfde naar school, maar soms eerder.

Werkers op kindercentra en leerkrachten – gemakshalve vatten we vanaf hier beide groepen samen met de term leerkrachten – hebben met het voorlezen diverse doelen voor ogen. De belangrijkste doelen zijn onder drie noemers samen te vatten: bevordering van de taal- en leesontwikkeling, cultuuroverdracht en persoonlijkheidsgroei (Baudoin et al., 1990; Nijmeegse Werkgroep Taaldidactiek, 1992). Voorbeelden van doelen op het terrein van de taal- en leesontwikkeling zijn: vergroting van de taalbeheersing, kennismaking met de functie van geschreven taal en met de opbouw van boeken en verhalen. Voorbeelden van doelen op het terrein van de cultuuroverdracht zijn: wereldoriëntatie, overdracht van waarden en normen (incl. omgangsvormen) en inwijding in de wereld van de literatuur. Een voorbeeld op het terrein van persoonlijkheidsgroei: kinderen inzicht geven in de eigen gevoels- en ervaringswereld.

De vraag kan worden gesteld of voorlezen een geschikt middel is voor het eerste doel:

* Deze literatuurstudie is uitgevoerd met financiële steun van het Instituut voor Onderzoek van Onderwijs (SVO), project 95729.

bevordering van de taal- en leesontwikkeling. Er is uitgebreid empirisch onderzoek gedaan naar deze vraag. Het grootste deel van de studies heeft betrekking op het voorlezen in gezinsverband. Recent is een tweetal meta-analyses verschenen, namelijk Bus, IJzendoorn en Pellegrini (1995) en Scarborough en Dobrich (1994). Deze overzichtstudies komen in belangrijke opzichten overeen. De bestudeerde onafhankelijke variabele is het voorlezen door ouders aan hun kinderen in de voorschoolse leeftijd. De gemeenschappelijke vraag is welke effecten deze variabele heeft op de taal- en leesontwikkeling van kinderen. Gemeenschappelijk is ook het grootste deel van het bronnenmateriaal: ruim 30 studies die in de afgelopen dertig jaar verschenen zijn.

Een verschil is er in de manier waarop de resultaten van de studies gecombineerd zijn. Bus et al., (o.c.) voerden een voornamelijk kwantitatieve meta-analyse uit waarbij studies met een verschillende opzet (correlationeel, longitudinaal, experimenteel) bij elkaar gevoegd zijn. Scarborough en Dobrich (o.c.) hanteerden naast kwantitatieve maten ook kwalitatieve overwegingen en analyseerden de studies per afzonderlijk onderzoekstype. Ondanks het methodische verschil stemmen de conclusies sterk met elkaar overeen. Bus et al. rapporteren een algemene effectgrootte (Cohens d) van .59. Scarborough en Dobrich rapporteren een mediane effectgrootte (pmc) van .26. Beide conclusies samengenomen duiden erop dat 7 tot 10 procent van de variantie in taal- en leesvaardigheden samenhangt met het voorlezen door ouders in de voorschoolse periode.

In beide meta-analyses wordt tevens de conclusie getrokken dat de onafhankelijke variabele, het voorlezen door de ouders, in de meeste studies onvoldoende precies beschreven is. Er bestaat daardoor onzekerheid over de modale 'treatment' die nodig is om, bijvoorbeeld, een effectgrootte van .6 te realiseren. Die onzekerheid bestaat er in drie belangrijke opzichten: de frequentie waarmee voorgelezen wordt, de leeftijd waarop met voorgelezen wordt aangevangen (en bijgevolg de duur van de 'treatment') en de stijl van voorlezen. Wat de voorleesstijl betreft hebben alleen Scarborough en Dobrich een vijftal relevante studies apart geanalyseerd. Zij vonden hierin geen aanwijzingen dat de voorleesstijl sterker met de taal-

en leesontwikkeling samenhang dan de frequentie van voorlezen.

Hoewel beide besproken overzichtstudies het inzicht in de mogelijke effecten van voorlezen verdiepen, kan men de besproken conclusies niet zonder meer generaliseren naar het voorlezen op school. Tussen voorlezen op school en voorlezen in gezinsverband bestaan niet onaanzienlijke verschillen. Het meest kenmerkende verschil is dat ouders als regel aan maar één kind, of hoogstens enkele kinderen tegelijk voorlezen, terwijl op school meestal aan een hele klas tegelijk wordt voorgelezen. Dit verschil heeft consequenties voor het aantal interacties tussen voorlezer en luisteraar en de aard van deze interacties. Leerkrachten hebben tijdens het voorlezen veel minder mogelijkheden hun gedrag af te stemmen op de individuele leerling dan ouders dat hebben ten opzichte van hun eigen kind(eren). Op zijn best bevatten de meta-analyses van Bus et al. en Scarborough en Dobrich alleen indirecte evidentie aangaande het vraagstuk van de effectiviteit van voorlezen op school: de 7 tot 10 procent gemeenschappelijke variantie is mogelijksterwijs een bovengrensschatting.

Een meer direct antwoord kan men proberen te krijgen door een analyse van de literatuur over voorlezen op school. Maar hierbij stuit men op twee moeilijkheden. De eerste is dat er nog maar weinig studies gedaan zijn naar de effecten van voorlezen op school. De tweede moeilijkheid is dat over deze studies geen review beschikbaar is. Het onderhavige artikel is bedoeld om in deze leemte te voorzien. Het doet verslag van een review van elf studies die het resultaat zijn van een uitgebreide speurtocht.

2 Probleemstelling

2.1 Kernbegrippen

Idealiter dicteert een theorie de hypothesen die in een review worden geëvalueerd. In het geval er geen sprake is van een coherente theorie, moet men volstaan met een minder samenhangend stelsel van uitspraken over de factoren die een beschrijving of verklaring geven van de bestudeerde verschijnselen. Hier wordt een poging gedaan zo'n theorie, in een noodzakelijk nog rudimentaire vorm, te specificeren.

Taalontwikkeling is het object van de te ontwikkelen theorie. De snelheid waarmee taalontwikkeling plaatsvindt, en met name de individuele variatie daarin, is de te verklaren grootheid. Taalontwikkeling wordt hier in een ruime betekenis gehanteerd, betrekking hebbend op zowel de gesproken als geschreven taal. Onderscheiden componenten zijn: fonologische vaardigheden, woordkennis, kennis van woord- en zinsbouw, lees- en schrijfvaardigheid. De volgorde van deze opsomming representeert grofweg een ontwikkelingsvolgorde, althans wat betreft het beginpunt van elk der componenten. De ontwikkeling van vaardigheden in het herkennen en produceren van spraakklanken kent bijvoorbeeld een vroegere start dan de ontwikkeling van een woordenschat. Evenzo kent de ontwikkeling van een woordenschat een eerdere start dan de ontwikkeling van syntactische en morfologische vaardigheden. Maar wat betreft de periode waarin de ontwikkelingen plaatsvinden, is er een ruime overlap. De ontwikkeling van een woordenschat verloopt simultaan met ontwikkelingen van vaardigheden waarvan het startpunt later ligt.

Waarvan is de taalontwikkeling, preciezer gezegd de snelheid waarmee het kind zijn taalbeheersing ontwikkelt, afhankelijk? De hoofdstelling is dat de taalontwikkeling van kinderen een functie is van twee grootheden: het mentale vermogen om taal te leren en de ervaringen die kinderen tijdens hun leven met taal opdoen.

Op het *taalleervermogen* wordt hier slechts kort ingegaan. Het is in dit verband van minder belang, omdat voorlezen een exponent is van de andere grootheid, de ervaringen met taal. Men veronderstelt dat kinderen door hun aanleg over een aangeboren vermogen beschikken om taal te leren: kinderen zijn daartoe als het ware voorgeprogrammeerd. Biologische groei en ervaringen met taal (het taalaanbod) zorgen ervoor dat het taalleervermogen tot ontwikkeling komt. Verschillen tussen kinderen wat betreft hun taalleervermogen zijn derhalve op drie bronnen terug te voeren: ze zijn deels door aanleg en biologische groei bepaald, maar deels ook het resultaat van de mate waarin het taalaanbod het taalleervermogen stimuleert. Deze zienswijze impliceert dat de beide grootheden, het taalleervermogen en de ervaringen met taal, niet onafhankelijk van elkaar opere-

ren maar in een interactie de taalontwikkeling bepalen.

