

boekbespreking

B. Meuffels LEERLINGEN BEOORDELEN ELKAARS OPSTELLEN

Doel van het onderzoek

In november van het afgelopen jaar promoveerde Gert Rijlaarsdam op een proefschrift, getiteld *Effecten van leerlingenrespons op aspecten van stelvaardigheid*.

Rijlaarsdam is als wetenschappelijk onderzoeker verbonden aan de SCO (Stichting Centrum voor Onderwijsonderzoek van de Universiteit van Amsterdam), een instituut dat de afgelopen jaren een zekere faam en reputatie heeft opgebouwd op het gebied van praktisch relevant onderwijsonderzoek. Ook het empirisch onderzoek waarop Rijlaarsdam promoveerde, beoogt praktische inzichten op te leveren die voor moedertaalleraren relevant zijn. In een empirisch effect-onderzoek gaat Rijlaarsdam na of leerlingen, wanneer ze commentaar moeten leveren op elkaars opstellen, beter gaan schrijven. Zelfs als die leerlingen niet beter kunnen schrijven dan leerlingen die via een 'traditionele' didactiek zijn onderwezen waarin de docent de opstellen becommentarieert, dan nog kan aan dat resultaat een zeker praktisch belang niet onzegd worden. De werkdruk van de Neerlandicus — het is genoegzaam bekend — is als gevolg van het vele correctiewerk in de regel vrij hoog. Als het toch niet uitmaakt

of leerkrachten dan wel leerlingen opstellen beoordelen, wat ligt er dan meer voor de hand dan dat men de leerlingen elkaars opstellen laat becommentariëren?

Het idee dat het inschakelen van leerlingen als beoordelaars van elkaars teksten een positieve invloed uitoefent op de stelvaardigheid, is niet nieuw. Omstreeks 1970 werd deze zogenaamde *peer-evaluation* door met name Amerikaanse onderzoekers gepropageerd als een veelbelovende didactiek voor het stellen. In de loop der jaren is er heel wat empirisch onderzoek naar de effecten van peer-evaluation verricht. Rijlaarsdam toont in zijn dissertatie overtuigend aan dat al dat onderzoek laboreert aan een aantal methodologische feilen die ondubbelzinnige conclusies over de effectiviteit van peer-evaluation uitsluiten. In een grootscheeps onderzoek waaraan acht scholengemeenschappen meewerkten (met in totaal 22 klassen, 11 docenten en 264 leerlingen van de derde klas van het havo en vwo) probeert Rijlaarsdam dergelijke manco's te voorkomen en dat maakt dat zijn studie zeker niet kan worden afgedaan als een loutere replicatie, een loutere herhaling van eerder verricht onderzoek. Bovendien tracht hij een theoretische bijdrage te leveren aan de fundering van deze steldidactiek.

Theoretische achtergrond

Waarom zou peer-evaluation positieve effecten op de stelvaardigheid van leerlingen hebben? Volgens Rijlaarsdam zijn eventuele positieve effecten te verklaren uit het feit dat leerlingen zoveel teksten van elkaar moeten lezen. Het gevolg daarvan is onder meer, dat ze gevoeliger worden voor fouten, ook die welke voorkomen in hun eigen teksten. Deze grotere gevoeligheid resulteert in een intensiever en zorgvuldiger corrigeren en herschrijven van de eigen tekst. Bovendien lezen de leerlingen teksten, geschreven aan de hand van een opdracht die ze ook zelf hebben ontvangen. Doordat de leerlingen die teksten over dezelfde opdracht onderling vergelijken, kunnen ze abstracte modellen construeren van 'de ideale tekst'.

Een positief effect van peer-evaluation zou ook verklaard kunnen worden uit het feit dat leerlingen *commentaar leveren* op elkaars schrijfproducten. Zodoende leren ze impliciet dan wel expliciet criteria hanteren voor de ideale tekst. De kennis van deze criteria zal ook ge-

mobiliseerd worden als zij zelf een tekst moeten schrijven.

Verder zou een positief effect van peer-evaluation op de stelvaardigheid verklaard kunnen worden uit de specifieke aard van het ontvragen commentaar: meer dan één persoon beoordeelt een schrijfprodukt; bovendien is die beoordelende persoon een mede-leerling die relatief snel feedback geeft. Krachtens deze specifieke kenmerken van het commentaar kennen de leerlingen dat commentaar meer betekenis toe, zodat ze met de gegeven feedback meer gaan doen. Herschrijven wordt dan ook niet meer als een strafmaatregel beschouwd, maar als een uitdaging om te voldoen aan de commentaren.

