

7#?...G»>...FFFFFF&fFåFåFåFå
F ñ F ¶ F ¶ B F Ë x F å G `
GÄGû*G»FGûGÇGûGûG»GûGûGûGûGûGû13
980

[80-9-13_verbanden_studieresultaten.rtf](#)

[80-9-13_verbanden_studieresultaten.pdf](#)

Oorzakelijke verbanden tussen
persoonskenmerken en studieresultaten.

[13-9-80 Voorspellen resultaten 9-80]

i n l e e i d i n g

Algemeen stukje over het belang van inzicht in
oorzakelijke samenhangen. Voor het
onderwijsbeleid is dat evident, en daarmee ook
voor het onderzoek voorzover dat
beleidsrelevant wil zijn.

Vandaar de zeeën van onderzoek naar effecten
van verschillende onderwijsmethoden,
beoordelingsprocedures, etc. Het 'oorzakelijke'
pleegt hier veelal experimenteel onderzocht te
worden, en levert in de regel eenduidige (zij het
ook vaak teleurstellende) resultaten op.

Een volgende stap van onderzoek is dan nogal eens het gaan onderzoeken van modererende, storende, etcetera variabelen, in de hoop tenminste specifiek effectieve methoden te kunnen vinden (ATI research is daar een voorbeeld van).

Een soort onderzoek dat wel iets weg heeft van dat laatste, ATI-achtige soort van onderzoek is het klassiek psychometrische paradigma waarin m.b.v. tests, toetsen, biografische gegevens e.d. onderzocht wordt welke persoonlijke verschillen samenhangen met verschillen op één of meer 'criterium' variabelen. Dit onderzoek is enkele stappen verwijderd van rechttoe-rechtaan experimenteel onderzoek, en het zal duidelijk zijn dat resultaten van dit soort onderzoek niet zonder meer oorzakelijk te interpreteren zijn.

Toch bestaat er voor dit tijdschrift [TOR, Tijdschrift voor Onderwijsresearch] een bepaalde traditie, ingezet door Crombag, Gaff & Chang (1975) waarin er stilzwijgend vanuit gegaan wordt dat op basis van correlationeel onderzoek vergaande uitspraken gedaan kunnen worden over oorzakelijke samenhangen in (in dit geval) het wetenschappelijk onderwijs.

Het is mijn bedoeling deze 'traditie' kritisch te reviewen, een aantal varianten van oorzakelijke samenhangen te onderscheiden, en aan te geven welk soort onderzoek nodig is om welk soort oorzakelijk verband te kunnen opsporen. Geprobeerd wordt tenslotte om uit bekend onderzoek over het functioneren van het wetenschappelijk [onderzoek] een voorlopige evaluatie op te maken over de kwaliteiten daarin, én de tekortkomingen.

Het probleem zoals gedemonstreerd aan de literatuur.

"Scores op schaaltes voor studiegewoonten werden gerelateerd aan studieresultaat na één en na vier jaar. Daarbij werden maar enkele significante verbanden gevonden. Maar géén schaalte, 'Herformuleren van de leerstof', toonde een negatief verband in alle vier onderzochte (sub)faculteiten, wat er op wijst dat het in eigen woorden formuleren van de leerstof een gevaarlijke gewoonte is. Andere gegevens in dit onderzoek wezen erop dat ook intellectuele capaciteiten en prestatie motivatie slechte voorspellers van studieresultaten zijn. Omdat dit een kenmerkend onderzoekresultaat

schijnt te zijn (Lavin, 1965), wordt geconcludeerd dat het tijd wordt een vraagteken te zetten achter de rationaliteit van het onderwijsstelsel, en dat we misschien uit moeten gaan zien naar mogelijkheden om betekenisvolle correlaties te bewerkstelligen waar die nu vaak afwezig blijven." Crombag, Gaff, en Chang 1975).