De *ervaringen die kinderen met taal opdoen* is de tweede belangrijke grootheid. De aard en de hoeveelheid van deze ervaringen bepalen mede hoe snel taalontwikkeling plaatsvindt. Wat betreft de aard van de ervaring is het nuttig enkele onderscheidingen aan te brengen, omdat deze laten zien dat de ervaringen met taal van grote diversiteit zijn.

Een eerste onderscheiding is die tussen taal die mondeling en taal die schriftelijk aangeboden wordt. Vanzelfsprekend is in de aanvangsfase alleen het mondelinge taalaanbod van belang. Maar als kinderen eenmaal zelf kunnen lezen, neemt hun vaardigheid toe om via schriftelijk taalaanbod hun taalontwikkeling te versnellen. In feite is er bij voorlezen nog een derde vorm van taalaanbod in het geding, geschreven taal die mondeling wordt aangeboden. Door voorlezen ontdekken kinderen allerlei eigenschappen aangaande de vorm en functie van geschreven taal en komen ze in aanraking met het typische schrijftaalregister. Onderzoek naar ontluikende geletterdheid heeft duidelijk gemaakt dat veel kinderen al vóórdat ze als vierjarige de basisschool betreden, een zeker inzicht hebben in vorm en gebruik van geschreven taal (Sulzby & Teale, 1991; Verhoeven, 1994).

Een tweede onderscheiding is die tussen gezin en school. Traditioneel is wat betreft de eerste taalontwikkeling – zeg tot ongeveer vier jaar – de aandacht vooral naar het gezin uitgegaan. Gezien het feit dat jonge kinderen wat betreft hun taalontwikkeling onder sterke invloed van hun ouders staan, valt dat te begrijpen. Bij de bestudering van de latere taalontwikkeling is er vooral aandacht voor de beïnvloeding vanuit de school. Men vergeet daarbij wel eens dat kinderen ook in de schoolleeftijd nog steeds een invloed ondervinden van de talige omgeving thuis. Het lijkt waarschijnlijk dat het gezin, óók voor kinderen in de schoolleeftijd, een belangrijke determinant van de taalontwikkeling is (Wells, 1986).

In het verleden werd wel een derde onderscheid gemaakt, namelijk tussen ongestuurde taalverwerving, die zich buiten de schoolmuren zou afspelen en gestuurde taalverwerving, die uitsluitend binnen school zou plaatsvinden. Maar deze beide termen – gestuurd en onge-

stuurd – suggereren een tegenstelling die er in werkelijkheid niet is. Immers, ook ouders ondersteunen bewust en actief de taalverwerking van hun kinderen, zo goed als kinderen op school ook via ongestuurde situaties hun taalbeheersing vergroten.

2.2 Verbale interactie

Een kind ontwikkelt zijn taalvaardigheid door middel van de taal die hem wordt aangeboden. In dat taalaanbod vallen twee uitersten te onderscheiden: taal die in een directe interactiesituatie rechtstreeks tot het kind wordt gericht en taal die het kind uit de omgeving oppikt, maar die niet rechtstreeks tot het kind is gericht. Algemeen heerst het inzicht dat het eerste type – de verbale interactie – voor de taalontwikkeling van veel groter belang is dan het tweede type, de omgevingstaal. Anders gezegd, verbale interacties vormen de sleutel tot de taalontwikkeling. Er zijn vier redenen aan te geven waarom de verbale interactie een zoveel krachtiger 'motor' is dan de omgevings-taal:

- a. door de interactie wordt de aandacht van het kind gericht;
- b. de noodzaak de interactie op gang te houden, dwingt volwassenen tot begrijpelijke taal, aangepast aan het taalniveau van het kind;
- c. de interactie fungeert als een sterke uitnodiging aan het kind om zelf taal te gebruiken en met taalvormen te experimenteren;
- d. de interactie voorziet het kind van uiteenlopende vormen van feedback.

In de literatuur worden uiteenlopende voorbeelden gegeven van effectieve interactievormen of -strategieën. Zo maken ouders in de allereerste fase van de taalontwikkeling gebruik van een specifieke taalvariant, het 'motherese', waarvoor in ons land als term het 'Moeders' of 'Mammiaans' is voorgesteld. Deze variant kenmerkt zich onder meer door een grote expressiviteit, een nadrukkelijke articulatie, een laag spreektempo, een sterk vereenvoudigde syntaxis en een specifiek idioom (Gleitman, Newport & Gleitman, 1984). Bruner (1983) wijst op het grote belang van routinematige procedures ('routines'), die als het ware uitnodigen tot herhaling van interacties: spelletjes, vaste gebeurtenissen zoals aankleden, wassen, eten en dergelijke. Die

routinematige procedures lijken ook een essentieel onderdeel te vormen bij het gebruik van aanwijsprentenboeken, zoals beschreven door Ninio en Bruner (1978). Schaerlaekens en Gillis (1987) geven een ander belangrijk aspect aan: semantische contingentie. Ouders proberen in de interactie als regel de inhoudelijke relatie te bewaken; zij borduren – inhoudelijk gezien – voort op de bijdragen van het kind. Strategieën die zij hierbij gebruiken zijn onder andere: expansie (herhaling van de uiting van het kind, maar in een meer volledige vorm), bevestigen van beweringen van het kind, vragen om verduidelijkingen, aanbrengen van correcties, enzovoort. Dergelijke strategieën vormen als het ware de tijdelijke steigers die kinderen in staat stellen aan hun taal te bouwen. Bruner (1983) spreekt in dit verband van 'scaffolding', een techniek waarmee ouders uitbreiding geven aan wat Vygotskij de 'zone van de naaste ontwikkeling' heeft genoemd. Zodra het kind laat blijken de aangeboden steun niet meer nodig te hebben, wordt deze steun ook niet meer aangeboden.

2.3 Taalontwikkeling door voorlezen

Om te begrijpen waarom voorlezen de taalontwikkeling van jonge kinderen kan versnellen, is het nodig aandacht te schenken aan de verschillende manieren van voorlezen. Hierin zijn twee uitersten te onderscheiden: simpel gezegd, 'praten met het kind' en 'praten tegen het kind'. Bij de eerste manier van voorlezen is er zeer veel interactie tussen voorlezer en kind over de tekst, bij de tweede manier is die interactie nagenoeg afwezig; de voorlezer draagt de tekst voor en de rol van het kind is teruggebracht tot luisteren. De eerste manier duiden we aan als een interactieve voorleesstijl, de tweede als een passieve voorleesstijl.

Er zijn redenen om van een interactieve voorleesstijl een veel sterker effect op de taalontwikkeling te verwachten dan van een passieve voorleesstijl. In feite zijn dit ongeveer dezelfde redenen als waarom het aangaan van verbale interacties voor de taalontwikkeling veel gunstiger is dan het aanbieden van omgevingstaal. Voorlezen vindt veelal plaats in een specifieke 'setting' waarin de handelingen met een grote mate van routine vervuld worden. Die routinematigheid zorgt voor een bepaalde mate van voorspelbaarheid. Door die voorspel-

baarheid worden kinderen in staat gesteld langzamerhand een steeds groter aantal handelingen, eerder nog verricht door de voorlezer, zelfstandig te verrichten. Men kan denken aan het aanreiken van het boek, het omslaan van de bladzijden, het aanvullen van het laatste woord, of, in een later stadium, het parafraseren van delen van het verhaal. Ondanks het routinematige karakter is voorlezen geen rigide activiteit. De 'setting' biedt voldoende mogelijkheden voor een zekere variatie in de herhaling.

