

VERSCHENEN PROEFSCHRIFT

Anouk van Erden en Mik van Es, *Meten en maximaliseren van basale schrijfvaardigheid bij eerstejaarsstudenten in het hoger beroeps onderwijs*. (Proefschrift). Groningen: Universiteit van Groningen. ISBN: 978-90-367-6710-1 (pdf), ISBN: 978-90-367-6709-5 (boek), 487 blz.

Enige tijd geleden hebben de kranten vol gestaan over de slechte schrijfpredaties van eerstejaars studenten. De aanleiding hiervoor was bovenstaand proefschrift van Anouk van Erden en Mik van Es. Het is goed dat er weer aandacht is voor schrijfvaardigheid, maar met dit onderzoek wordt het schrijfonderwijs noch het schrijfonderzoek een dienst bewezen.

Het proefschrift begint met een inleiding over de tekortschietende schrijfvaardigheid van eerstejaarsstudenten, waarna de auteurs de stand van zaken in het schrijfvaardigheidsonderzoek bespreken. Daarbij zijn zij uitermate kritisch over het tot dusver verrichte onderzoek. Ik kan mij bij deze bespreking niet aan de indruk onttrekken dat grote delen van de relevante literatuur niet aan de orde komen, en dat wat wel besproken wordt tendentius weergegeven is. Daarnaast worden verschillende onbewezen stellingen ter verklaring naar voren geschoven. Van alle drie de problemen een voorbeeld. Proefschriften waarin de schrijfvaardigheid van studenten onder de loep genomen is, zoals van Van den Berg (2003), of De Jong (2006), of proefschriften waar de meetproblematiek centraal staat, bijvoor-

beeld Wesdorp (1974), De Glopper (1988), of Schoonen (1991), worden nergens in dit proefschrift genoemd. De bespreking van de literatuur is op zijn minst tendentius te noemen; in veel gevallen twijfel ik of ik wel dezelfde publicaties gelezen heb als de beide auteurs. Zo zouden Bonset en Hoogeveen (2007) voorstanders zijn van alleen maar plannen: niet schrijven maar plannen! Ik heb een andere conclusie getrokken op grond van hun werk. Of wat te denken van de bespreking van de meta-analyse van Graham en Perrin (2007): de auteurs trekken de conclusie dat volgens hen 'iedere vorm van schrijfonderwijs effectief [is], behalve daadwerkelijk schrijven' (p. 54). Die conclusie kan ik met de beste wil van de wereld niet uit deze meta-analyse halen. Een voorbeeld van een onbewezen conclusie van de auteurs: 'onderzoek heeft (...) vaker een negatieve rol gespeeld bij het tekortschieten van het schrijfonderwijs' (p. 54). Dit is niet alleen een onbewezen maar ook een on-onderzochte stelling, en ze kan in elk geval niet gebaseerd worden op Broekkamp en Van Hout-Wolters (2007). Mijn conclusie is dat de mening van de auteurs aan de basis ligt van de selectie en de weergave van de onderzoeksliteratuur. Men had een idee en dat moest coûte que coûte aangetoond worden (overigens aardig om te vermelden dat dit een verwijt van de auteurs aan andere schrijfonderzoekers is).

In de volgende hoofdstukken staat de constructie en beproefing van een nieuwe lesmethode (TAVAN) centraal. Terecht stellen de auteurs dat het verstandig is om bij

onderwijs altijd na te denken over het criterium waarmee dat onderwijs geëvalueerd kan worden. De auteurs constateren dat fouten hiervoor heel goed bruikbaar zouden kunnen zijn, waarbij fouten gedefinieerd zijn als dat 'wat een expert signaleert als fout' (p. 82). Dit is natuurlijk geen formele definitie, maar veeleer een operationele definitie die nodig is om verschillen tussen teksten aan te kunnen tonen, in casu meetbaar te maken. Conform de definitie zijn in een eerste deelstudie dertig teksten beoordeeld (waarvan tien geschreven door eerstejaars wo-studenten). De vier beoordelaars bleken duidelijk te verschillen in aantallen geconstateerde fouten. Vandaar dat de definitie aangescherpt is tot bevestigde fouten, i.e. fouten die door tenminste één andere beoordelaar bevestigd zijn. Later wordt de definitie van fout nogmaals aangescherpt tot kloppende bevestigde fouten, omdat niet alle bevestigde fouten ook echte fouten bleken. Omdat lange teksten meer fouten bevatten dan korte teksten staat verder het aantal fouten per honderd woorden centraal. (Terzijde zij opgemerkt dat de lezer op grond van de definiëring en alle daarop volgende aanpassingen geen idee heeft van wat een fout is; een algemene omschrijving helpt niet meer).

