

# De effectiviteit van computerondersteund schrijfonderwijs: een meta-analyse

ERIK VAN SCHOOTEN, RUBEN FUKKINK, KEES DE GLOPPER

*De tekstverwerker is als schrijfgereedschap niet meer uit het onderwijs weg te denken. Veel minder gangbaar is het gebruik van computerprogramma's die het schrijfonderwijs ondersteunen. Zo zijn er programma's die oefeningen bieden voor deelvaardigheden of specifieke deelprocessen ondersteunen. Andere programma's analyseren schrijfproducten en geven commentaar, of bevorderen de sociale en dialogische aspecten van het schrijfproces. Is zulke technologie voor het onderwijs even nuttig als de tekstverwerker dat is voor het schrijven? In dit artikel wordt verslag gedaan van een meta-analyse van de effecten van computerondersteund schrijfonderwijs (COS).<sup>1</sup>*

Van alle vormen van taalvaardigheid is schrijven waarschijnlijk de meest complexe en veel-eisende. Schrijven stelt vaak hoge eisen aan de kennis, vaardigheden en motivatie van auteurs, niet alleen van beginners, maar ook van gevorderden en experts. Uit onderzoek naar schrijfprocessen (Bereiter & Scardamalia 1987, Emig 1971, Flower & Hayes 1981, Levy & Ransdell 1996) is duidelijk geworden wat er allemaal komt kijken bij de productie van geschreven teksten. Schrijvers zijn, zeker wanneer hun taken uitdagender zijn, vaak

op meerdere fronten tegelijk aan het werk: zij moeten ideeën ontwikkelen en kennis 'ontdekken' of construeren, hun gedachten in schrijftaal onder woorden brengen en deze met de pen op papier of via het toetsenbord in schrift uitdrukken. Daarbij moeten zij hun aandacht verdelen over verschillende niveaus (woord, zin, alinea en tekst) en afwisselend plannen, structureren, evalueren en reviseren. Schrijven gaat verder gepaard met een aanhoudend beroep op allerlei vormen van kennis: over taal, teksten, concepten en feiten. Doordat er zoveel verschillende aspecten een rol spelen, doet schrijven een groot beroep op het werkgeheugen (Chanquoy & Alamargot 2002, Kellogg 1994, McCutchen 1996). Complexe schrijftaken leiden niet zelden tot vormen van cognitieve overbelasting.

Schrijven wordt meestal niet vanzelf geleerd. De ontwikkeling van schrijfvaardigheid is sterk afhankelijk van onderwijs. Dat onderwijs kent vele vormen, afhankelijk van de leeftijd en het ontwikkelingsniveau van schrijvers, van de doelen en de context van de instructie en van de gekozen didactiek. Zo kan schrijfinstructie proces- of productgeoriënteerd zijn, kan het strategieën bieden of tekstmodellen, uitgaan van individueel of

samenwerkend leren, zich richten op conceptualiseren of formuleren, enzovoort.

Naar de effectiviteit van verschillende vormen van schrijfonderwijs zijn vele honderden studies verricht. Een belangrijk deel van dat onderzoek is inmiddels al weer heel wat jaren geleden op de rij gezet, eerst door Wesdorp (1983) en korte tijd later door Hillocks (1986). Deze meta-analyses, kwantitatieve samenvattingen en analyses van de uitgevoerde experimenten, hebben overeenkomstige resultaten. Wesdorp laat zien dat instructie in *pre-writing activities* (het genereren van ideeën, het structureren van tekst, discussie vooraf) effectief is. Hetzelfde geldt voor instructie gericht op het formuleren (via oefeningen in het combineren van zinnen), op beoordeling door medeleerlingen en op revisie. De beste leeromgeving is er volgens Hillocks één waarin de leerdoelen duidelijk en specifiek zijn en waarin leerlingen veel samenwerken.

In studies die Wesdorp en Hillocks onder de loep namen, speelde de computer geen rol. Computers speelden in het schrijfonderwijs tot ver in de jaren tachtig van de vorige eeuw sowieso een marginale rol. Inmiddels is het gebruik van met name tekstverwerkers enorm toegenomen (Daiute 1986). Tekstverwerkers vergemakkelijken het schrijfproces (Bangert-Drowns 1993, Cochran-Smith 1991), met name het reviseren wordt gefaciliteerd (Butterfield, Hacker & Albertson 1996). Het gebruik van tekstverwerkers leidt ook tot betere schrijfproducten (MacArthur, Feretti, Okolo & Cavelier 2001). De computer kan tevens gebruikt worden als instructiemiddel ten dienste van het schrijfonderwijs. Computerondersteund schrijfonderwijs is echter niet gangbaar, ondanks enkele decennia van ontwikkelwerk en onderzoek.

Wat voor vormen heeft COS in de loop der jaren zoal aangenomen? Sterkel, Johnson en Sjogren (1986) onderscheiden drie vormen van COS. De eerste toepassingen waren programma's met *drills*: oefeningen voor onder-

delen als spelling, grammatica of zinsbouw. Volgens Williams (2001) neemt de populariteit van zulke programma's tegenwoordig weer toe. De eerste generatie COS-programma's richten zich vaak op toepassing van taalregels, maar kunnen ook gecombineerd worden met een globaler perspectief. Het *Writing-to-Read*-programma bijvoorbeeld onderwijst jonge leerlingen foneem-grafeemrelaties, maar laat ze het geleerde ook toepassen in zelf geschreven verhaaltjes (Leahy 1991). Andere programma's bieden voorbeeldmodellen, oefening en direct individueel commentaar. Soms is dit commentaar adaptief, want afhankelijk van de door de leerling gegeven input.

De tweede generatie betreft interactieve COS-programma's, vaak geïnspireerd op het idee van 'procedurele facilitatie' (oefening in deelprocessen door gelijktijdige ondersteuning of neutralisatie van andere deelprocessen, zie Bereiter & Scardamalia 1987). Meestal richten deze programma's zich op hogere schrijfprocessen en proberen ze leerlingen bewuster te maken van communicatieve doelen. Het programma van Reynolds en Bonk (1996) bijvoorbeeld bevat adviezen gericht op stofvinding ('Wat zou je publiek nog meer willen weten over dit onderwerp') en adviezen gericht op evaluatie ('Blijft je tekst bij het door jou aangekaarte probleem?'). Gebruikers van het programma kunnen adviezen soms oproepen middels functietoetsen, andere keren worden deze ongevraagd geboden. In het programma CAC-2 van Woodruff, Bereiter en Scardamalia (1981) bijvoorbeeld wordt leerlingen eerst de vraag 'Heb je een mening over dit onderwerp?' gesteld. Indien het antwoord bevestigend is, volgt 'Ok, laten we dat dan aan de lezer vertellen'.