Een ander belangrijk kenmerk is dat de interactieve voorleesstijl volop gelegenheid biedt tot de al genoemde 'scaffolding'. De voorlezer biedt het kind ondersteuning bij de betekenisreconstructie, waartoe het kind zelfstandig nog niet in staat is. Die ondersteuning biedt het kind als het ware de mogelijkheid de eigen beperkingen te overstijgen. Van belang is dat de voorlezer in opeenvolgende situaties de mate van ondersteuning vermindert, waardoor het kind de mogelijkheid krijgt steeds grotere delen van de betekenisreconstructie zelfstandig te verrichten.

Rest hier de bespreking van de vraag welke taalontwikkelingscomponenten bij voorlezen aan bod kunnen komen. De verwachting is dat de effecten zich op een breed terrein kunnen manifesteren: inzicht in verhaalstructuren, woordenschat, vaardigheden op het gebied van woord- en zinsbouw, meta-linguïstisch bewustzijn en vaardigheden op het gebied van de vroege geletterdheid. Op voorhand is het moeilijk om aan te geven op welk gebied de grootste effecten te verwachten zijn. De effectgrootte zou kunnen samenhangen met de leeftijd van de kinderen en mogelijk ook met de taalcomponent die in de interactie over de tekst centraal gesteld wordt. Een betrekkelijk algemene verwachting is dat de effecten op mondelinge taalvaardigheidsaspecten groter zullen zijn dan op de schriftelijke. Deze verwachting is ingegeven door het feit dat voorlezen – althans voor de kinderen – een voornamelijk mondelinge aanleggenheid is. Ze oefenen veel meer hun luister- en spreekvaardigheid (inclusief woordenschat en grammaticale vaardigheden) dan hun geletterdheid. Enige steun voor deze algemene verwachting werd gevonden door Bus et al. (1995), die voor mondelinge taalvaardigheid een iets sterker effect rapporteren ($d = .67$) dan voor ontlukende geletterdheid ($d = .58$) of

voor lezen ($d = .55$).

2.4 Hypothesen

Het voorgaande leidt tot de algemene verwachting dat het effect van voorlezen op school op de taalontwikkeling van jonge kinderen aan diverse invloeden bloot staat. Deze algemene verwachting is in vijf hypothesen gespecificeerd.

Hypothese 1: voorlezen is gunstiger voor de taalontwikkeling naarmate er tussen voorlezer en kind meer interactie is over de voorgelezen tekst.

Hypothese 2: in vergelijking met het voorlezen door ouders thuis valt te verwachten dat het effect van voorlezen door leerkrachten op school geringer zal zijn, geringer derhalve dan .6 (Cohens d), het resultaat uit de meta-analyses van Bus et al. (1995) en Scarborough en Dobrich (1994).

Hypothese 3: de grootte van de groep waaraan wordt voorgelezen hangt negatief samen met de grootte van het effect.

Hypothese 4: het effect van voorlezen zal groter zijn op de mondelinge taalontwikkeling dan op de schriftelijke taalontwikkeling.

Hypothese 5: naarmate het voorlezen op school vaker gebeurt, zal de effectgrootte toenemen.

3 Opzet

3.1 Verzameling en selectie van de studies

Searches in elektronische literatuurbestanden hebben een belangrijke basis gelegd voor de verzamelde literatuurverwijzingen. Deze searches zijn uitgevoerd aan de hand van zoekprofielen op basis van drie groepen descriptoren: een groep ter aanduiding van de onafhankelijke variabele ('storybook reading, story reading, reading aloud to others'), een groep ter aanduiding van de afhankelijke variabelen ('reading comprehension, listening comprehension, vocabulary development') en een groep ter aanduiding van de leeftijd van de kinderen ('early childhood education, preschool education, primary education'). Deze keuze van descriptoren is mede afgestemd op de descriptoren die aan reeds beschikbare kernartikelen bleken te zijn toegekend. De searches zijn uitgevoerd in juni 1995 op vier bestanden:

ERIC (vanaf 1965 tot en met maart 1995), Dissertation Abstracts International (vanaf 1980, het jaar waarin men begonnen is met het opne- men van abstracts, tot en met mei 1995), Language Learning en Behavior Abstracts (1973 tot en met december 1994) en PsycLit (vanaf 1974 tot en met maart 1995). Het leeu- wendeel van de relevante referenties werd in ERIC gevonden. De andere drie bestanden leverden weinig bruikbare verwijzingen op.

Naast de searches is ook handmatig gezocht in literatuurlijsten van reeds beschikbare publi- caties, waaronder de eerdere bibliografieën van McCormick (1981) en Radecki (1987). Naar- mate de aan opsporing bestede tijd toenam, nam het aantal gevonden referenties af. Omdat is doorgezocht totdat geen nieuwe referenties meer gevonden werden, bestaat de indruk dat de populatie van relevante referenties bij bena- dering volledig verzameld is. In totaal handel-

Tabel 1
Overzicht van gecodeerde studiegenmerken

Studiegenmerk	Schaal
1. Jaar van publiceren	jaartal
2. Publicatietype	1. artikel 2. ongepubliceerd (these, rapport, congresbijdrage, etc.)
3. Omvang van de experimentele groep	aantal leerlingen
4. Omvang van de controlegroep	aantal leerlingen
5. Sociaal-economische status van de leerlingen	1. achterstandssituatie 2. gemengd 3. voornamelijk niet uit achterstandssituatie
6. Natie waar het onderzoek verricht is	1. Verenigde Staten 2. elders
7. Leeftijd leerlingen, gemiddeld bij de start van het onderzoek	aantal maanden
8. Onderzoeksopzet	1. experimenteel 2. correlatieel
9. Controle op samenstelling experi- mentele en controle groep	1. matching 2. randomisatie
10. Duur van het experimentele pro- gramma	aantal weken
11. Aantal contacturen in het experi- mentele programma	aantal uren
12. Gemiddelde groeps grootte tijdens voorlezen	aantal leerlingen
13. Scholing van de voorlezers	1. geen speciale scholing 2. alleen schriftelijke scholing 3. schrifte- lijke en mondelinge vormen van scholing
14. Aard van voorlezen	1. alleen voorlezen 2. 'aangekleed' met activiteiten direct voor of na het voorlezen 3. ingebed in verdergaande activiteiten
15. Herhaalde aanbieding van hetzelf- de boek	0. nee 1. ja
16. Bijzondere aandacht voor de selec- tie van aan te bieden boeken	0. nee 1. ja
17. Aard van de voorlezers	1. leerkrachten aan eigen leerlingen 2. 'vreemden'
18. Bezigheden controlegroep	1. standaardprogramma 2. leesrijheidscurriculum
19. 'Bias' in effectgrootte	0. geen 'bias' of 'bias' onbekend 1. positieve 'bias', bijvoorbeeld als gevolg van matching of covariantie-analyse
20. Domein	1. mondelinge taal 2. lezen
21. Geschatte effectgrootte	uitgedrukt als het gestandaardiseerde verschil in gemiddelde tus- sen experimentele en controle groep
22. Standaardfout van de geschatte effectgrootte	uitgedrukt op dezelfde schaal als de effectgrootte

de het om 63 literatuurverwijzingen.

Niet alle betrokken literatuur is daadwerkelijk verzameld, bestudeerd of in de review opgenomen. Soms bleek de taal een beletsel (drie publicaties waren in het Hebreeuws), zes publicaties (met name oudere dissertaties en congresbijdragen) konden niet geleverd worden, drie onderzoeksrapporten beschreven onderzoek dat later ook als tijdschriftartikel verscheen, in zeven gevallen bleek voorlezen op school geen onderdeel van de 'treatment' te zijn, twee publicaties betroffen onderzoek onder uitsluitend migrantenkinderen. Bij nadere bestudering vielen nog 22 publicaties af, ofwel omdat er van empirisch onderzoek geen sprake was (in zeven gevallen), ofwel omdat er geen controlegroep in de onderzoeksopzet aanwezig was (in twaalf gevallen), ofwel omdat de rapportage zo onvolledig was dat er geen effectgroottes geschat konden worden (in drie gevallen). Acht publicaties zijn buiten beschouwing gebleven, omdat als 'treatment' slechts één prentenboek was voorgelezen. Ten slotte, Morrow (1988) is verwijderd vanwege het afwijkende karakter van haar afhankelijke variabelen. Zij meet het effect van voorlezen op de taalontwikkeling uitsluitend af aan de 'responsiveness', geoperationaliseerd als het aantal opmerkingen van kinderen tijdens het voorlezen.