In hun discussie vragen deze auteurs zich af *"waarom studieprestaties niet beïnvloed worden door de studiegewoonten van studenten, door hun capaciteiten en motivatie, zoals zoveel docenten, studenten, en ouders nalevelijk aannemen? Misschien wordt het tijd de feiten te erkennen: studiesucces is in hoge mate een kwestie van geluk, of grotendeels afhankelijk van irrelevante factoren, of beide."*

Dit zijn krachtige uitspraken, die dan ook niet nagelaten hebben hun sporen te trekken in onderzoek dat sindsdien verricht is. Te vrezen valt dat deze woorden ondertussen hun werk ook gedaan hebben in het bevestigen van vooroordelen of misvattingen, zowel bij onderzoekers van onderwijs als bij docenten en bestuurderen. Misschien is het tijd de logische basis voor deze zwartgallige

onderwijsopvatting eens aan een kritisch onderzoek te onderwerpen,

Om te beginnen zullen Crombag c.s. zeker de eersten zijn om toe te geven dat het voor de student zeker niet onverschillig is of hij méér dan wel minder tijd in de voorbereiding van zijn tentamens steekt; dat het zeker verschil maakt of je goed gemotiveerd studeert of niet; dat het wel degelijk van belang is over een goed stel hersens te kunnen beschikken. Wanneer uit het onderzoek van Crombag c.s. déze samenhangen niet in de onderzoekresultaten tot uitdrukking zijn gekomen, dan kan alleen maar geconcludeerd worden dat kennelijk de aard van het onderzoek geen uitspraken toelaat op het niveau van het individu.

Er zijn tenminste twee wezenlijke problemen met de methodische opzet die Crombag c.s. gekozen hebben. Ten eerste wordt het criterium 'studieresultaten' geaggregeerd over meerdere afzonderlijke studieonderdelen. Daardoor wordt het moeilijk om op basis van de onderzoekresultaten uitspraken te doen die betrekking hebben op studiegedragingen die

telkens op één studieonderdeel (tentamen) gericht zijn. En ten tweede wordt de samenhang onderzocht tussen verschillen tussen personen in studiegedragingen, en verschillen tussen personen in studieresultaten. Uit dergelijk onderzoek kunnen geen conclusies getrokken worden die betrekking hebben op de samenhang tussen veranderend studiegedrag van de individuele student, en zijn behaalde studieresultaten. Op beide methodologische punten kom ik in volgende paragrafen nog terug.

Van Vliet (1975) onderzoekt de samenhang tussen capaciteitsvariabelen en tijdbesteding enerzijds, en tentamenresultaat anderzijds met eveneens weer correlationele technieken (regressieanalyse). Ook hier worden samenhangen **tussen** verschillen tussen personen onderzocht, en interpretaties gehanteerd die oorzakelijke samenhangen voor **individuele** studenten betreffen. Onderzoeksresultaten worden door Van Vliet gebruikt om tot bepaling van de cesuur op het tentamen te komen. Hij komt daarbij in verlegenheid door een significant negatieve samenhang tussen tijdbesteding en

tentamenresultaat. Omdat die negatieve correlatie hem niet bevalt, laat hij hem uit de regressieanalyse weg. Dat moet wel tot irrelevante resultaten leiden, die hij desalniettemin toch voor zijn cesuurbepaling gebruikt (vanwege een zekere schijnvaliditeit?). De negatieve relatie voor tijdbesteding had hem op het spoor van de misspecificatie van zijn model moeten zetten.

Oud (1977), in een artikel over de analyserichting bij causale hypothesen, kritiseert Crombag c.s.: in de analyse nemen Crombag c.s. studiegewoonten als afhankelijk van studieresultaat, en interpreteren de resultaten als oorzakelijk in de omgekeerde richting. Oud geeft geen heranalyse van de gegevens op de correcte wijze, dat is in dit geval bijvoorbeeld een analyse van het studieresultaat voor twee deelgroepen die extreem (hoog/laag, of gunstig/ongunstig) scores op de studiegewoonten schaaltes. Het is ook niet te verwachten dat zo'n heranalyse tot resultaten leidt die Crombag c.s. beter zouden bevallen. De eerder genoemde methodologische bezwaren tegen de studie en interpretatie van Crombag c.s. worden door Oud niet genoemd.