Programma's uit de derde groep analyseren schrijfproducten en geven commentaar. De oudste programma's van deze groep betreffen programma's voor spelling- en

grammaticacontrole, die momenteel standaard zijn opgenomen in tekstverwerkers. Een programma als *Writer's Workbench* zoekt fouten in de spelling en de zinsbouw en stelt verbeteringen voor. Van deze programma's werd oorspronkelijk verwacht dat ze de vaardigheid van gebruikers vergroten (Kiefer & Smith 1983, Sterkel, Johnson & Sjogren 1986). Programma's uit zowel de derde als de tweede groep beogen een vermindering van de cognitieve belasting die tijdens het schrijven optreedt (zie bijvoorbeeld Zeller Mayer, Salomon, Globerson & Givon 1991).

Recentelijk is er een vierde groep COS-programma's ontwikkeld, veelal gebaseerd op sociaal-constructivistische en neo-Vygotskyaanse theorieën, die de sociale en dialogische aspecten van het schrijf(ler)proces benadrukken. Dergelijke programma's bieden de beginnende schrijver bijvoorbeeld via e-mail of internet een reëel publiek en stimuleren samenwerking (Harris & Wambeam 1996).

Of COS werkt is al jaren onderwerp van discussie (zie Clark 1983, Kozma 1991). Naar de effecten van COS-programma's zijn veel studies verricht, met steeds wisselende uit-

komsten. Omdat een meta-analyse van dit onderzoek ontbrak, hebben wij deze verricht. Onze analyse moest antwoord geven op de volgende vragen:

1. Hoe groot is het effect van COS vergeleken met niet-computerondersteunde schrijfinstructie?
2. Wat zijn de kenmerken van de onderzochte COS-programma's en leerlingen en welke eigenschappen hebben de experimenten en de wetenschappelijke publicaties die daarover verslag doen?
3. Wat is de relatie tussen de gevonden effecten van COS enerzijds en de kenmerken van de COS-programma's, leerlingen, experimenten en publicaties in kwestie?

*Identificatie van de bronnen en selectie van de studies*

Studies naar de effecten van COS zijn gezocht in de volgende digitale bestanden van wetenschappelijke literatuur: Current Contents, Educational Resources Information Center (ERIC), Education Library, Library and Information Science

effect	communication* technology	learning	composition
empirical	computer*	teaching	grammar
experiment	desktop		handwriting
research	electronic mail		punctuation
treatment	information processing		spelling
	information technology		writing
	internet		written
	word processing		
	world wide web		
	education*		
	instruction*		

Tabel 1: Profiel van zoektermen voor studies naar effecten (kolom 1) van computerondersteund (kolom 2) onderwijs (kolom 3) van schrijven (kolom 4). Een \* bij een zoekterm is een jokerteken dat staat voor iedere mogelijke continuering van de term.

Abstracts, Linguistics & Language Behavior Abstracts (LLBA), MLA (Modern Language Association) International Bibliography, PAIS International, Periodical Abstracts, PsycINFO (voorheen PsychLIT), Sociological Abstracts en Ulrich's International Periodicals). Daarbij is een zoekprofiel gebruikt waarin vier groepen zoektermen werden onderscheiden. Het profiel is in tabel 1 weergegeven. Alleen publicaties die gekenmerkt werden door tenminste één zoekterm uit ieder van de vier groepen rekenden we tot onze bronnen.

Met het zoekprofiel werden 867 bronnen geïdentificeerd. Veel bronnen bleken onbruikbaar, bijvoorbeeld omdat ze alleen een beschouwing over het nut van COS bevatten, in plaats van een verslag over een empirische studie naar effecten. In de meta-analyse zijn alleen studies opgenomen die verslag deden van echte experimenten (waarin individuele leerlingen via loting aan één van de condities of onderwijsvormen worden toegewezen) of quasi-experimenten (waarin niet individuele leerlingen aan de condities worden toegewezen, maar groepen leerlingen dan wel klassen). Er waren ook bronnen waarin sprake was van andere effecten dan die op schrijfvaardigheid. Die zijn uitgesloten: we hebben ons beperkt tot studies waarin een maat voor schrijfvaardigheid als afhankelijke variabele werd gebruikt.

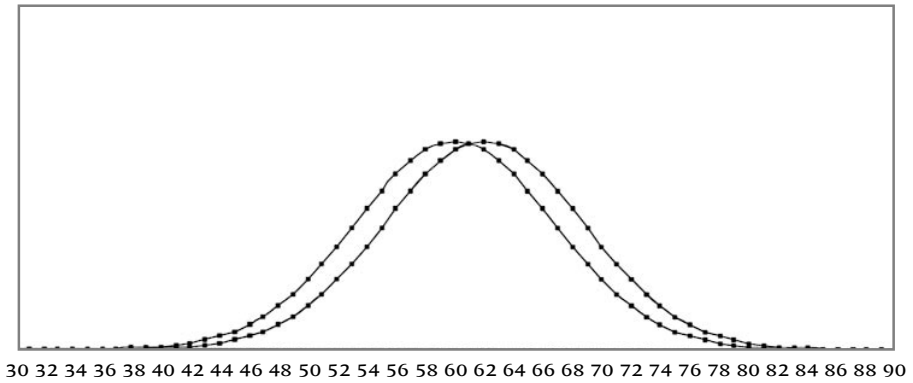
Er bleken 72 studies aan onze selectiecriteria te voldoen. Veel van deze studies rapporteerden over meer dan één vergelijking, bijvoorbeeld omdat twee experimentele vormen van COS met één controlegroep vergeleken werden of omdat er meer dan één maat voor schrijfvaardigheid als afhankelijke variabele werd gebruikt. De 72 studies bevatten in totaal 334 experimentele vergelijkingen. Deze vergelijkingen vormden het vertrekpunt voor onze meta-analyse.