In totaal zijn derhalve 52 van de 63 publicaties terzijde geschoven en bleven er elf bruikbare studies over. Deze zijn in de literatuurlijst met een asterisk aangegeven.

3.2 Selectie en codering van variabelen

Er zijn 22 variabelen geselecteerd voor codering (zie Tabel 1). De variabelen kunnen als volgt worden gerubriceerd: publicatiestatus (variabelen 1 en 2), beschrijving steekproef (variabelen 3 tot en met 7), onderzoekstype (variabelen 8 en 9), typering van voorlezen (variabelen 10 tot en met 17), typering activiteiten in controlegroep (variabele 18).

Variabele 19 ('bias' in effectgrootte) is toegevoegd, omdat in een aantal onderzoeken gebruik is gemaakt van technieken om de variantie in de afhankelijke variabele te reduceren. Het gaat met name om onderzoeken waarbij leerlingen gematcht zijn of waarin covariantieanalyse is toegepast. Zulke technieken zijn weliswaar verdedigbaar vanuit de wens de

gevoeligheid van de onderzoeksopzet te vergroten. Maar het effect is een positieve 'bias' van de effectgrootte, een 'bias' die groter is naarmate de matchingsvariabelen of covariaten sterker samenhangen met de afhankelijke variabelen. Opname van variabele 19 maakt het mogelijk om achteraf een schatting te doen van de grootte van de 'bias'.

Variabele 20 (domein) is het resultaat van een post-hoc classificatie. De afhankelijke variabelen zijn onderverdeeld naar het domein waarop ze betrekking hebben: mondelinge taal (inclusief woordenschat) en lezen. In beginsel verdient een fijnere indeling de voorkeur, maar het geringe aantal studies maakte een fijnere indeling onpraktisch.

Variabelen 21 en 22 betreffen respectievelijk de effectgrootte en de bijbehorende standaardfout. Gebruik is gemaakt van Cohens *d*, waarbij de standaardisatie is uitgevoerd via de 'within group' standaardafwijking. De standaardfout is geschat volgens Hedges en Olkin (1985, p. 86, Eq. 15). De standaardfout is een functie van de steekproefgrootte en de effectgrootte. Naarmate de steekproef kleiner is of de effectgrootte groter, is de standaardfout van de effectgrootte kleiner.

De codering van de studies volgens de in Tabel 1 aangegeven variabelen bleek in het algemeen eenvoudig. Eén kwestie vroeg evenwel extra aandacht. Aangezien de meeste publicaties geen schatting geven van de effectgrootte, moest deze dikwijls worden afgeleid uit de wèl vermelde gegevens. Hierbij is zoveel mogelijk gebruik gemaakt van omzettingstabellen in leerboeken voor statistiek. In de enkele gevallen dat zulks onmogelijk bleek, is gebruik gemaakt van benaderingsformules zoals gegeven in Mullen (1989).

3.3 Integratie van effecten

Het resultaat van de coderingsfase was een bestand met resultaten uit 12 studies¹. In dit bestand waren 59 geschatte effectgroottes aanwezig. Om te vermijden dat studies met meer uitkomsten een zwaarder stempel zouden drukken, zijn de effectgroottes in twee stappen naar een hoger aggregatieniveau gebracht. In de eerste stap is de 'studie x domein' combinatie als eenheid gekozen. Anders gezegd, de uitkomsten zijn binnen elke studie per domein vastgesteld. Dit resulteerde in een bestand met 12 stu-

dies en 20 uitkomsten (elf voor het domein 'mondelinge taal', negen voor het domein 'lezen'. In de tweede stap zijn de 20 uitkomsten geaggregeerd door per studie de uitkomsten samen te nemen. Dit resulteerde in een bestand met twaalf uitkomsten, voor elke studie één gecombineerde uitkomst.

Beide aggregaties zijn verricht volgens het model van de 'weighted integration method', waarbij de uitkomsten worden gewogen naar hun standaardfout: hoe groter de standaardfout, des te geringer het gewicht. De geaggregeerde effectgroottes en de bijbehorende standaardfouten zijn geschat volgens Hedges en Olkin (1985, p. 112, Eqs. 8 en 9). Aanname bij deze manier van aggregeren is dat de uitkomsten binnen één studie homogeen zijn en alleen van elkaar verschillen op basis van toevallige verschillen tussen de afhankelijke variabelen.

Voor het integreren van de uitkomsten van verschillende studies lijkt deze aanname te beperkend. Verschillen in uitkomsten kunnen dan immers ook het gevolg zijn van systematische verschillen in de experimentele opzet of verschillen tussen de populaties waaruit steekproeven getrokken zijn. Daarom is voor het integreren van de uitkomsten van verschillende studies het 'random effects' model toegepast (Hedges & Olkin, 1985, p. 189 e.v.). In dit model wordt de effectgrootte voor studie j , d_j , gesplitst in twee componenten: de ware effectgrootte δ_j en de foutencomponent ϵ_j :

$$d_j = \delta_j + \epsilon_j \quad (1)$$

Dankzij het feit dat de varianties van d_j bekend zijn, biedt het model de mogelijkheid de variantie van δ_j , de variantie van de ware effectgrootte te onderscheiden van de variantie van ϵ_j , de foutenvariantie. Het is vervolgens mogelijk te toetsen of de ware parametervariantie statistisch significant van nul verschilt. Indien dat inderdaad het geval blijkt te zijn, wordt de verzameling effectgroottes als heterogeen beschouwd. Het loont dan de moeite te trachten de heterogeniteit in verband te brengen met de diverse studiekekenmerken.

De beide aggregaties op basis van de 'weighted integration method' zijn verricht met het programma META (Schwarzer, 1989). De analyses volgens het 'random effects' model zijn uitgevoerd met behulp van een spe-

ciale versie van HLM (namelijk VKHLM, zie Bryk, Raudenbusch & Congdon, 1994). In tegenstelling tot META biedt VKHLM de mogelijkheid studiekekenmerken als predictoren in de analyse op te nemen.

De resultaten van de beide aggregaties zijn in Tabel 2 opgenomen. De gerapporteerde effectgroottes zijn erg heterogeen: ze vertonen een groot bereik (van -1.24 tot en met 1.59).

4 Resultaten

Allereerst wordt, in paragraaf 4.1, een kwalitatieve beschrijving en evaluatie gegeven van de geselecteerde studies. In paragraaf 4.2 worden de studieuitkomsten kwantitatief samengevat en in verband gebracht met de studiekekenmerken.

4.1 De studies kwalitatief beschouwd

De geselecteerde studies zijn over het algemeen van recente datum. Ze zijn gepubliceerd tussen 1968 en 1994 (mediaan: 1990). Zeven studies zijn als artikel gepubliceerd, vier behoren er tot wat wel als grijze literatuur wordt aangeduid.

Qua onderzoekstype zijn de studies tamelijk homogeen: in elf van de twaalf gevallen is er sprake van een experimentele opzet met een controlegroep. Meyer et al. (1994) zijn de enigen die over een survey-opzet rapporteren. De studies zijn over het algemeen tamelijk kleinschalig. De gemiddelde omvang van de experimentele groep bedraagt 50 kinderen (minimum: 10, maximum: 155). In de meeste studies wordt extra zorg besteed aan de vergelijkbaarheid van de experimentele en controlegroep, soms via 'matching', maar vaker post-hoc via covariantieanalyses.