Deze eerste deelstudie heeft de pers ruimschoots gehaald. Op grond van tien teksten menen de auteurs aangetoond te hebben dat de gemiddelde wo-student 42 fouten maakt op één A4. Zij menen dat hun steekproef aselekt getrokken is. Hoewel dit strikt genomen niet waar is, moeten we ze daar niet al te hard over vallen (in welk onderzoek is wel een volledige random steekproef voorhanden?). Het gaat bij zo'n uitspraak vooral om de precisie van de schatting. En daar valt wel het een en ander op af te dingen. Ten eerste is een steekproef van tien teksten wel heel erg weinig. Indien een andere steekproef (teksten en/of studenten) getrokken was, zou men tot een ander gemiddeld aantal

fouten komen. Ten tweede verschillen de teksten in lengte (gemiddeld 279 woorden). Daarom zijn de gegevens geëxtrapoléerd naar een tekst ter lengte van een A4 (van 500 woorden). Omdat de geobserveerde tekstlengte gemiddeld een half A4 is en er een negatieve relatie is tussen tekstlengte en het aantal fouten per honderd woorden ($r = -0.65$; $p.104$), hebben we hier wellicht te maken met een overschatting van het werkelijke aantal fouten dat gemaakt zou worden als de studenten echt een A4 vol geschreven zouden hebben. Het enige dat we op grond van deze gegevens zeker weten is dat de 2% studenten die 500 woorden of meer geschreven hebben minder dan het geschatte aantal van 42 fouten gemaakt hebben.

Vervolgens worden taalmethodes voor hbo en wo bekeken en beoordeeld op mogelijke effectiviteit. Een inventarisatie en beoordeling van andere, oudere, methodes is altijd nuttig, en voor de ontwikkeling van een nieuwe methode een uiterst leerzame onderneming. In dit proefschrift worden de oordelen van één van de auteurs over verschillende taalmethoden uitgebreid voor het voetlicht gebracht. Het is natuurlijk jammer dat het maar de oordelen van één van de auteurs betreft, want in hoeverre deze oordelen generaliseerbaar zijn, is niet duidelijk. De auteurs zijn zich hiervan bewust, en daarom heeft de andere auteur ook één methode beoordeeld. Hoewel er duidelijke verschillen zijn tussen de oordelen van beide beoordelaars (p.173), concluderen zij dat beide beoordelingen eenzelfde beeld laten zien. Dit is op zijn minst een voorbarige conclusie. Immers, de verschillen in kwaliteit tussen de methodes worden geïnterpreteerd op basis van absolute scores, en die verschillen tussen de beoordelaars. Kortom, er wordt blind gevaren op de oordelen van één beoordelaar, waardoor de geldigheid van de conclusies wel enigszins in het gedrang komt.

De volgende stap is onderzoek naar de effectiviteit van de auteurs ontwikkelde en gepropageerde methode TAVAN. TAVAN is een digitale methode waarbij de studenten oefenen met zinnen waarin de fouten opgemerkt en verbeterd moeten worden. De controlegroep in dit experiment werd gevormd door studenten die werkten met het reguliere taalprogramma TAVA, waarin directe instructie gegeven wordt in een aantal voor schrijvers lastige linguïstische fenomenen. Hier wordt een onderzoeksopzet gebruikt waarbij de studenten als voormeting de fouten in een tekst verbeteren, en als nameting de fouten in een andere tekst verbeteren. Een dergelijke opzet wordt vaak gekenmerkt door een contaminatie van meetmoment en te verbeteren tekst. Om dit probleem het hoofd te bieden, heeft (circa) de helft van de studenten bij de voormeting fouten in de ene tekst en bij de nameting in de andere tekst verbeterd. Voor de andere helft van de studenten is de volgorde van de teksten omgedraaid. Zo is een onderzoeksontwerp ontstaan waarbij voor het tekst-volgorde-effect gecontroleerd kan worden.