De geanalyseerde studies zijn gepubliceerd tussen 1975 en 1999. Vóór 1975 werd geen onderzoek naar de effecten van COS uitgevoerd. Dat we studies na 1999 uitsluiten komt

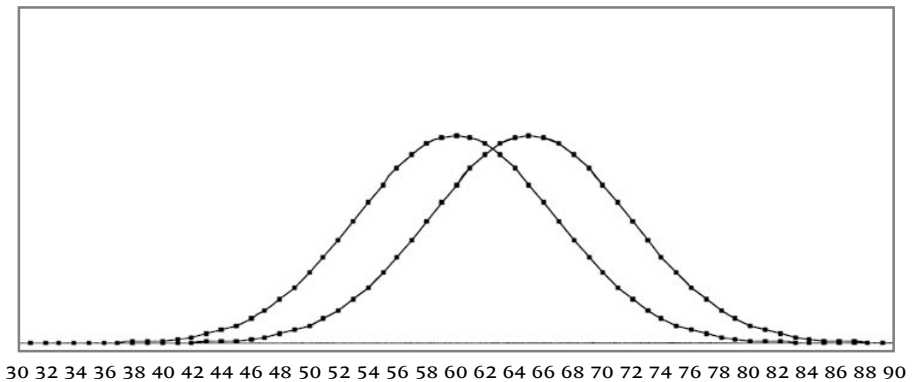
doordat we de collectie van experimenten in 2000 hebben aangelegd. Inmiddels zijn we bezig ons corpus te actualiseren, overigens zonder veel vrucht, omdat er na 1999 nauwelijks nieuwe studies zijn gepubliceerd. Het zijn er zo weinig dat ze nauwelijks invloed kunnen hebben op de resultaten en conclusies die we op het huidige corpus baseren.

#### *Berekening van de effectmaat en verdere analyses*

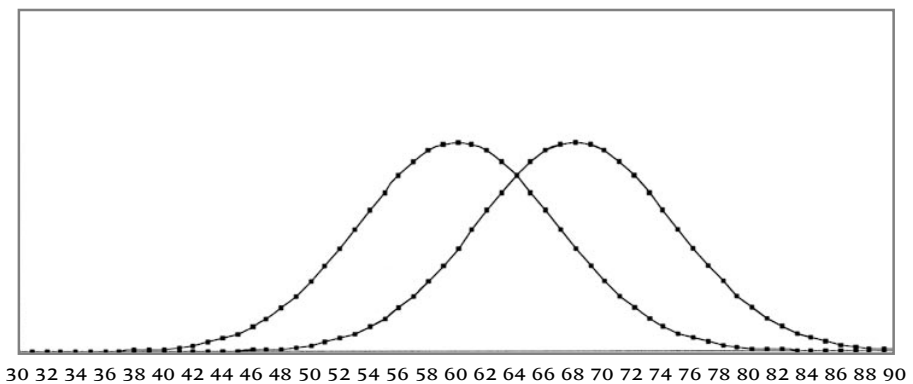
We hebben, zoals gebruikelijk, alle uitkomsten uit de experimentele vergelijkingen onder één noemer gebracht, dat wil zeggen uitgedrukt in één universele effectmaat. Wat moet de lezer zich bij die effectmaat voorstellen? De gebruikte maat, Hedges  $d$  (Hedges & Olkin 1985), drukt uit hoe de verschillen tussen vormen van onderwijs (+COS en -COS bijvoorbeeld) zich verhouden tot de verschillen binnen vormen van onderwijs. Enkele voorbeelden kunnen dit verduidelijken. Stel dat de uitkomsten van een COS-experiment laten zien dat er tussen twee vormen van onderwijs geen verschil in effect is, terwijl er binnen beide vormen wel verschil is tussen de effecten op individuele leerlingen. Hedges  $d$  is dan gelijk aan nul, want de verschillen tussen onderwijsvormen zijn nul en die binnen vormen zijn groter dan nul. Zijn er tussen onderwijsvormen wel verschillen, maar daarbinnen niet (dit laatste komt eigenlijk nooit voor), dan is Hedges  $d$  oneindig groot, want delen door nul geeft levert oneindig op. Normaal is het derde type situatie, waarin er zowel tussen als binnen onderwijsvormen verschillen zijn. De effectmaat is dan een getal dat de reële verhouding van tussen- en binnengroepsverschillen uitdrukt. Bij een  $d$  van 0.20 hebben we te maken met een klein effect; een  $d$  van .50 wijst op een middelmatig groot effect; bij een  $d$  van .80 dan noemen we het effect groot (Cohen 1992).



**Klein effect: gemiddelde (- COS) = 60; gemiddelde (+ COS) = 62**



**Middelmatig effect: gemiddelde (- COS) = 60; gemiddelde (+ COS) = 65**



**Groot effect: gemiddelde (- COS) = 60; gemiddelde (+ COS) = 68**

*Figuur 1: Kleine, middelmatige en grote effecten*

In figuur 1 worden deze drie situaties in beeld gebracht. Daarbij is de standaarddeviatie, de gebruikelijke maat voor de verschillen binnen groepen, steeds op 10 gesteld en het gemiddelde van de controlegroep (-COS) steeds op 60; het gemiddelde van +COS, de experimentele groep, staat op 62, 65 of 68.

Effectmaten konden berekend worden voor 278 vergelijkingen tussen een experimentele conditie (+COS) en een controleconditie (-COS). De overige studies met in totaal 56 vergelijkingen meldden alleen of de verschillen tussen de experimentele condities statistisch significant waren (dat wil zeggen niet aan toeval toe te schrijven).

Aan de hand van onze universele effectmaat konden we het gemiddelde effect van de COS-programma's schatten en ook de verschillen in effecten bestuderen die zich tussen de verschillende programma's voordoen. We zijn met name nagegaan of wisselingen in effecten samenhangen met kenmerken van de onderzochte COS-programma's en leerlingen, en met eigenschappen van de experimenten en de publicaties in kwestie. Daarbij hebben we gebruik gemaakt van multiniveau regressieanalyse (Bryk & Raudenbusch 1992, Hox 2002). Nadere details omtrent de verschillende analyses zijn te vinden in de appendix bij dit artikel.