Desalniettemin vertonen diverse studies op het vlak van de onderzoeksoptzet duidelijke tekortkomingen. De studie van Cohen (1968) schiet tekort doordat er geen controle was op de uitvoering van de experimentele conditie. Maar er zijn meer studies die weinig inzicht geven in de implementatie van het experimentele voorleesprogramma. In twee studies – Feitelson, Kita & Goldstein, 1986; Feitelson et al., 1993 – wordt een sterke oppositie van de leerkrachten tegen de voorleesconditie gemeld. In de eerste van de genoemde studies zijn om deze reden

Tabel 2

De geselecteerde studies: aantal proefpersonen in experimentele en controlegroep (N_{exc}), domein (1 = mondelinge taal, 2 = lezen), effectgrootte (d) en standaardfout van de effectgrootte (se_d)

Studie	N_{exc}	Domein	d	se_d
Cohen, 1968	155+130	1	.31	.07
		2	.15	.08
		gecombineerd	.23	.08
Feitelson et al., 1991	19+19	1	.91	.15
Feitelson et al., 1993	86+49	1	.53	.07
Feitelson et al., 1986	31+57	1	.19	.10
		2	.67	.16
		gecombineerd	.42	.16
Karweit, 1989; pre-K	43+43	1	.52	.16
		2	.52	.22
		gecombineerd	.52	.13
idem; Kindergarten	60+60	1	.43	.13
		2	1.33	.12
		gecombineerd	.70	.11
Khare, 1992	38+38	2	.30	.23
Meyer et al., 1994	325	1	.45	.11
		2	-1.24	.05
		gecombineerd	-.33	.08
Morrow et al., 1990	32+30	1	1.59	.21
		2	.08	.10
		gecombineerd	.72	.19
Peterman, 1988	60+60	1	.61	.13
		2	.16	.13
		gecombineerd	.38	.13
Sullivan, 1988	14+14	1	.57	.28
		2	.00	.38
		gecombineerd	.27	.27
Valdez-Menchaca & Whitehurst, 1992	10+10	1	1.49	.36

zelfs twee van de drie experimentele klassen voortijdig uit het onderzoek gestapt. De rapportage heeft alleen op de overblijvende klas betrekking. Zoiets kan de uitkomsten sterk vertekenen, indien bijvoorbeeld de overblijvende leerkracht over bijzondere capaciteiten zou beschikken. De correlatieve studie van Meyer et al. (1994) schiet ernstig tekort op het vlak van de data-analyse. In de eerste plaats wordt voorbij gegaan aan de multiniveau structuur van de data. Ernstiger is dat er alleen 'zero-order' correlaties worden gerapporteerd. Veel krachtiger multivariate analysetechnieken – waaronder multipale regressie en lineair-structurele modellen – worden niet toegepast.

Over de proefpersonen valt het volgende te melden. In acht van de twaalf studies kwamen de kinderen uit een achterstandssituatie. Het accent ligt op kinderen van vijf jaar oud. De gemiddelde leeftijd bedraagt 66 maanden (minimum: 31 maanden, maximum: 90 maanden). Acht van de twaalf studies zijn in de Verenigde Staten verricht. De steekproef van Val-

dez-Menchaca en Whitehurst (1992) steekt af bij die in andere onderzoeken. Het betreft tweejarige Mexicaanse kinderen uit achterstandssituaties die een kinderdagopvang bezochten.

De gemiddelde duur van het experimentele programma bedroeg 24 weken (minimum: 3 weken, maximum: 40 weken). Uitgedrukt in contacturen bedraagt de gemiddelde duur 42 uur (minimum: 5 uur, maximum 140 uur). De gemiddelde grootte van de groep waaraan werd voorgelezen bedroeg 19 kinderen. Slechts in drie studies was de gemiddelde groepsgrootte minder dan tien kinderen. In zes studies werd alleen maar aan de kinderen voorgelezen, zonder 'follow-up' activiteiten. In acht studies ontvingen de voorlezers geen aparte scholing. Meestal waren het de eigen leerkrachten die aan de kinderen voorlezen. Over de bezigheden van de controlegroep bevatten tien studies geen of bijna geen nadere aanduidingen.

Wat betreft de afhankelijke variabelen is een algemene tekortkoming in de rapportages dat over de betrouwbaarheid van de afhankelijk-

ke variabelen nagenoeg geen gegevens vermeld worden. Een tweede probleem is de aard of de betekenis van de variabelen. De hier gehanteerde indeling in twee domeinen (mondelinge taal, lezen) verbergt in feite een enorme variëteit aan afhankelijke variabelen. De indruk bestaat dat met name sommige instrumenten voor de mondelinge taalbeheersing in standaardisatie tekort schieten. Morrow, O'Connor en Smith (1990) en Peterman (1988) maken gebruik van 'free recall' en 'probed recall' taken, waarbij kinderen een voorgelezen boek zo goed mogelijk moeten proberen na te vertellen. Behalve een gebrek aan standaardisatie bestaat tegen deze instrumenten nog een tweede bezwaar. Ze leunen sterk op wat de kinderen in de experimentele conditie is aangeleerd: luisteren naar en praten over voorgelezen boeken. De standaardisatie van de maten voor het domein 'lezen' is van een hoger niveau. Maar hier doet zich het verschijnsel voor dat de maten erg sterk uiteenlopen: van toetsen voor ontluikende geletterdheid (o.a. 'Sulzby's scale for reenactments' en de toets 'Concepts about print') tot en met testen voor technisch en begrijpend lezen. Die diversiteit zou gemakkelijk kunnen verbergen dat voorlezen voor sommige aspecten van geletterdheid veel gunstiger is dan voor andere aspecten.

Als laatste punt van deze paragraaf wordt aandacht besteed aan afwijkende effectgroottes, waaronder hier de effectgroottes begrepen worden die groter dan 1.00 (absoluut) zijn. Het zijn er vier, twee voor het domein mondelinge taal en twee voor het domein lezen. De studie van Morrow, O'Connor en Smith (1990) resulteerde in een effectgrootte van 1.59 voor het domein mondelinge taal. In het voorgaande is al naar voren gebracht dat de afhankelijke maten in deze studie dicht tegen de experimentele conditie aanliggen. De effectgrootte voor het domein lezen in dezelfde studie, geoperationaliseerd via maten die veel verder afliggen van de experimentele conditie, bedroeg slechts .08. Het sterke effect (domein mondelinge taal; effectgrootte 1.49) in de studie van Valdez-Menchaca en Whitehurst (1992) is gerealiseerd in een experimentele groep van slechts tien kinderen aan wie in een één-op-één situatie werd voorgelezen. Het is de enige studie waarin één-op-één voorgelezen is. Opvallend is verder dat de experimentele conditie niet erg

intensief is geweest: 30 sessies van circa tien minuten interactief voorlezen. In feite is in deze studie het laagste aantal voorleesuren gerealiseerd.

De studie van Karweit (1989) in de kindergarten-steekproef resulteerde in een effectgrootte van 1.33 voor lezen. Het effect is sterk beïnvloed door een effectgrootte van 3.75 op de 'Woodcock word attack test', waarbij de auteur aangeeft dat de controlegroep in de door de test gemeten vaardigheid – in tegenstelling tot de experimentele groep – geen enkele oefening heeft gehad. Deze opmerking suggereert dat de verschillen tussen de experimentele en controle conditie veel verder reikten dan alleen het voorlees-programma, een suggestie die wordt versterkt door het feit dat de experimentele scholen proefscholen waren voor het 'Success for All' programma van Slavin. De vierde afwijkende effectgrootte (domein lezen; effectgrootte -1.24) is gebaseerd op de correlatieve studie van Meyer et al. (1994). Het is in feite een gecombineerde effectgrootte, op basis van vier tests voor technisch lezen die alle vier sterk negatief correleerden met de tijd die leerkrachten aan voorlezen besteedden. Diezelfde tests correleerden sterk positief met de tijd die leerkrachten aan specifieke oefeningen voor aanvankelijk lezen besteedden.

4.2 De gerapporteerde effecten

Voor het domein 'mondelinge taal' blijkt de gecombineerde effectgrootte .64 te bedragen (zie Tabel 3). Gezien de bijbehorende standaardfout verschilt deze waarde statistisch significant van nul. Voor het domein 'lezen' bedraagt de gecombineerde effectgrootte .22, een waarde die statistisch niet significant van nul verschilt. De 'overall' effectgrootte, waarin de resultaten over de twee domeinen samengevoegd zijn, bedraagt .47. Deze waarde verschilt statistisch significant van nul. In alle drie de gevallen wijst de heterogeniteitstoets uit dat de uitkomsten heterogeen zijn. Dit rechtvaardigt het zoeken naar studiekenmerken die met de effectgrootte samenhangen.