Elshout (1977) kapittelt Crombag c.s. op weer een ander punt. Conclusies zoals die van Crombag c.s., willen ze zinvol zijn, moeten berusten op enige kennis van zaken omtrent de voorwaarden waaronder voorspellende validiteiten zich ook kunnen manifesteren. Elshout geeft een inventarisatie van die 'kennis van zaken', en vat een en ander samen in de uitspraak dat validiteiten hoger zullen zijn naarmate de gebruikte tests en toetsen meer gelijk zijn aan de gehanteerde (te voorspellen) criteria. De benadering van Elshout is niet helemaal bevredigend. Hij beperkt zich er eigenlijk toe op te merken dat verkregen validiteiten geïnterpreteerd moeten worden naar wat er in de gegeven omstandigheden als maximaal haalbare validiteit in zit, verwacht mag worden. Op zich is dat correct, maar ook Elshout laat de wezenlijke methodologische problemen in Crombag c.s. relatie tussen onderzoek en interpretatie ongenoemd.

Elshout en Van Loo (1977) slaan de door Elshout (1977) aangegeven weg in, en laten zien dat voor mannelijke en vrouwelijke studenten de patronen in validiteitscoëfficiënten verschillend

kunnen zijn. Ook deze studie rommelt daarmee in de methodologische marge, in plaats van de grondfouten in de aanpak van Crombag c.s. aan te pakken. noot ¹⁾)

noot 1) de data-analyse vindt in de 'verkeerde' (Oud, 1977) richting plaats, zoals ook Crombag c.s. deden. De auteurs konden van de studie van Oud nog niet op de hoogte zijn, wél echter de redactie van het TOR.

Tromp en Wilbrink (1977) wijzen er, mede in verband met de studie en conclusies van Crombag c.s., op dat beschrijvende uitspraken niet rechtlijnig in voorschrijvende uitspraken te vertalen zijn. Wie conclusies wil trekken over de causale samenhang van variabelen, kan zich daarvoor niet beroepen op correlaties of partiële correlaties. Een eenvoudig padanalytisch model wordt gegeven, waaruit het wél mogelijk is om voor het onderwijsbeleid sterke conclusies te trekken. In een latere paragraaf zal dat model verder besproken worden.

Vorst (1978) is de volgende onderzoeker die de uitdaging gesteld door Crombag c.s. beantwoordt. Hij acht het mogelijk dat het gebrek aan sterke verbanden tussen relevante

variabelen en studiesucces (Crombag c.s. en Elshout en Van Loo) te wijten is aan het ongedifferentieerde criterium. Hij twijfelt er, mét Elshout (1977), ernstig aan of het uiteindelijk succes van de studie wel betrouwbaar valt te voorspellen of dat achteraf te achterhalen is welke de relevante invloeden op dat studiesucces zijn geweest. Vorst kiest in een nieuw onderzoek om die reden tussentijds studiesucces, en met name studietempo, als criterium-variabele. Ook Vorst probeert daarmee een wezenlijk fout design te verbeteren door aan de ingrediënten te sleutelen. Vorst analyseert zijn gegevens wél in de correcte (oorzakelijke) richting (Oud, 1977), maar slaagt er ook niet in de sterke samenhangen die hij op voorhand meende te moeten verwachten, in zijn onderzoek boven water te halen. Ook ingrijpende onderwijsveranderingen die mede in het onderzoek betrokken waren, bleken geen verschil te maken. Vorst gaat desondanks niet van harte mee in de cynische visie van Crombag c.s. op het onderwijsgebeuren: in zijn discussie probeert hij de rationaliteit van dat onderwijsgebeuren boven water te houden door uitgebreid aandacht te besteden aan

mogelijk relevante redenen voor het uitblijven van de verwachte verbanden.