Hoe belangwekkend deze, summier weergegeven, theoretische overwegingen ook mogen zijn, de in praktisch opzicht meest klemmende vraag blijft toch: gaan leerlingen nu echt beter schrijven als niet hun docent, maar zichzelf de opstellen beoordelen?

Resultaten

Rijlaarsdam verwacht dat [1] leerlingen onderwezen via de peer-evaluation didactiek, betere opstellen schrijven dan leerlingen die van de leerkracht feedback hebben ontvangen (beter, voor zover dat betrekking heeft op doel- en publiekgerichtheid, stijl en opbouw van een tekst), [2] deze leerlingen meer vertrouwen krijgen in eigen kunnen, meer plezier in het schrijven, en minder angst om negatief beoordeeld te worden op de schrijftaak, [3] deze leerlingen hun schrijfprocessen ten positieve zullen veranderen (bijvoorbeeld: tijdens het schrijven zullen ze meer alternatieven genereren op het gebied van de inhoud, de presentatiestructuur en de formuleringen).

Anders dan veel empirisch-onderwijskundig onderzoek zijn de resultaten van de studie van Rijlaarsdam verrassend ondubbelzinnig. Afgezien van enkele veranderingen in schrijfprocessen kon geen van de verwachte effecten worden aangetoond. Al met al zijn deze resultaten enigszins teleurstellend.

Rijlaarsdam heeft welbewust — en mijns inziens terecht — gekozen voor een *empirische* aanpak om de effecten van peer-evaluation te beproeven. Mededelingen van docenten over de effectiviteit van peer-evaluation acht hij geen doorslaggevend bewijs voor de kwaliteit van peer-evaluation. Overigens is Rijlaarsdam

realist genoeg om niet de illusie te koesteren dat empirisch-onderwijskundig onderzoek tot 'fool-proof' didactieken leidt.

Alternatieve verklaringen

Het is niet eenvoudig om empirisch-onderwijskundig onderzoek op deugdelijke wijze op te zetten en uit te voeren, zeker wanneer dat onderzoek ook nog praktisch relevante uitkomsten moet opleveren. Tal van storende variabelen kunnen gewenste leereffecten vertroebelen; alternatieve verklaringen van de gevonden onderzoeksuitkomsten liggen in de regel voor het oprapen. Kunnen de negatieve onderzoeksuitkomsten van Rijlaarsdam misschien alternatief verklaard worden? Heeft hij wellicht belangrijke storende invloeden over het hoofd gezien?

In dit verband is het van belang op te merken dat Rijlaarsdam zijn onderzoek op voorbeeldige wijze heeft opgezet en uitgevoerd, waarbij vrijwel niets aan het toeval is overgelaten. Dat hij geen positieve effecten van peer-evaluation aangetroffen heeft,¹ kan dus niet (alternatief) verklaard worden uit bijvoorbeeld een gebrekke en rommelige onderzoeksopzet of uit de onbetrouwbaarheid van meetinstrumenten. Twee korte voorbeelden mogen voldoende zijn om deze stelling te illustreren.

Eén van de meest netelige problemen bij effect-onderzoek van stelvaardigheid betreft de omstandigheid dat die stelvaardigheid op betrouwbare wijze gemeten moet worden. Het is algemeen bekend dat opstelbeoordelaars het vaak niet met elkaar eens zijn over de kwaliteit van de opstellen en — sterker — dat ze het vaak ook niet met zichzelf eens zijn wanneer ze dezelfde serie opstellen na enkele weken opnieuw moeten beoordelen. In zulke gevallen schort het aan de betrouwbaarheid van de beoordeling, en in die gevallen kan ook nooit een effect van peer-evaluation gevonden worden. Via een apart vooronderzoek is Rijlaarsdam nagegaan of opstelbeoordeling aan de hand van zogenaamde opstelschalen betrouwbaar is. Opstelschalen zijn in kwaliteit oplopende voorbeeldopstellen (bijvoorbeeld 7), die elk voorzien zijn van een cijfer (bijvoorbeeld 4, 5, 6, 6½, 7, 8 en 9) en die de beoordelaar als referentiepunt dient te gebruiken bij het toekennen van een cijfer. Blijkens de resultaten levert deze beoordelingsprocedure een extreem hoge betrouwbaarheid op, vergeleken

althans met conventionele beoordelingsprocedures. Dat Rijlaarsdam geen effecten van peer-evaluation gevonden heeft, kan dus niet alternatief verklaard worden uit de onbetrouwbaarheid van de opstelbeoordeling.