Gezien het geringe aantal studies zijn de heterogeniteitsanalyses op een terughoudende manier verricht. Per afzonderlijke analyse is steeds maar één studiekenmerk als predictor ingevoerd. Bestudeerd is of opname van een predictor leidde tot een reductie van de hete-

Tabel 3

Uitkomsten van de analyse volgens het 'Random effects' model (model zonder predictoren), waarbij δ de ware effectgrootte is

Domein (aantal studies)	'Fixed effect'		'Random effect'			
	δ	se_{δ}	$Var(\delta_j)$	df	χ^2	p -waarde
Mondelinge taal (11)	.64	.12	.14	10	80.7	.00
Lezen (9)	.22	.24	.48	8	946.9	.00
Gecombineerd (12)	.47	.11	.13	11	135.0	.00

rogeniteit (i.e. de variantie van δ). De reductie is uitgedrukt als een percentage, waarbij de heterogeniteit van het nulmodel – het model zonder predictoren – op 100 procent gesteld is. Bedraagt de reductie voor een predictor bijvoorbeeld 10 procent, dan betekent dit dat toevoeging van de predictor aan het nulmodel de heterogeniteit van de ware effectgrootte met 10 procent doet afnemen. De analyses zijn verricht voor bijna alle in Tabel 1 opgenomen studiekekenmerken, uitgezonderd kenmerk 4 (de omvang van de controlegroep; deze variabele overlapt sterk met kenmerk 3, de omvang van de experimentele groep) en kenmerk 8 (er was slechts één studie met een correlatieve opzet).

Uit Tabel 4 komt naar voren dat diverse studiekekenmerken sterk samenhangen met de effectgrootte. Maar tevens blijkt dat het beeld voor de drie domeinen (mondelinge taal, lezen, gecombineerd) niet gelijk is. Opvallend voor

het domein 'mondelinge taal' is dat het kenmerk 'groeps-grootte' sterk negatief geassocieerd is met de effectgrootte: de sterkere effecten worden gevonden in de studies waarin aan kleinere groepen kinderen wordt voorgelezen. Dit kenmerk verklaart maar liefst 98 procent van de heterogeniteit. Nog één ander studiekekenmerk verklaart meer dan 25 procent van de variantie: de leeftijd van de kinderen waaraan wordt voorgelezen (grotere effecten bij jongere kinderen).

Voor het domein 'lezen' geldt dat er vier studiekekenmerken zijn die elk op zich omstreeks 25 procent van de heterogeniteit kunnen verklaren. Eén daarvan is de omvang van de experimentele groep (kenmerk 3), die – evenals bij het domein 'mondelinge taal' – negatief geassocieerd is met de effectgrootte. De andere drie zijn de sociaal-economische status van de leerlingen (grotere effecten bij leerlingen met een lagere status), kenmerk 9 (grotere effecten

Tabel 4

Reductie van de heterogeniteit (in procenten) die het gevolg is van de opname van een studiekekenmerk in het nulmodel; tussen haakjes wordt de richting van het effect aangegeven

Studiekekenmerk ^a	Mondelinge taal (n=11)	Lezen (n=9)	Gecomb. (n=12)
1. Jaar van publiceren	0	0	0
2. Publicatietype	0	0	0
3. Omvang exp. groep	20 (-)	28 (-)	85 (-)
5. SES leerlingen	0	24 (-)	41 (-)
6. Natie	0	0	15 (+)
7. Leeftijd	28 (-)	0	12 (-)
9. Matching/randomisatie	0	25 (-)	34 (-)
10. Duur programma	0	0	0
11. Aantal contacturen	0	0	0
12. Groeps-grootte	98 (-)	0	31 (-)
13. Scholing	0	0	0
14. Aard van voorlezen	0	13 (+)	0
15. Boekherhaling	0	9 (+)	0
16. Boekselectie	0	29 (+)	9 (+)
17. Aard voorlezers	20 (+)	0	15 (+)
18. Bezigheden C-groep	11 (+)	0	0
19. 'Bias' in effectgrootte	0	9 (+)	33 (+)

^a Voor een toelichting op de studiekekenmerken raadplege men Tabel 1

in de studies waarin de leerlingen gematcht zijn) en kenmerk 16 (grotere effecten in de studies waarin expliciete aandacht gegeven werd aan de selectie van de voor te lezen boeken).

Bezien we de gecombineerde uitkomsten, dan valt op dat de steekproefomvang een sterk negatieve relatie vertoont met de effectgrootte (sterkere effecten bij onderzoeken met een kleinere experimentele groep). Dit kenmerk verklaart de heterogeniteit voor 85 procent. Nog vier andere kenmerken vertonen op dit niveau van aggregatie een sterke relatie met de effectgrootte: de sociaal-economische status van de leerlingen (grotere effecten bij leerlingen met een lagere status), kenmerk 9 (grotere effecten in de studies waarin de leerlingen gematcht zijn), de groepsgrootte (grotere effecten in studies waarin aan kleinere groepen kinderen wordt voorgelezen) en kenmerk 19 (grotere effecten in de studies waarin de effect-schattingen waarschijnlijk positief onzuiver zijn).

Al met al is het in Tabel 4 gepresenteerde beeld niet erg eenduidig. Het kleine aantal studies, het kleine aantal leerlingen in de meeste studies en de hiervoor besproken tekorten in de kwaliteit van de afzonderlijke studies dragen hieraan waarschijnlijk bij. In feite is er slechts één kenmerk dat consistent negatief samenhangt met de effectgrootte, namelijk de omvang van de experimentele groep: hoe kleiner de groep, des te groter het gerealiseerde effect.

5 Conclusies en discussie

5.1 Conclusies

Tot welke conclusies leidt deze review van onderzoek naar de effecten van voorlezen op school aan jonge kinderen? Allereerst moet worden vastgesteld dat de empirische basis tamelijk smal is. De basis wordt gevormd door twaalf onafhankelijke steekproeven, verzameld in elf studies. Van belang is dat een aantal van deze studies tamelijk kleinschalig is en qua opzet ernstige tekortkomingen vertoont.

De meest algemene, over de twaalf steekproeven en de twee domeinen gegeneraliseerde effectschatting bedraagt .47, met een standaardfout van .11. Uitgesplitst naar domein bedraagt de effectschatting voor mondelinge

taal .64 (standaardfout .12) en voor lezen .22 (standaardfout .24). In termen van Cohen (1988): voorlezen op school heeft een middelmatig effect op de mondelinge taalontwikkeling van kinderen en een zwak effect op de leesontwikkeling van kinderen. Niet uitgesloten kan worden dat het effect op lezen nul bedraagt, aangezien de nul binnen het 95-procents betrouwbaarheidsinterval valt.

De gegeneraliseerde schattingen verbergen een aanzienlijke variatie in onderzoeksuitkomsten. Bij statistische toetsing blijkt er in alle drie de gevallen (mondelinge taal, lezen, gecombineerd) sprake te zijn van heterogeniteit. Het is een ongunstige omstandigheid dat de heterogeniteit slechts voor een beperkt gedeelte samen blijkt te hangen met de in de review betrokken studiekekenmerken. Voor het domein mondelinge taal zijn er twee kenmerken die de effectgrootte modereren: de effecten zijn sterker bij voorlezen aan kleinere groepen en aan jongere kinderen. Voor het domein lezen zijn er vier modererende kenmerken: de effecten zijn sterker in studies met een kleinere steekproef en in studies waarin kinderen gematcht zijn. Bovendien worden sterkere effecten gerapporteerd in studies met kinderen uit achterstandssituaties en in studies waarin de voorgelezen boeken met zorg gekozen zijn.

De conclusies verlenen steun aan twee van de vijf in paragraaf 2.4 geformuleerde hypothesen. Gesteund wordt de hypothese dat er een 'overall' positief effect is, evenwel zwakker dan het effect van voorlezen door ouders. Men vergelijk de hier gerapporteerde effectgrootte van .47 met de door Bus et al. (1995) gerapporteerde effectgrootte van .59 en de door Scarborough en Dobrich (1994) gerapporteerde mediane correlatie van .26. Bevestigd wordt ten tweede ook de hypothese dat het effect van voorlezen op school voor mondelinge taal sterker is dan voor lezen.

