

## Notities en Commentaren

### WAT BEWIJST WILBRINK EIGENLIJK?

H.G. Schmidt, M.L. de Volder, W.H. Gijselaers en L.M.M. Kerkhofs.

*Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch, Rijksuniversiteit Limburg.*

Maastrichtse medische studenten nemen in de eerste vier jaar van hun studie gewoonlijk om de zes weken deel aan een toets, bedoeld om kennis, verworven in die periode, te meten. In totaal gaat het daarbij om 20 toetsen. De functie van deze toetsen is een formatieve, dat wil zeggen dat zij studenten (en docenten) een inzicht bieden in de kwaliteit van het doorlopen leerproces, zonder dat een slechte prestatie op deze toetsen consequenties heeft voor de studievoortgang. (Summatieve beoordeling van studiestatistiek vindt plaats op basis van andere informatiebronnen.) Studenten worden dan ook aangemoedigd niet te raden naar het goede antwoord, maar door omcirkeling van een vraagteken aan te geven wanneer ze een bepaald item niet kunnen beantwoorden. Een dergelijk antwoordgedrag verschaft onderwijsmakers en individuele studenten een inzicht in waar de 'zwakke plekken' zitten en is dus in beider belang.

In een onderzoek dat enige tijd geleden in dit tijdschrift gepubliceerd werd (Schmidt et al., 1984) ontdekten wij dat het aantal goede antwoorden dat studenten op deze toetsen geven (omgezet in een percentagescore om vergelijking van resultaten tussen toetsen mogelijk te maken) over de loop der jaren zowel transversaal als longitudinaal een stijgende tendens vertoont. Het verband tussen studievordering (= het aantal 6-weeks blokken dat op een bepaald moment doorlopen is) en studiestatistiek, uitgedrukt in een product-moment correlatiecoëfficiënt, bleek waarden aan te nemen tussen de 0,40 en 0,50. We probeerden dit verschijnsel te verklaren door te veronderstellen dat gedurende de studie voorkennis toeneemt, die studenten in staat stelt nieuwe informatie die zij in de loop van hun studie moeten bestuderen, steeds makkelijker te verwerken. Ter ondersteuning van deze hypothese werden enige aanvullende gegevens gepresenteerd.

De resultaten van dit onderzoek verontrustten Wilbrink (1985) blijkbaar zozeer dat hij het nodig vond een nogal fors stuk te schrijven dat ons werk op een aantal punten kritiseert. Wilbrinks kritiek laat zich ongeveer als volgt samenvatten: er is geen zinnige theorie denkbaar die een positief verband tussen studiejaar en studiestatistiek zou kunnen verklaren, en dus kan dat verband er ook niet zijn. Wat Schmidt en zijn collega's presenteren is niets meer dan een artefact, een 'significante toevalligheid' zonder betekenis. En ja hoor, als je nog eens aan het rekenen slaat, en bijvoorbeeld de totaalscores van studenten corrigeert voor raden, dan vind je al snel een 'bijna symbolische correlatie' (de woorden zijn van Wilbrink) van  $-0,01$ .

Onze reactie hierop komt – ook kort samengevat – op het volgende neer: (1) er is wél een zinnige theorie denkbaar die een positief verband tussen studiejaar en -prestatie kan verklaren, en (2) in zijn pogen te bewijzen dat een dergelijk verband niet kan bestaan, maakt Wilbrink gebruik van een – hem blijkbaar onbekend – statistisch artefact, dat op kan treden wanneer verschillende, onderling sterk gecorreleerde, variabelen tot één index gecombineerd worden. Hij haalt daarmee zijn eigen redenering onderuit.

Deze twee punten worden hieronder nader toegelicht.

- (1) In onze bijdrage hebben we een poging gedaan een psychologische leertheorie die sterk de rol van voorkennis in het leerproces benadrukt (Anderson, 1977; Spilich et al., 1979) toe te passen op het onderwijs. We redeneerden daarbij als volgt. Men mag veronderstellen dat kennis opgedaan in een eerdere fase van de studie het bestuderen van nieuwe leerstof, dat wil zeggen leerstof uit een ander vakgebied binnen hetzelfde domein, op een later tijdstip vergemakkelijkt. Zo zullen bijvoorbeeld medische studenten die in het derde of vierde jaar het vak pathologie bestuderen daarbij steun ondervinden van kennis van anatomie en fysiologie of biochemie, opgedaan eerder in hun studie. Als we nu aannemen dat de moeilijkheidsgraad van de aangeboden leerstof en de daarop gebaseerde toetsen over de jaren heen niet systematisch varieert, dan kan verwacht worden dat studenten als gevolg van toenemende voorkennis in de loop van hun studie betere prestaties gaan leveren. Immers, bij min of meer constante moeilijkheidsgraad van de leerstof zal die voor hen relatief steeds 'makkelijker' worden naarmate de studie vordert en zullen hun prestaties stijgen. Uiteraard zal het op deze wijze gehypothetiseerde verband tussen studievordering en prestatie niet bijzonder sterk zijn, want leerstof variëert nou eenmaal wat in moeilijkheidsgraad (er bestaan 'erkend moeilijke' vakken en voor sommige verderop in een studie aan de orde komende vakken hebben studenten soms weinig voorkennis). Zolang echter die variatie niet systematisch is, moet enig verband teruggevonden worden.