#### *Het gemiddelde effect van COS en de wisselingen per experimentele vergelijking*

Het gemiddelde effect van COS bedraagt .46. Deze waarde kan, gezien het aantal vergelijkingen, niet aan toeval worden toegeschreven. Het gemiddelde effect is niet alleen statistisch significant, het is ook substantieel. COS heeft een middelmatig groot effect op de schrijfvaardigheid van leerlingen. Als we bedenken dat de duur van werken met de COS-programma's gemiddeld slechts 42 uur was (zie hieronder), dan is dit zeker een interessant effect te noemen.

Verder blijkt dat de experimentele vergelijkingen substantieel heterogeen zijn. Met andere woorden: de onderlinge verschillen tussen de uitkomsten van de vergelijkingen zijn zo groot dat zij niet aan het toeval kunnen worden toegeschreven.

#### *De kenmerken van de programma's, leerlingen, experimenten en publicaties*

Wat waren eigenlijk de kenmerken van de onderzochte COS-programma's en leerlingen en welke eigenschappen hadden de experimenten en de wetenschappelijke publicaties die daarover verslag doen? Een antwoord op deze vraag is nodig om het bovenvermelde effect op waarde te kunnen schatten. We presenteren daartoe gegevens die gemiddeld zijn over de 334 experimentele vergelijkingen. Voor nadere informatie over de wijze waarop de kenmerken zijn vastgesteld, verwijzen we de lezer naar de appendix bij dit artikel. We bespreken hieronder eerst de programma-kenmerken.

#### *Kenmerken van de COS-programma's*

Alle COS-programma's werden op school aangeboden. De instructie duurde gemiddeld 21 weken en nam in doorsnee 42 uur in beslag. In de vergelijkingen kwamen de volgende tekstgenres aan bod: expressief (26%), narratief (10%), descriptief (2%), argumentatief (21%), en beschouwend (24%). De programma's richtten zich op tekstinhoud (69%), tekstorganisatie (35%), formuleren (30%), vervoegingen (4%), spelling (58%), interpunctie (20%) en handschrift (4%). De programma's waren gericht op het tekstniveau (61%), het alineaniveau (39%), het zinsniveau (27%), het woordniveau (36%) of het foneemen- en grafeemniveau (39%). Publiekgerichtheid was de focus in 19% van de vergelijkingen.

De drie schrijffases kwamen alle aan bod: voorbereidende activiteiten (27%), het schrijven zelf (86%) en de revisie (31%). In slechts 13% van de gevallen kwamen de fases alledrie aan bod. Er werd bijna zonder uitzondering geschreven in een taal die op school als de instructietaal en in de samenleving als de meerderheidstaal gold (99%).

Wat betreft de didactiek is er sprake van het aanleren van regels (20%), van het aanbieden van voorbeeldmodellen (35%), van procedurele facilitatie (54%), of van training van strategieën (18%). Samenwerkend leren kwam slechts in 8% van de vergelijkingen voor. In 7% van de gevallen werden via internet of een encyclopedie inhoudsbronnen geboden. Linguïstische bronnen (woordenboek, thesaurus, spellingcontrole, grammaticaconrole) waren in 8% van de gevallen beschikbaar. Programma's leverden meestal instructie op het scherm (92%) en geregeld (ook) gesproken feedback (28%). Bij 8% was er via e-mail of internet sprake van echt publiek. Leerlingen konden meestal invloed uitoefenen op het tempo of de inhoud van de instructie (66%). Of er sprake was van adaptieve instructie, waarbij de antwoorden van de leerlingen bepalend zijn voor de volgende stap in instructie, konden we niet bepalen (zie appendix).

Het COS-programma werd in 37% van de gevallen niet alleen tijdens de instructiefase gebruikt, maar was ook tijdens de nameting als instrument of schrijfstun beschikbaar. Bij deze vergelijkingen zijn daarom het effect van de instructie met en het effect van het geïnstrueerde gebruik van de digitale schrijffomgeving niet te scheiden. Voor de goede orde, dit is niet alleen een louter methodologische kwestie. Het is immers in de eerste plaats een didactische vraag of het nodig en wenselijk is om een instrument na de instructiefase weg te halen.

### Kenmerken van de leerlingen

De gemiddelde leeftijd van de onderzochte leerlingen was 13 jaar. De proporties mannelijke en vrouwelijke proefpersonen waren respectievelijk .51 en .49. De leerlingen waren meestal moedertaalsprekers (86%); soms was sprake van gemengde groepen van eerste- en tweede-taalleerders (14%). De sociaal-economische status van de leerlingen was gemiddeld (75%) of laag (25%). Hun vaardigheidsniveau was meestal gemiddeld (66%), soms laag (27%) en zelden hoog (7%). Van leerproblemen was zelden sprake (6%). De typevaardigheid van de leerlingen was meestal gemiddeld (86%).

### Kenmerken van de experimenten

De gemiddelde steekproefgrootten voor experimentele groepen en controlegroepen bedroegen respectievelijk 119 en 101. De leerlingen kregen meestal instructie van de eigen leraar (84%) en soms van een onderzoeker (16%). De leerlingen in de experimentele condities kregen computerondersteund schrijfonderwijs (+COS) en maakten tijdens de toetsmomenten gebruik van een tekstverwerker (+TVW). De leerlingen in de controlecondities kregen meestal ook schrijfonderwijs en dat dan natuurlijk zonder computerondersteuning (-COS); daarbij hadden zij vrijwel nooit een tekstverwerker tot hun beschikking (-TVW). In een enkel geval kregen de leerlingen in de controleconditie geen instructie (-S). In tabel 2 zijn de verschillende condities weergegeven en de frequenties van de soorten vergelijkingen.

Classificatie van condities	Instructie met computer	Instructie zonder computer	Geen instructie
Schrijven met tekstverwerker zonder tekstverwerker	+CO	-COS+TVW -COS-TVW	-S+TVW -S-TVW
Frequenties van vergelijkingen	Vergelijking		Aantal
Experimenteel vs. controle	+COS	↔ -COS+TVW	71
	+COS	↔ -COS-TVW	260
	+COS	↔ -S+TVW	1
	+COS	↔ -S-TVW	1
Geen controlegroep, alleen COS	Aantal oefeningen v toename schrvdh		1

Tabel 2: Classificatie van experimentele conditie (+COS) en controlecondities (overige) en frequenties van vergelijkingen

De opzet van de studies was niet altijd even sluitend: de interne validiteit van de vergelijkingen werd door ons meestal als goed beoordeeld (78%), soms als middelmatig (13%), soms als laag (10%). Willekeurige toewijzing aan condities van individuele leerlingen vond bij slechts 32% van de vergelijkingen plaats.