Maar er zijn ook twee hypothesen die in de data geen steun vinden. Het verwachte modererende effect van de grootte van de groep waaraan wordt voorgelezen is alleen voor het domein mondelinge taal aangetroffen. Het verwachte modererende effect van de duur en de intensiteit van het voorleesprogramma is zelfs in geen van beide domeinen waargenomen. Met name de laatste bevinding roept serieuze twijfel op. Immers, als voorlezen op school de

taal- en leesontwikkeling van kinderen ondersteunt, dan mag men toch verwachten dat er een relatie bestaat tussen de duur van de 'treatment' en het daarmee gerealiseerde effect.

Tenslotte, er is één hypothese die in de onderhavige review niet getoetst kon worden. Dat is de verwachting dat interactief voorlezen tot een sterker effect zal leiden dan een passieve voorleesstijl. Vrijwel alle rapportages gaven onvoldoende inzicht in de gehanteerde voorleesstijl. Dit punt brengt ons naar het discussiegedeelte van deze paragraaf, waarin drie onderwerpen worden besproken.

5.2 Discussie

Elke review – dus ook deze – wordt serieus bedreigd door het feit dat de gevonden studies geen getrouwe afspiegeling van al het verrichte onderzoek hoeven te vormen. Rosenthal (1979) spreekt in dit verband over het 'file drawer' probleem, het verschijnsel dat studies met een statistisch significant resultaat een grotere kans hebben om in een tijdschrift gepubliceerd te worden dan studies met een nul-uitkomst. Hoe ernstig deze bedreiging voor de onderhavige review is, valt moeilijk te zeggen. Weliswaar is gebleken dat het studiekekenmerk 'publicatietype' geen modererend effect had op de effectgrootte. Maar hierbij moet worden aangetekend dat met een aantal van slechts twaalf steekproeven geen gevoelige analyses mogelijk zijn. Een andere benadering is een schatting te maken van het zogenaamde 'fail-safe number', het aantal studies met nul-uitkomsten dat nodig is om de effectgrootte van .64 (zoals vastgesteld voor het domein mondelinge taal) terug te brengen tot bijvoorbeeld .20, een waarde die correspondeert met nog slechts een gering effect. Volgens een door Orwin (1983) gegeven formule bedraagt dit 'fail-safe' aantal in het onderhavige voorbeeld 24 studies. Anders gezegd, er zouden 24 niet gepubliceerde studies met een nul-resultaat moeten zijn om het gemiddelde effect van .64 terug te brengen tot .20. Hoe exact dit aantal ook is, het is moeilijk te evalueren bij gebrek aan inzicht in het werkelijke aantal ongepubliceerde studies en de uitkomsten daarvan.

Een andere bedreiging voor de onderhavige review betreft de betrouwbaarheid van de afzonderlijke rapportages. In een aantal gevallen is er reden om hier aan te twijfelen. De stu-

die van Morrow et al. (1990) is ook gepubliceerd als 'doctoral thesis' van O'Connor (1989). Waar in het artikel sprake is van een steekproef van 62 kinderen, rept de dissertatie stelselmatig over een steekproef van 53 kinderen. Zo'n verschil kan een onschuldige verklaring hebben. Maar vreemder is dat de gerapporteerde gemiddelden en standaarddeviaties tot op twee decimalen nauwkeurig gelijk zijn. Feitelson et al. (1993) rapporteren drie grootheden waaruit effectgroottes te destilleren zijn: gemiddelden en standaarddeviaties, F-waarden en percentages verklaarde varianties. Als men ze evenwel alle drie gebruikt, komt men tot verschillende geschatte effectgroottes. In de onderhavige review is overigens gebruik gemaakt van de (na covariantieanalyse) gecorrigeerde gemiddelden en standaarddeviaties. Karweit (1989) maakt een opmerking – waarnaar in het voorafgaande al is verwezen – die twijfel oproept over de verschillen tussen de programma's voor de experimentele en de controlegroep. Het lijkt erop alsof die verschillen meer omvatten dan alleen het voorlezen. Meyer et al. (1994) schrijven dat ze beschikken over de gegevens uit twee cohorten, elk ruim 300 leerlingen bevattend, maar rapporteren vervolgens alleen over het eerste cohort. Al met al roept de kwaliteit van vele rapportages vraagtekens op. Hier ligt zeker een taak voor tijdschriftredacties en beroepsverenigingen om met meer kracht te streven naar uniformiteit en kwaliteit van rapportage.

Het is te hopen dat de onderhavige review geen aanleiding zal geven tot de voldane constatering dat voorlezen op school gunstig is voor de mondelinge taalontwikkeling van kinderen. De empirische basis voor deze constatering is nog veel te smal: te weinig onderzoek, te veel onderzoek met kwalitatieve beperkingen, te veel onbegrepen variatie in de onderzoeksuitkomsten. Beter zou het zijn de onderzoeksinspanningen te vergroten en beter op elkaar af te stemmen. In dat licht worden – ter afsluiting – nog de volgende suggesties gedaan.

Er is behoefte aan studies waarin de duur en/of de intensiteit van de voorleesprogramma's experimenteel wordt gevarieerd. Is het bijvoorbeeld zo dat voorleesprogramma's die zich over een heel schooljaar uitstrekken tot grotere effecten aanleiding geven dan programma's die slechts enkele maanden beslaan? Het

zou verontrustend zijn als de duur van de programma's de effecten niet zou modereren, zoals op dit moment nog de conclusie lijkt te zijn.

Er is behoefte aan studies waarin de groeps-grootte systematisch wordt gevarieerd. Elsäcker (1994) vergeleek de effecten van het voorlezen aan groepen van vijf kleuters met het voorlezen aan klassikale groepen. Morrow en Smith (1990) vergeleken drie condities met elkaar: één-op-één, groepjes van drie kleuters en integrale kleuterklassen. Beide studies leidden tot de conclusie dat voorlezen in kleine groepen (van vijf, respectievelijk drie kleuters) het gunstigst is.

Er is behoefte aan studies waarin de stijl van voorlezen experimenteel wordt gevarieerd. Corcoran Nielsen (1990) vergeleek de effecten van vier verschillende voorleesstijlen, zonder overigens een nul-conditie in haar onderzoek te betrekken. Dickinson en Smith (1994) vergeleken in een survey-opzet drie voorleesstijlen wat betreft de effecten op de taalontwikkeling van kinderen. Burton (1993) demonstreerde in een aantrekkelijk 'multiple baseline design' dat de voorleesstijl van leerkrachten beïnvloedbaar is.

Naast de gesignaleerde behoeften aan experimenteel onderzoek is er ook behoefte aan onderzoek naar processen op micro-niveau. Hoe groot is bijvoorbeeld de kans dat voorgelezen kinderen de betekenis van een onbekend woord oppikken en vasthouden? Welke factoren beïnvloeden die kans? Relevant onderzoek in deze richting is gerapporteerd door Eller, Pappas en Brown (1988), Elley (1989), Leung en Pikulski (1990), Robbins en Ehri (1994), Sénéchal en Cornell (1993) en Sénéchal, Thomas en Monker (1995). Deze onderzoeken tonen aan dat voorlezen – veelal in een één-op-één situatie – inderdaad aanleiding geeft tot woordenschatuitbreiding. Maar ze roepen tevens de vraag op of hetzelfde effect op zou treden bij het voorlezen aan grotere groepen.

Noot

- 1 De studie van Karweit (1989), die over twee steekproeven rapporteert – een 'pre Kindergarten' en een 'Kindergarten' steekproef – is in het gegevensbestand als twee onafhankelijke studies beschouwd.

Literatuur

(De in de review opgenomen studies zijn met een asterisk aangeduid.)