Helaas is de rapportage van de statistische analyse onvolledig. De auteurs rapporteren een 2x2 variantie-analyse (onderwijsmethode x tekstvolgorde) op de verschilscore. Dat is een gemiste kans, want eigenlijk is de vraag of aangetoond kan worden dat studenten in de ene conditie meer vooruitgaan dan in de andere conditie. Dat wil zeggen: is er gegeven de onzekerheden in de beginmeting én de onzekerheid in de nameting een effect aantoonbaar van de experimentele didactiek? Door alleen de verschilcores te analyseren wordt in feite de onzekerheid in de beginmeting niet in de analyse meegenomen. Kortom, een analyse herhaalde-metingen, waar de interactie tussen meetmoment en conditie centraal staat, was meer op zijn plaats geweest.

Daarnaast kenmerkt de onderzoeksop-

zet zich door een denkfout. Eigenlijk is de vraag 'als we studenten na de experimentele didactiek een willekeurige tekst laten schrijven, maken zij daarin dan minder fouten?'. Let wel, een willekeurige tekst. De tekst, de opdracht, is dan ook een zogenaamde random factor en geen fixed factor; we willen immers aantonen dat de experimentele methode 'werkt' bij alle teksten. Kortom: we willen generaliseren over teksten. Dit is een lastig en pijnlijk punt in veel experimenteel schrijfonderzoek. Maar om dit probleem geheel onder het tapijt te vegen is onverstandig. Zeker als men denkt met een simpel voortoets-natoets-ontwerp de schrijfonderzoekswereld veel te kunnen leren.

In de resterende hoofdstukken wordt getracht het belang van fouten voor basale schrijfvaardigheid verder te onderbouwen. Dit wordt (o.a.) gedaan door de relatie tussen holistische oordelen van de auteurs over de teksten en het aantal kloppende bevestigde fouten per honderd woorden te bestuderen. Een schoolvoorbeeld van een slager die zijn eigen vlees keurt! Ook wordt de relatie tussen fouten en zeer onbetrouwbare studentoordeelen geanalyseerd. Helaas is de overeenstemming tussen de oordelen van de studenten zo laag dat hier in mijn ogen geen zinvolle conclusie aan verbonden zou mogen worden. Daarnaast wordt de stabiliteit van fouten in tekstshelften geanalyseerd. Opmerkelijk is dat de conclusies in dit proefschrift duidelijk afwijken van dat van alle andere onderzoek naar de stabiliteit van schrijfvaardigheid of de stabiliteit van spelfouten. Maar daar ligt vast weer de gebrekkige conceptualisatie en analyse van de hele onderzoeksgemeenschap aan ten grondslag....

Samenvattend: ik vind dit boek niet van het niveau dat van een proefschrift verwacht zou mogen worden.

HUUB VAN DEN BERGH

LITERATUUR

- Berg, B. A. M. van den (2003). *Peer assessment in universitair onderwijs. Een onderzoek naar bruikbare ontwerpen*. Universiteit Utrecht, Utrecht. (diss.)
- Bonset, H., & Hoogeveen, M. (2007). *Schrijven in het basisonderwijs. Een inventarisatie van empirisch onderzoek in het perspectief van leerplannontwikkeling*. Enschede: SLO.
- Broekkamp, H., & Hout-Wolters, B. van (2007). The gap between educational research and practice. *Educational Research and Evaluation*, 13, 209-220.
- Glopper, K. de (1988). *Schrijven beschreven: inhoud, opbrengsten en achtergronden van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs SVO. (diss.)
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of educational psychology*, 99(3), 445-476.
- Jong, J. de (2006). *Uitgesproken complex. Interactie tussen scriptieschrijvers en begeleiders*. Universiteit Utrecht, Utrecht. (diss.)
- Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen. Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs*. Amsterdam: UvA. (diss.)
- Wesdorp, H. (1974). *Het meten van de productief-schriftelijke taalvaardigheid. Directe en indirecte methoden: 'opstelbeoordeling' versus 'schrijfvaardigheidstoetsing'*. Purmerend: Muusses. (diss.)