De afhankelijke variabele waaraan het effect van COS werd afgemeten betrof meestal het schrijfproduct: de globale of holistische kwaliteit (25%), de hoeveelheid geschreven tekst (5%), de kwaliteit van gebruikte argumenten (2%), de grammaticale correctheid (11%), de spelling (30%) of het handschrift (1%). In andere gevallen ging het om onderdelen of aspecten van het schrijfproces: het metacognitief schrijfbewustzijn (4%), de hoeveelheid tijdens de voorbereiding geproduceerde ideeën (5%), de kwaliteit van de tijdens de voorbereiding gegenereerde ideeën (5%) of de schrijfsnelheid (2%).

Bij 20% van de vergelijkingen werd gecontroleerd voor onderwerpskennis door leerlingen over verschillende onderwerpen te laten schrijven. In 66% van de gevallen ontbrak een voormeting of pretest. Bij deze vergelijkingen kon daarom niet bepaald worden of er al op

de pretest sprake was van verschillen tussen de condities. Dat is jammer, want zulke verschillen kunnen de conclusies van experimenten ondermijnen.

In 88% van de gevallen waren de verschillende condities genest binnen docenten: men gaf slechts binnen één vorm van onderwijs les. In de overige 12% was sprake van volledige kruising: dezelfde leerkrachten gaven instructie in ieder van de condities. Aan beide opties zitten voor- en nadelen. Bij nesting van condities binnen leerkrachten kunnen de experimentele docenten en de controledocenten van elkaar verschillen in, bijvoorbeeld, belangstelling, motivatie, inzet, leeftijd of ervaring. Bij kruising zijn ook problemen denkbaar. Omdat docenten met beide onderwijsvormen te maken hebben kunnen zich interacties tussen de condities voordoen. En ook hier is niet gezegd dat ieders motivatie en inzet in beide condities hetzelfde zullen zijn.

#### Kenmerken van de wetenschappelijke publicaties

De opgenomen studies zijn gepubliceerd van 1975 tot 1999, maar een meerderheid is



verschenen tussen 1987 en 1993. Van de studies verscheen 38% in wetenschappelijke tijdschriften. De resterende 62% is zogeheten 'grijze' literatuur.

*Relaties tussen de grootte van de effecten en afzonderlijke kenmerken van de programma's, leerlingen, experimenten en publicaties*

Zoals gezegd bedraagt het gemiddelde effect van COS .46 en is er sprake van substantiële variatie: in sommige vergelijkingen zijn de effecten (veel) groter, in andere (veel) kleiner. Er is dus zowel een aanleiding als een basis voor een nadere analyse van de relaties tussen de grootte van de effecten en de kenmerken van de programma's, leerlingen, experimenten en publicaties. Wij hebben daarom voor ieder afzonderlijk kenmerk bepaald of er sprake is van een niet toevallige (in onderzoekstaal: statistisch significante) relatie met de effecten. Zulke kenmerken lijken belangrijk, al kunnen we niet bewijzen dat ze de verschillen in de effecten veroorzaken.

COS-programma's die gericht waren op het narratieve genre hebben minder positieve effecten dan de programma's die de andere genres betreffen. Dat geldt ook voor programma's die zich richtten op handschrift of schrijfsnelheid. De programma's die zich richten op het schrijven zelf, hebben kleinere effecten dan de programma's die bedoeld zijn voor voorbereidende activiteiten of revisie. Van belang lijkt verder of het COS-programma ook tijdens de nameting als instrument of schrijfstun beschikbaar is: in die vergelijkingen waarin het programma beschikbaar bleef zijn de effecten positiever.

Wat betreft de leerlingen: de effecten zijn groter wanneer zij ouder waren, wanneer hun vaardigheid groter was en wanneer zij beter konden typen.

Twee kenmerken van de experimenten doen ertoe. Hoe groter de verschillen op de

pretest ten gunste van de COS-groep waren, hoe positiever de effecten (zo'n 'valse' start komt nogal eens voor). Verder zijn de effecten positiever wanneer de leerkrachten en de condities volledig gekruist werden. We herinneren de lezer eraan dat dit kenmerk voor meer dan één uitleg vatbaar is. De publicatiekenmerken vertonen geen samenhang met de effecten.

Wat moet de lezer zich nu bij de gevonden relaties tussen de grootte van de effecten en de kenmerken van de programma's, leerlingen en experimenten voorstellen? In de onderstaande figuren proberen we hun interpretatie door illustratie te vergemakkelijken.

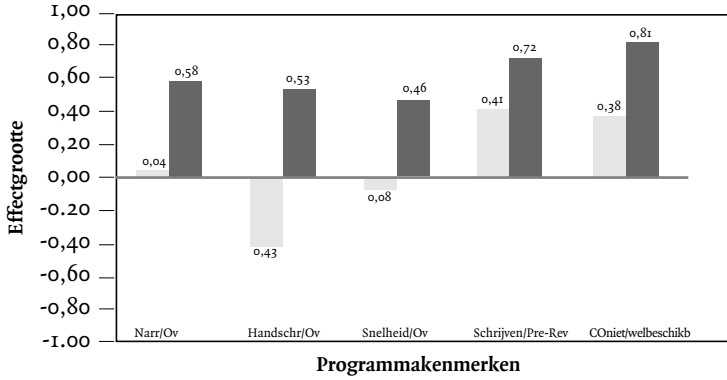
In figuur 2 is afgebeeld met welke effectgrootten de potentieel belangrijke programmakekenmerken gepaard gaan.

In figuur 3 is te zien hoe de effectgrootten met de potentieel belangrijke leerlingkenmerken samenhangen.