Een ander probleem waarmee je als effect-onderzoeker soms geconfronteerd wordt is dat de leerkrachten, die een nieuw ontwikkelde didactiek (in dit geval: peer-evaluation) moeten geven, zich vaak niet of in onvoldoende mate aan de instructies van de onderzoeker houden. Staat er in de instructie dat er met de cursus 6 maal geoefend moet worden, dan oefent de een slechts 5 maal, de ander 3 maal, enzovoort. Als de leerkrachten zich niet houden aan de instructies, dan kan er überhaupt geen effect aangetoond worden. Via observaties, leerlingenvragenlijsten en docentenlogboeken kon Rijlaarsdam concluderen, dat de 'overtredingen' tegen het voorgeschreven lessenplan minimaal waren. Opnieuw sluit dit gegeven een mogelijk alternatieve verklaring van de gevonden onderzoeksuitkomsten uit.

Hoe kunnen die negatieve onderzoeksuitkomsten dan verklaard worden? Of moeten we wellicht voetstoots aannemen dat peer-evaluation, vergeleken met traditionele steldidactieken, niet tot positievere effecten leidt?

Tussen Scylla en Charibdis

Ik heb toch wel enkele bedenkingen tegen de onderzoeksopzet die Rijlaarsdam gekozen heeft. Juist die onderzoeksopzet maakt dat hij nauwelijks positieve effecten van peer-evaluation kan aantreffen, ook al zouden ze er 'in werkelijkheid' zijn.

In elk effect-onderzoek moeten de resultaten van een experimentele groep vergeleken worden met die van een controle-groep. Zou je bijvoorbeeld willen weten of een pil tegen hoofdpijn effectief is, dan kun je niet volstaan met een onderzoeksontwerp waarin aan een aantal hoofdpijnlijders de pil wordt toegediend en vervolgens, na enkele weken, wordt nagegaan of die personen minder last hebben van hun kwaal. Zo'n ontwerp sluit een aantal alternatieve, rivaliserende verklaringen van een positief resultaat niet uit. Niet de medicijn per se, maar de speciale aandacht en zorg van de arts die het betreffende medicijn verstrekke, zou de afname in hoofdpijn teweeg gebracht kunnen hebben. Om zulke, vaak triviaal aandoende alternatieve verklaringen uit te kunnen sluiten

dient zo'n onderzoeksontwerp naast een experimentele groep een controle-groep te bevatten, i.c. een groep bestaande uit patiënten die alleen een placebo (een 'neppil') ontvangen. Blijkt nu dat de experimentele groep minder last heeft van hoofdpijn dan de controle-groep, dan kan die winst in ieder geval niet meer toegeschreven worden aan de invloed van de arts.

Ook in onderwijskundig onderzoek dient, naast een experimentele groep, een controle-groep op te treden teneinde alternatieve verklaringen uit te sluiten (bijvoorbeeld het feit dat leerlingen ouder, 'rijper' worden en als gevolg daarvan beter gaan schrijven).

Rijlaarsdam heeft naar mijn oordeel geen erg gelukkige keuze gedaan wat betreft de vorm en inrichting van de experimentele en de controle-groep. Beide groepen volgen precies dezelfde, door Rijlaarsdam ontwikkelde steldidactiek, die zonder meer intensief genoemd kan worden (sommige bij het onderzoek betrokken docenten vreesden zelfs hun gewone programma voor het derde leerjaar niet af te kunnen krijgen, zoveel lesuren slokte de steldidactiek op). Het enige aspect waarin de twee groepen verschilden betrof de aard van de feedback: in de experimentele groep (de peer-evaluation groep) lazen de leerlingen elkaars opstellen, leverden er aan de hand van standaard-reactieformulieren schriftelijk commentaar op, en ontvingen schriftelijk commentaar van drie leerlingen — in de controle-groep ontvingen de leerlingen eveneens schriftelijk commentaar, op precies dezelfde standaard-reactieformulieren, nu echter ingevuld door de docent. Met zulke uiterst minieme verschillen tussen beide condities zal men niet licht forse verschillen in schrijfvaardigheid aantreffen en dus ook niet positieve effecten van peer-evaluation.