- Baudoin, T., Meelis-Voorma, T., Moolenaar, P. Overmeijer, H., & Ros, B. (1990). *Jeugdliteratuur voor de beroepspraktijk; een leergang voor het HBO*. Groningen: Wolters-Noordhoff.
- Blok, H. (1992). *Lezen voor je plezier, hoe staat het daarmee bij leerlingen op de basisschool?* (SCO-rapport 286). Amsterdam: Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam.
- Bruner, J. (1983). *Child's talk: learning to use language*. New York: Norton.
- Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1994). *HLM 2/3; Hierarchical linear modeling with the HLM/2L and HLM/3L programs*. Chicago: Scientific Software International.
- Burton, Sister A.M. (1993). *Influence of teacher-mediated strategies on the verbal responses of at risk first grade students during small group storybook read-aloud* (doct. diss.). Temple University.
- Bus, A.G., Ijzendoorn, M.H. van, & Pellegrini, A.D. (1995). Joint book reading makes for success in learning to read: a meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, 65, 1-21.
- *Cohen, D.H. (1968). The effects of literacy on vocabulary and reading achievement. *Elementary English*, 45, 209-213, 217.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corcoran Nielsen, D.E. (1990). *The effects of four approaches to group storybook reading in kindergarten on story comprehension, story structure knowledge, and concepts about print* (doct. diss.). University of Minnesota.
- Dickinson, D.K., & Smith, M.W. (1994). Long-term effects of preschool teachers' book readings on low-income children's vocabulary and story comprehension. *Reading Research Quarterly*, 29, 104-122.
- Eller, R.G., Pappas, C.C., & Brown, E. (1988). The lexical development of kindergartners: learning from written context. *Journal of Reading Behavior*, 20, 5-24.
- Elley, W.B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly*, 24, 174-187.

- Elsäcker, W. van (1994). *Prentenboeken voorlezen aan allochtone en autochtone kleuters; een onderzoek naar het effect van groeps grootte op de interactie tijdens het voorlezen en de mate waarin het voorgelezen verhaal begrepen en onthouden wordt*. Tilburg: Werkverband Taal en Minderheden, Letterenfaculteit K.U.B. Tilburg.
- *Feitelson, D., Goldstein, Z., Eshel, M., Flasher, A., Levin, M., & Sharon, S. (1991). *Effects of exposure to literary language on kindergartners' listening comprehension and use of language* (unpublished manuscript). Haifa: University of Haifa.
- *Feitelson, D., Goldstein, Z., Iraqi, J., & Share, D. (1993). Effects of listening to story reading on aspects of literacy acquisition in a diglossic situation. *Reading Research Quarterly*, 28, 70-79.
- *Feitelson, D., Kita, B., & Goldstein, Z. (1986). Effects of listening to series stories on first graders' comprehension and use of language. *Research in the Teaching of English*, 20, 339-356.
- Gleitman, L., Newport, E., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, 11, 43-79.
- Hedges, L.V., & Olkin, I. 1985. *Statistical methods for meta-analysis*. San Diego: Academic Press.
- *Karweit, N. (1989). The effects of a story-reading program on the vocabulary and story comprehension skills of disadvantaged prekindergarten and kindergarten students. *Early Education and Development*, 1, 105-114.
- *Khare, P. (1992). *Effects of oral versus read stories on children's creativity and sense of story structure* (doct. diss.). Birmingham, AL: University of Alabama.
- Leung, C.B., & Pikulski, J.J. (1990). Incidental learning of word meanings by kindergarten and first-grade children through repeated read-aloud events. In J. Zutell & S. McCormick (Eds.), *Literacy theory and research: analyses from multiple paradigms, thirty-ninth yearbook of the National Reading Conference* (pp. 231-240). Chicago: National Reading Conference.
- McCormick, S. (1981). *Reading aloud to preschoolers age 3-6: a review of the research* (paper presented at the Annual Meeting of the Southwest Regional Conference of the International Reading Association; ED 199 657).
- *Meyer, L.A., Stahl, S.A., Linn, R.L., & Wardrop, J.L. (1994). Effects of reading storybooks aloud to children. *Journal of Educational Research*, 88, 69-85.
- Morrow, L.M. (1988). Young children's responses to one-to-one story readings in school settings. *Reading Research Quarterly*, 23, 89-107.
- *Morrow, L.M., O'Connor, E.M., & Smith, J.K. (1990). Effects of a story reading program on the literacy development of at-risk kindergarten children. *Journal of Reading Behavior*, 22(3), 255-275.
- Morrow, L.M., & Smith, J.K. (1990). The effects of group setting on interactive story book reading. *Reading Research Quarterly*, 25, 213-231.
- Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale (NJ): Lawrence Erlbaum Associates.
- Nijmeegse Werkgroep Taal didactiek (1992). *Taal didactiek aan de basis* (4e druk). Groningen: Wolters-Noordhoff.
- Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of Child Language*, 5, 1-15.
- O'Connor, E.M. (1989). *The effect of story reading as an instructional strategy on kindergarten children's literacy skill development*. New Brunswick, NJ: Rutgers State University of New Jersey.
- Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. *Journal for Educational Statistics*, 8, 157-159.
- *Peterman, C.L. (1988). *Successful storyreading procedures: Working with kindergarten teachers to improve children's story understanding* (paper presented at the Annual Meeting of the National Reading Conference (38th, Tucson, AZ, November 29-December 3, 1988; ED 314 739).
- Radecki, K.K. (1987). *An annotated bibliography of the literature examining the importance of adults reading aloud to children* (ED 296 274). South Bend, IN: Indiana University.
- Robbins, C., & Ehri, L.C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *Journal of Educational Psychology*, 86, 54-64.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Scarborough, H.S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14, 245-302.
- Schaerlaekens, A.M., & Gillis, S. (1987). *De taalverwerving van het kind*. Groningen: Wolters-Noordhoff.
- Schwarzer, R. (1989). *Meta-analysis program; program manual*. Berlin: Institut für Psychologie, Freie Universität Berlin.
- Sénéchal, M., & Cornell, E.H. (1993). Vocabulary acquisition through shared reading experiences.

Reading Research Quarterly, 28, 360-374.

Sénéchal, M., Thomas, E., & Monker, J. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, 87, 218-229.

*Sullivan, J.L. (1988). *Reading aloud to children and its effect on their attention span* (Master's Thesis, ED 292 077). Kean College (NJ).

Sulzby, E., & Teale, W. (1991). Emergent literacy. In R. Barr, M.L. Kamil, P. Mosenthal, & P.D. Pearson (Eds.), *Handbook of Reading Research, Vol. II* (pp. 727-757). White Plains (NY): Longman.

*Valdez-Menchaca, M.C., & Whitehurst, G.J. (1992). Accelerating language development through picture book reading: A systematic extension to day-care. *Developmental Psychology*, 28, 106-114.

Veen, A., Overmars, A., & Glopper, K. de (1995). *Boekenpret belicht: onderzoek naar een project Leesbevordering voor 0- tot 6-jarigen*. Utrecht: ISOR/Onderwijsonderzoek, Universiteit Utrecht.

Verhoeven, L. (1994). *Ontluikende geletterdheid*. Lisse: Swets & Zeitlinger.

Wells, G. (1986). *The meaning makers: children learning language and using language to learn*. Portsmouth: Heinemann.

Manuscript aanvaard 7-1-1997

Auteur

H. Blok is verbonden aan het SCO-Kohnstamm Instituut van de Faculteit Pedagogische en Onderwijskundige Wetenschappen van de Universiteit van Amsterdam.

Adres: Wibautstraat 4, 1091 GM Amsterdam. E-mail: henkb@educ.uva.nl

Abstract

Reading to young children in educational settings: a review of research

H. Blok. *Pedagogische Studiën*, 1997, 74, 233-248.

Reading to young children is an almost daily activity in kindergartens. One of the perceived benefits is the facilitation of literacy acquisition. This article reviews 11 studies into the effects of reading to young children in schools. The age of the children varied between 31 and 90 months. Dependent variables were classified in two domains: oral language and reading skills. The combined effect size for the oral language domain was $d = .64$, and for the reading domain $d = .22$. Although the effect in the oral language domain looks promising, caution is needed because the empirical evidence appears to be meager. Not only is the number of studies small, but a critical analysis of the design of the studies reveals general poor quality.