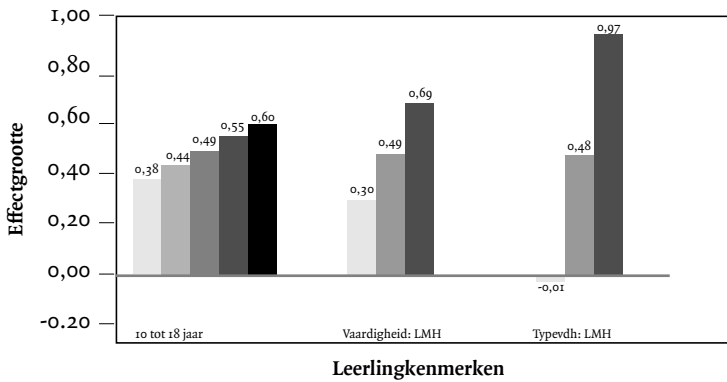
In figuur 4 tenslotte zijn de samenhangen tussen de effectgrootten en de potentieel belangrijke methodologische kenmerken weergegeven.

*Succes bij het verklaren van de heterogeniteit van de effecten*

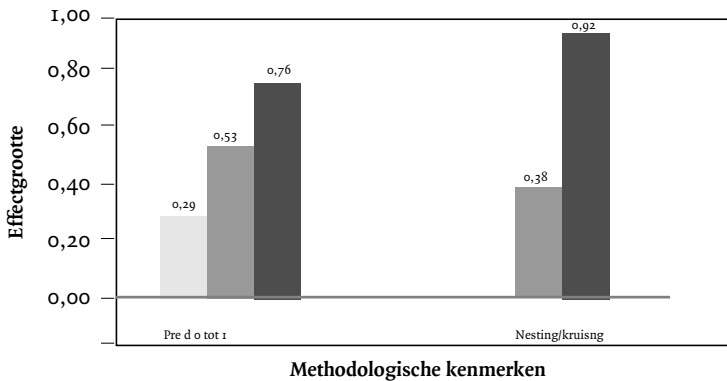
Een laatste belangrijke constatering betreft ons succes in het beschrijven of (statistisch) verklaren van de substantiële heterogeniteit in de effectgrootten. Indien we daarbij alle elf eerder opgevoerde kenmerken tegelijk gebruiken, kunnen we van slechts 14% van de verschillen in de effectgrootten rekenschap afleggen. We zijn er met andere woorden nauwelijks in geslaagd de effectieve COS-programma's van de minder of niet-effectieve COS-programma's te onderscheiden. Het blijft daarom goeddeels onduidelijk waardoor sommige COS-programma's wel en andere nauwelijks of geen positieve effecten op de schrijfvaardigheid laten zien.



Figuur 2: Gemiddelde effectgrootten bij potentieel belangrijke programmakenmerken



Figuur 3: Gemiddelde effectgrootten bij potentieel belangrijke leerlingkenmerken



Figuur 4: Gemiddelde effectgrootten bij potentieel belangrijke methodologische kenmerken

## Conclusie

Onze meta-analyse laat een middelmatig groot effect van computerondersteund schrijfonderwijs zien. De gemiddelde effectgrootte van COS is .46, wat nauw overeenkomt met de middelste grafiek in figuur 1. Dit gemiddelde effect is beslist niet verwaarloosbaar, zeker wanneer wordt bedacht dat het verbonden is aan 21 weken en 42 uur COS. Bij langduriger gebruik van COS zou het effect nog groter kunnen zijn. Dit blijkt overigens niet op voorhand uit onze meta-analyse. Daarin hebben we namelijk geen verband gevonden tussen de hoeveelheid computerondersteund onderwijs (in weken en uren) en de effectgrootte. Er is met andere woorden tussen studies of vergelijkingen geen samenhang tussen de duur en de effecten van COS. Maar dit zegt niet alles. Binnen studies kan daarvan wel degelijk sprake zijn. Wij hebben helaas maar één studie waaruit zo'n verband kan worden afgelezen, die op de onderste rij van tabel 2. Uit dit onderzoek (Stumhofer 1988) blijkt dat er een zwak verband was tussen het aantal COS-oefeningen en de toename van de schrijfvaardigheid: de correlatie bedroeg .20.

Een tweede boodschap betreft de heterogeniteit van de effecten van COS. De variatie in de effectgrootten is zeer fors. Er is sprake van veel spreiding, onder en boven het gemiddelde van .46. De range van effectgrootten loopt van  $-3.90$  tot  $+4.72$ . COS kan dus zeer effectief zijn, maar ook sterk negatief werken. Er is dus alle reden om te zoeken naar een nadere verklaring van de wisselende effecten. Dat hebben we gedaan en dat is tot op zekere hoogte gelukt. We hebben verbanden gevonden tussen de effectgrootten en afzonderlijke kenmerken van programma's, leerlingen en experimenten. We bespreken ze een voor een.

COS-programma's die gericht waren op het narratieve genre hebben minder positieve

effecten dan de programma's die de andere genres betreffen. Het is niet duidelijk hoe dit resultaat geïnterpreteerd moet worden. Past COS niet goed bij het narratieve genre? Dat valt op voorhand moeilijk in te zien. Waarom zou computerondersteuning minder goed passen bij het schrijven van verhalen dan bij het schrijven van bijvoorbeeld een betoog? Zowel bij het narratieve als het argumentatieve genre is het voorstelbaar dat de computer hints geeft die helpen bij het generen van inhoud of het aanbrengen van structuur. Was de kwaliteit van de computerondersteuning bij het schrijven van verhalen dan niet aan de maat? Of werden de verhalen met name beoordeeld op creatieve kenmerken die niet voor COS vatbaar waren? Zonder nadere analyse van de gebruikte software en verder onderzoek kunnen we alleen speculeren over een verklaring van dit genre-effect.

Minder positieve kenmerken vinden we ook bij programma's die zich richten op het handschrift of de schrijfsnelheid. Deze negatieve effecten kunnen misschien als volgt verklaard worden. Bij het gebruik van COS die gericht is op het verbeteren van het handschrift wordt het schrijven met de hand misschien teveel verdrongen of onderdrukt. De schrijfsnelheid zou te lijden kunnen hebben onder het feit dat een computeromgeving al snel complex is en meer aandacht en controle voor zichzelf vraagt dan pen en papier.