Ook in een heel ander opzicht acht ik de manipulatie van de experimentele en de controle-conditie niet geheel geslaagd, althans ongelukkig gekozen. De generaliseerbaarheid van de onderzoeksuitkomsten van Rijlaarsdams studie naar de schoolse praktijk wordt sterk gemitigeerd door de omstandigheid dat in de peer-evaluation groep telkens drie leerlingen schriftelijk (en anoniem) commentaar leverden op hetzelfde opstel, en dat onafhankelijk van elkaar. Mondeling overleg tussen deze drie beoordelaars of tussen de drie beoordelende leerlingen en de schrijver-leerling vond niet plaats.

Het komt mij voor dat peer-evaluation als steldidactiek juist een inspirerende en motiverende invloed op de leerlingen kan uitoefenen krachtens het feit dat er in de klas in kleine groepjes gezamenlijk met de schrijver-leerling gesproken en gediscussieerd wordt.

Nu zou men kunnen tegenwerpen dat Rijlaarsdam juist voor deze stringente opzet gekozen heeft teneinde de *interne validiteit* van zijn onderzoek te optimaliseren, zodat alternatieve verklaringen van een positief effect radicaal uitgesloten kunnen worden en dus precies is na te gaan welke factor verantwoordelijk is voor dit positieve effect. Maar ook in de door Rijlaarsdam gekozen opzet, hoe stringent deze ook is, kunnen we niet precies aangeven welke factor debet is aan een (eventueel) positief effect van peer-evaluation: de hoeveelheid feedback (drie commentaren in plaats van slechts één), de aard van de feedback (commentaar van een mede-leerling versus commentaar van een docent) of het lezen van elkaars opstellen. Toegegeven, elke onderwijskundige onderzoeker die met zijn onderzoek praktisch relevante resultaten wil bereiken, moet als het ware laveren tussen de Scylla van de interne validiteit en de Charibdis van de *externe validiteit* (i.c. de generaliseerbaarheid van de onderzoeksuitkomsten naar de alledaagse, schoolse situatie). Hoe sterker het accent op de optimalisering van de interne validiteit komt te liggen, hoe strakker in de regel de onderzoeksopzet, maar ook: hoe moeilijker de resultaten te generaliseren zijn.

De deskundigheid van de Neerlandicus als opstelbeoordelaar

In bepaalde wetenschappelijke kringen is het tegenwoordig *bon ton* de terzake kundigheid van de Neerlandicus als opstelbeoordelaar te discrediteren. In zekere zin kan men Rijlaarsdam's studie — ook al heeft Rijlaarsdam allerm minst die bedoeling — opvatten als een frontale aanval op de competentie van de Neerlandicus: het maakt immers niet uit, voor zover het om de verbetering van de schrijfvaardigheid gaat, of men leerlingen de opstellen laat beoordelen dan wel een specifiek daartoe opgeleide docent. Leerlingen zijn kennelijk even deskundig als docenten in het beoordelen van opstellen.

Twee relativerende opmerkingen zijn hier op hun plaats. Zowel de leerlingen als de docen-

ten die aan het onderzoek van Rijlaarsdam meewerkten konden uitsluitend schriftelijk commentaar leveren (aan de hand van 5-puntsschalen) op de volgende zes aspecten van een opstel: doelgerichtheid, publiekgerichtheid, structuur, betrouwbaarheid, nieuwswaarde en taalgebruik. Op alle andere aspecten van een opstel die mede de kwaliteit bepalen, kon niet worden gereageerd. Dat betekent dat een docent die peer-evaluation in de praktijk wil toepassen, nog steeds de opstellen van de 14- à 15-jarige leerlingen zal moeten corrigeren op vaak voorkomende fouten (spel- en interpunctiefouten, drogredenen, grammaticale onregelmatigheden, discrepanties tussen bedoelde en uitgedrukte betekenis, enzovoort). Ook in de peer-evaluation didactiek blijft de docent als specifiek deskundige onmisbaar. Men kan en mag immers niet van elke 14- à 15-jarige leerling van de derde klas van het havo/vwo verwachten dat hij bijvoorbeeld de spellingregels tot in de puntjes beheerst.