Het verbaast ons dat Wilbrink deze toch niet volstrekt onzinnige redenering in zijn reactie negeert en volstaat met de conclusie dat hij niet ziet op grond van welke theoretische overwegingen de "noodzaak" van zo'n grillige relatie" gepostuleerd kan worden.

- (2) Wilbrink construeert vervolgens een nieuwe index voor studieprestatie, die een lineaire combinatie is van drie variabelen: het percentage goede antwoorden (% goed), het percentage foute antwoorden (% fout) en het percentage niet beantwoorde vragen (% vraagteken). En wel zo:  $I = \% \text{ goed} - \% \text{ fout} + \frac{1}{2}\% \text{ vraagtekens}$ . Hij duidt deze scoringswijze aan als 'gangbaar' wanneer je wilt corrigeren voor raden.<sup>1</sup> I blijkt  $-0,01$  te correleren met studievordering, en op grond hiervan meent Wilbrink 'bewezen' te hebben dat het door ons gevonden verband gewoonweg niet bestaat.

Nu is er een probleem met samengestelde variabelen als I. Wanneer je een dergelijke variabele correleert met een andere variabele dan wordt de grootte van die correlatie in belangrijke mate bepaald door de *gewichten* die je aan de samenstellende variabelen toekent. Je kunt een dergelijke correlatie zelfs naar believen manipuleren door gewoon de betreffende gewichten te wijzigen<sup>2</sup>. Iedereen die wel eens een multipele regressie heeft uitgevoerd, zal dat kunnen beamen.

Een voorbeeld. Stel dat Wilbrink bij toeval de volgende formule had gebruikt:  $I' = \% \text{ goed} + \% \text{ fout} - \frac{1}{2}\% \text{ vraagtekens}$ .  $I'$  is identiek aan I; alleen zijn twee tekens omgewisseld.  $I'$  correleert echter  $0,63$  met studievordering! Of neem  $I'' = \% \text{ goed} + \% \text{ fout} + \% \text{ vraagtekens}$ . Correleert precies  $0,00$  met studievordering om de doodeenvoudige reden dat  $I''$  voor iedere student gelijk is, namelijk  $100$ . De lezer wordt uitgenodigd zelf een index te construeren die een sterk negatieve correlatie met studievordering heeft.

Het zal duidelijk zijn dat de correlatie die Wilbrink met veel aplomb presenteert als 'bewijs' van de onzinnigheid van de uitkomst van het onderzoek van Schmidt, e.a. (1984) niet anders is dan een statistisch artefact. Een artefact dat vermeden had kunnen worden als Wilbrink de moeite had genomen een eenvoudige covariantiematrix van de betreffende variabelen te construeren (zie daarvoor bijvoorbeeld Nunnally, 1978, p. 165). Het had hem veel werk kunnen besparen.

## CONCLUSIE.

Wij zijn niet onder de indruk van de kanttekeningen die Wilbrink (1984) bij de resultaten van ons onderzoek plaatste. Zijn kritiek is deels ongerechtvaardigd en deels gebaseerd op een onkritisch gebruik van statistische methoden. Hij lijkt te geloven dat een verband als het door ons ontdekte, niet kan bestaan en gebruikt ongeoorloofde middelen – ongeoorloofd niet in juridische maar in wetenschappelijke zin – om aan te tonen dat dat verband dan ook niet werkelijk bestaat. Wij zijn daarom van mening dat de omschrijving die Wilbrink in vragende zin op ons onderzoek van toepassing verklaarde, in bevestigende zin op zijn reactie teruglaat: "Een rommelig onderzoek, om een geliefkoosde conclusie overeind te kunnen houden".

## Noten

1. Wij begrijpen niet waarom een correctieformule voor raden wordt voorgesteld voor een situatie waarin studenten geadviseerd wordt juist *niet* te raden. Ons is altijd geleerd dat scoringsmethoden een afbeelding moeten vormen van het feitelijke testgedrag van studenten. Overigens is de formule niet correct. Volgens De Groot en Van Naerssen (1973) had het moeten zijn:  

$$I = \% \text{ goed} + \frac{1}{2}\% \text{ vraagtekens.}$$
2. Dat is des te meer het geval als de samenstellende variabelen onderling sterk correleren, zoals dat het geval is voor de data die Wilbrink bekritiseert. % Goed en % niet-beantwoord correleren bijvoorbeeld  $-0,93$ . Nogal wieses natuurlijk, want het gaat om percentagescores en als % goed toeneemt, moet het % vraagtekens afnemen. Nogal wieses? Voor Wilbrink blijkbaar niet.

## Literatuur

- Anderson, R.C. (1977). The notion of schemata and the educational enterprise: general discussion of the conference. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.). *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum.
- Groot, A.D. de, & R.F. van Naerssen (1973). *Studietoetsen*, deel 2. Den Haag: Mouton.
- Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw Hill.
- Schmidt, H.G., M.L. de Volder, W.H. Gijsselaers en L.M.M. Kerkhofs (1984). Een positief verband tussen studiejaar en tentamenresultaat, en de rol van toenemende voorkennis. *Tijdschrift voor Onderwijsresearch*, 9, 183-188.
- Spilich, G.J., G.T. Vesonder, H.L. Chiesi & J.F. Voss (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18, 275-291.
- Wilbrink, B. (1985). Bewijs van het omgekeerde, voor Schmidt et al. (1985). *Tijdschrift voor Onderwijsresearch*, 10, 141-144.

Ontvangen 21-11-1984