Dat programma's die zich richten op het schrijven zelf, kleinere effecten hebben dan programma's die bedoeld zijn voor voorbereidende activiteiten of revisie valt goed te begrijpen. Ondersteuning van het 'voorwerk' lijkt ook eenvoudiger in software in te bouwen dan ondersteuning van het formuleerproces. Voor de leerlingen kan vooraf plannen van de inhoud of de structuur met computerondersteuning niet alleen eenvoudiger, maar ook productiever worden. Aantekeningen, schema's en aanzetten voor tekst kunnen door kopiëren en plakken heel gemakkelijk in

de te schrijven tekst gebruikt worden. Wat voor de voorbereiding op het schrijven geldt, lijkt ook op te gaan voor de revisie. Vragen die tot reflectie aanzetten of checklists voor revisie kunnen binnen COS gemakkelijk aangeboden en gebruikt worden.

Voor de effecten van COS lijkt het van belang dat de programmatuur ook tijdens de nameting als instrument of schrijfsteun beschikbaar is. Dit resultaat is niet verrassend. De teksten die leerlingen via COS schrijven worden weliswaar beter, maar de leerlingen zijn, worden of blijven daarbij wel afhankelijk van de geboden technologie. Of dat onwenselijk is, is nog maar de vraag. Is het wel nodig om de computerondersteuning na de instructiefase weg te halen? Tijdelijke hulpmiddelen die leerlingen achter zich kunnen en moeten laten, horen in een nameting vanzelfsprekend niet toegankelijk te zijn. Maar wanneer het gaat om cognitieve tools die ook na het leerproces (en ook buiten school) van nut kunnen zijn, is het antwoord minder eenduidig. In dit verband kan gedacht worden aan programmatuur voor spellingcontrole. Deze werd ooit als didactische hulpmiddel geïntroduceerd, maar is inmiddels een vast bestanddeel van software voor tekstverwerking. Over de betekenis en waarde van effecten met en door technologie is overigens ook in de wetenschappelijke literatuur het laatste woord nog niet gezegd (zie Jonassen & Reeves 1996, Salomon, Perkins & Globerson 1991).

Wat betreft de leerlingen is gebleken dat de effecten van COS groter zijn wanneer zij ouder waren, wanneer hun vaardigheid groter was en wanneer zij beter konden typen. Het laatste van deze drie effecten is waarschijnlijk tijdgebonden. Misschien valt het in veel groepen leerlingen nu of binnenkort al niet meer te constateren. Het gebruik van computers is immers vaak al zo vroeg zo gewoon en zo frequent dat er nauwelijks relevante verschillen in typevaardigheid overblijven. De eerste twee

effecten zijn heel gebruikelijk. Het is in het algemeen zo dat oudere en meer vaardige leerlingen het meest van instructie profiteren. Zij leren sneller, door meer ervaring, grotere kennis, betere strategieën of een grotere intelligentie.

Behalve kenmerken van de programma's en de leerlingen deden twee kenmerken van de experimenten er toe. Het effect van het eerste methodologische kenmerk valt direct te begrijpen. Hoe groter de verschillen op de pretest ten gunste van de COS-groep waren, hoe positiever de effecten. Het is heel gewoon dat leerlingen die hoger scoren op een pretest meer profijt trekken van de geboden instructie. Dat kan komen door een cumulatieve werking van hun voorsprong of doordat zij sowieso sneller leren.

Dat effecten van COS positiever zijn als binnen de experimenten de leerkrachten en de condities volledig gekruist werden, was niet op voorhand duidelijk. Het verschil in effect tussen nesting en kruising is aanzienlijk, zoals figuur 4 laat zien. Wellicht is er binnen docenten toch sprake geweest van verschil in motivatie en inzet voor de beide typen condities. Het lijkt erop dat de docenten in de COS-conditie al dan niet bewust meer hun best hebben gedaan dan in de controleconditie. Het kan ook zo zijn dat de leerlingen in de experimentele conditie enthousiaster waren en harder hebben gewerkt dan hun schoolgenoten in de controleconditie.

#### Verder met COS?

De verbanden tussen de effectgrootten en de hierboven besproken kenmerken zijn de best beschikbare aanwijzingen die we op dit ogenblik hebben voor de verklaring van de heterogeniteit in de effecten van COS. Toch moeten we hun belang relativeren. Wij hebben de variatie in effectgrootten nog maar slecht begrepen. We kunnen uiteindelijk niet meer

dan 14% van de verschillen verklaren. Voor een beter begrip van de effecten van COS is meer en ander onderzoek nodig. Het lijkt ons vooral van belang om na te gaan hoe leerlingen tijdens het leren, oefenen en (later) schrijven gebruik maken van de computeromgeving. Door nauwkeurige bestudering van hun schrijfleerprocessen kan wellicht beter begrepen worden wanneer COS goed werkt, voor wie en waarom.

Voor het schrijfonderwijs is de boodschap van onze meta-analyse de volgende. Gezien de hier gepresenteerde resultaten doen COS-programma's het vaak beter dan vormen van

schrijfinstructie zonder computerondersteuning. Er is daarom alle reden om de gebruiksmogelijkheden van COS verder te verkennen en om effectief gebleken software voor gebruik in het Nederlandse onderwijs toegankelijk te maken.

#### NOOT

1. Deze analyse is uitgevoerd in het kader van een onderzoeksproject dat gesubsidieerd is door de Programmaraad voor Onderwijs-onderzoek (PROO) van NWO.

#### Appendix

Bij het berekenen van de gemiddelde effectgrootte is a priori gekozen voor een *random effects* model (Hedges & Olkin 1985), vanwege het uiteenlopende karakter van de studies. Onderscheiden werden het studie- en het vergelijkingsniveau. Zo werd de afhankelijkheid van meerdere uitkomsten afkomstig uit één studie statistisch verdisconteerd. In vervolganalyses zijn door middel van multiniveau regressieanalyses (Bryk & Raudenbusch 1992, Hox 2002) verbanden onderzocht tussen de effecten en de kenmerken van de onderwijsprogramma's, leerlingen, experimenten en publicaties in kwestie. Deze analyses zijn verricht met MLwin (Rasbash e.a. 2000). De significantie van de residuele variantie is bepaald met de  $\chi^2$ -test (Bryk & Raudenbusch 1992).

Het gemiddelde effect, na verdiscontering van verschillen in steekproefgrootte, komt uit op een statistisch significante *d*-waarde van .458 ( $se=.084$ ). De homogeniteitanalyse laat zien dat de vergelijkingen die in het model zijn opgenomen, heterogeen zijn ( $\chi^2 = 3542$ ;  $df = 275$ ;  $p < .000$ ).