Een tweede argument tegen de stelling dat de studie van Rijlaarsdam een desavouering van de Neerlandicus als opstelbeoordelaar behelst is het volgende. De onderzoeksuitkomsten (geen verschil in effectiviteit tussen de leerlingen-feedback en docenten-feedback) kunnen immers heel goed alternatief verklaard worden uit de deskundigheid van de Neerlandicus. Het is zeer wel mogelijk dat de docenten in de controle-groep kwalitatief beter commentaar leverden dan de leerlingen in de experimentele groep. De leerlingen, onderwezen met de peer-evaluation didactiek, kunnen geleerd hebben van het commentaar geven, maar deze leerwinst ten opzichte van de leerlingen in de controle-groep zou teniet kunnen worden gedaan doordat leerlingen slecht en docenten goed commentaar op een opstel leveren. Rijlaarsdam heeft via vragenlijsten de leerlingen in beide condities gevraagd naar hun tevredenheid over de feedback. Er bleken geen verschillen tussen beide condities. Rijlaarsdam merkt echter op — naar mijn oordeel zeer terecht — dat gepercipieerde kwaliteit van ontvangen commentaar niet hetzelfde is als objectieve kwaliteit. Juist op dit punt zou vervolgonderzoek noodzakelijk zijn.

De onmisbare Neerlandicus

De bovenstaande kritische op- en aanmerkingen op de dissertatie van Rijlaarsdam moeten

niet beschouwd worden als een vernietigende vorm van kritiek, eerder als mogelijke vragen die gesteld kunnen worden en waarop men graag een ondubbelzinnig antwoord had willen krijgen. Eerder heb ik al opgemerkt dat Rijlaarsdam zijn onderzoek erg zorgvuldig heeft opgezet en uitgevoerd, en dat de problemen waarmee hij in zijn onderzoek geconfronteerd werd (de spanning tussen interne en externe validiteit) karakteristiek zijn voor elk onderwijskundig-empirisch onderzoek met praktische pretenties. Rijlaarsdams dissertatie is niet alleen de moeite waard voor collega-onderzoekers, maar ook voor docenten Nederlands die geïnteresseerd zijn in een didactisch programma voor de stelvaardigheid.

Rest de vraag: is het zinvol peer-evaluation in de praktijk toe te passen?

Wie peer-evaluation in de klas wil toepassen moet beseffen dat hij — ook in deze didactiek — niet van elk correctiewerk ontslagen wordt. Ten tweede dient hij te beseffen dat de door Rijlaarsdam ontwikkelde steldidactiek en daarmee samenhangend lesprogramma extreem veel tijd kosten. In Rijlaarsdams onderzoek domineerde het schrijfonderwijs zozeer, dat sommige van de betrokken docenten vreesden hun gewone programma voor het derde leerjaar niet af te krijgen. Naar Rijlaarsdams visie behoort het schrijfonderwijs de spil van het curriculum Nederlands te worden, waaraan andere vakonderdelen ten dienste worden gesteld. Tot slot dient hij zich — met het oog op de effectiviteit van peer-evaluation — ervan bewust te zijn dat er een bijzondere *interactie* bestaat

tussen de effectiviteit van een steldidactiek en de persoon van een docent — zo heeft Rijlaarsdam aangetoond. De betreffende interactie komt erop neer dat de ene docent in Rijlaarsdams onderzoek succes had met de peer-evaluation didactiek, de andere daarentegen juist met de 'traditionele' didactiek waarin de docent zelf de feedback gaf. Dit resultaat is een treffende illustratie van de inmiddels tot platitude verworpen stelling dat de docent, met zijn enthousiasme, zijn deskundigheid en specifieke kwaliteiten en voorkeuren een onmisbare schakel in elk leerproces vormt.

Noot

- 1 Wanneer hier gesproken wordt van 'geen positieve effecten van peer-evaluation', dan wordt daarmee niet bedoeld dat de leerlingen die onderwezen werden met de peer-evaluation didactiek, niet vooruitgingen in schrijfvaardigheid. Integendeel, zij gingen in de loop van het derde schooljaar flink vooruit. Die vooruitgang was echter, anders dan verwacht, niet substantieel groter dan die van leerlingen die van hun docent feedback ontvingen.

G.C.W. Rijlaarsdam *Effecten van leerlingenrespons op aspecten van stelvaardigheid* Dissertatie Universiteit van Amsterdam 1986. Uitgegeven als SCO-rapport nr. 88. ISBN 90 6813 105 2; prijs f 11,50. Te bestellen bij: Stichting Kohnstamm Fonds voor Onderwijsresearch, Grote Bickerstraat 72, 1013 KS Amsterdam.