Om te bepalen of de coderingen voldoende betrouwbaar zijn uitgevoerd, is een deelverzameling van 22 studies door een tweede beoordelaar gecodeerd. De percentages overeenstemming lopen van 53 tot 100. Ook zijn Cohen's Kappa (voor nominale variabelen) en Spearman's *r* (voor

ordinaire) of *r<sub>r</sub>* (voor continue variabelen) berekend (Orwin 1994). Slechts twee van de 67 percentages overeenstemming komen onder de ondergrens van 60%. Eén van deze variabelen (het aantal uren instructie) blijkt echter een *r<sub>r</sub>* van .94 te hebben. Blijkbaar ronden beide beoordelaars verschillend maar wel consistent af. De Kappa van de andere variabele, 'adaptieve instructie', is met .16 te laag. Deze variabele is (als enige) uit de analyses verwijderd. Door missende waarden op sommige variabelen zijn de steekproefgrootten niet bij alle analyses gelijk. Wij zijn nagegaan of onze meta-analyse te lijden heeft onder vormen van bias of vertekening. Er zijn geen significante verbanden gevonden tussen de aard van de publicatiebron (wetenschappelijk tijdschrift of andere bron) en de studie-uitkomst. Dit resultaat wijst dus niet op publicatiebias. Verder is nagetrokken of de uitkomsten van studies die alleen melden of resultaten wel of niet 'significant' zijn, afwijken van studies die wel preciezere statistische gegevens verschaffen (Bushman 1994). De resultaten van de 278 vergelijkingen die zijn opgenomen in de multiniveau regressieanalyses wijken niet significant af van die van de 56 vergelijkingen die onvoldoende statistische informatie gaven om Hedges *d* te berekenen, zoals blijkt uit een toets op de proporties significante en niet-significante vergelijkingen ( $\chi^2 (df=2) = 3.43$ ,  $p = .18$ ).

## LITERATUUR

- Bangert-Drowns, R.L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research*, 63, 1, 69-93.
- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.
- Bryk, A.S., & Raudenbusch, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.
- Bushman, B.J. (1994). Vote-counting procedures in meta-analysis. In H. Cooper & L.V. Hedges (eds), *The handbook of research synthesis* (pp. 193-213). New York: Russell Sage Foundation.
- Butterfield, E.C., Hacker, D.J., & Albertson, L.R. (1996). Environmental, cognitive, and metacognitive influences on text revision: Assessing the evidence. *Educational Psychology Review*, 8, 3, 239-297.
- Chanquoy, L., & Alamargot, D. (2002). Working memory and writing: Evolution of models and assessment of research. *Année Psychologique*, 102, 2, 363-398.
- Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53, 4, 445-459.
- Cochran-Smith, M. (1991). Word processing and writing in elementary classrooms: A critical review of related literature. *Review of Educational Research*, 61, 1, 107-155.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Daiute, C. (1986). Physical and cognitive factors in revision: Insights from studies with computers. *Research in the Teaching of English*, 20, 2, 141-159.
- Emig, J. (1971). *The composition process of twelfth graders*. Urbana, Ill.: NCTE.
- Flower, L.S., & Hayes, J.R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32, 365-387.
- Harris, L.D., & Wambeam, C.A. (1996). The internet-based composition classroom: A study in pedagogy. *Computers and Composition*, 13, 3, 353-371.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, Florida: Academic Press Inc.
- Hillocks, G. (1986). *Research on written composition*. Urbana, IL: National Council of Teachers of English.
- Hox, J.J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Jonassen, D.H., & Reeves, T.C. (1996). Learning with technology: Using computers as cognitive tools. In D.H. Jonassen (ed.), *Handbook of research for educational communications and technology* (pp. 693-719). New York: Simon & Schuster Macmillan.
- Kellogg, R.T. (1994). *The psychology of writing*. New York: Oxford University Press.
- Kiefer, K.E., & Smith, C.R. (1983). Textual analysis with computers: Tests of Bell Laboratories' computer software. *Research in the Teaching of English*, 17, 3, 201-214.
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research* 61, 2, 179-211.
- Leahy, P. (1991). A multi-year formative evaluation of IBM's 'Writing to Read' program. *Reading Improvement*, 28, 4, 257-264.
- Levy, C.M., & Ransdell, S. (1996). *The science of writing: Theories, methods, individual differences and applications*. Mahwah, NJ: Lawrence Erlbaum.
- MacArthur, C.A., Ferretti, R.P., Okolo, C.M., & Cavelier, A.R. (2001). Technology applications for students with literacy problems: A critical review. *The Elementary School Journal*, 101, 3, 273-301.
- McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, 8, 3, 299-325.
- Orwin, R.G. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (eds), *The handbook of research synthesis* (pp. 139-162). New York: Russell Sage Foundation.

Rasbash, J. et al. (2000). *A user's guide to MLwiN (version 2.1)*. London: University of London.

Reynolds, T.H., & Bonk, C.J. (1996). Computerized prompting partners and key-stroke recording devices: Two macro driven writing tools. *Educational Technology Research and Development*, 44, 3, 83-97.

Salomon, G., Perkins, D.N., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20, 3, 2-9.

Sterkel, K.S., Johnson, M.I., & Sjogren, D. (1986). Textual analysis with computers to improve the writing skills of business communication students. *Journal of Business Communication*, 23, 1, 43-61.

Stumhofer, N.C. (1988). *The impact of computer-assisted instruction on students' knowledge of basic writing skills*. ERIC Document Service (ED 293 600).

Wesdorp, H. (1983). *Schrijven in het voortgezet onderwijs*. Harlingen: Flevodruk.

Williams, J.P. (2001). Commentary: Four meta-analyses and some general observations. *The Elementary School Journal*, 101, 3, 349-354.

Woodruff, E., Bereiter, C., & Scardamalia, M. (1981). On the road to computer assisted compositions. *Journal of Educational technology Systems*, 10, 133-148.

Zellermayer, M., Salomon, G., Globerson, T., & Givon, H. (1991). Enhancing writing-related metacognitions from a computerized writing-partner. *American Educational Research Journal*, 28, 373-391.