Meerum Terwogt-Kouwenhoven (1980) gaat eveneens in op de uitdaging van Crombag c.s., door te proberen in dezelfde soort onderzoekopzet aan te tonen dat er wel degelijk 'zinnvolle' relaties zijn tussen studiesucces en de variabelen die daar redelijkerwijs mee samen moeten hangen. Het verschil met voorgaande onderzoeken is dat zij er inderdaad in slaagt, de tactische aanwijzingen van Elshout (1977) hanterend.

De aanbeveling van Elshout om testgedrag en criterium zoveel mogelijk op elkaar te doen lijken is hier goed uitgebuit. Je kunt je echter afvragen hoe zinvol deze Elshout-strategie is. In uiterste consequentie zou studenten gewoon gevraagd kunnen worden hoeveel kans ze denken te hebben (of nastreven) in één keer voor de propedeuse te slagen. Daarmee is een heel grote overeenkomst tussen 'test' en criterium bereikt. Zou dan ook een aanzienlijke 'predictieve' correlatie blijken, dan zou daarmee tenminste aangetoond zijn dat de onderwijssituatie er voor de betrokken

studenten allerminst irrationeel uitziet. Maar verder valt er t.a.v. het functioneren van het onderwijs, en de studenten daarin, geen relevante informatie uit te peuren, Hier lijkt dan ook het belangrijkste probleem met het onderzoek van Meerum Terwogt-Kouwenhoven te liggen: een hoge mate van voorspelbaarheid aantonen weerlegt de aanklacht van irrationaliteit van het onderwijs (Crombag c.s.), maar ook hier staan de gebruikte correlationele technieken het niet toe om de gevonden verbanden te interpreteren naar effecten die verandering in studiegedrag van de individuele student zal hebben op zijn of haar studiesucces. Het laatste blijkt in dit onderzoek alleen al uit het vinden van een aanzienlijke negatieve regressiecoëfficiënt voor gemiddeld schriftelijk eindexamencijfer in de voorspelling van het aantal afgelegde tentamens. Dat betekent, en die conclusie trekt de auteur helaas niet, dat de predictieve validiteiten ook in dit onderzoek niet overeenstemmen met het soort verbanden dat bedoeld wordt wanneer gesproken wordt over 'significante relaties tussen studieprestaties enerzijds en variabelen die daar redelijkerwijs mee samen zouden moeten hangen anderzijds'.

De gehanteerde methodologie past eenvoudig niet op de onderzoeksvraagstelling, hoe aardig de meeste resultaten er ook uit mogen zien.

Onderzoek en analyseer op hetzelfde niveau waarop je conclusies wilt trekken. (wilt concluderen).

Een valkuil waarin menig onderwijsonderzoeker ongemerkt belandt is de kuil met het label 'aggregation fallacy'. Het doet zich voor overal waar conclusies getrokken worden op een ander niveau dan waarop het onderzoek zich beweegt. Bijvoorbeeld: de gegevens hebben betrekking op groepen, en de onderzoeker trekt conclusies over individuen. Hoe fors een dergelijke vergissing erin kan hakken wil ik demonstreren aan de hand van de bekende dataset van Guilford (1954). Laat A t/m E vijf studieonderdelen voorstellen, gedaan door 7 studenten, voor wie zowel bestede tijd als het tentamenresultaat bekend zijn (tabel 1).

student	A	B	C	D	E
---------	---	---	---	---	---

totaal
(voor ieder vak telkens tijd en score)

1	5	6	5	5	3	4	5	6	3
3	21	24							
2	9	8	7	7	5	5	8	7	5
2	34	29							
5	3	4	3	5	3	3	7	6	1
6	17	24							
4	7	5	3	6	1	4	3	5	3
5	17	25							
5	9	2	7	4	7	3	8	2	5
3	36	14							
6	3	4	5	4	3	6	5	4	1
2	17	20							
7	7	3	7	3	5	5	5	5	5
4	29	20							
	43	32	37	34	27	30	41	35	2 3
25	171	156							

Tabel 1 (gefingeerde) gegevens over tijdbesteding en studieresultaten,

Wie beschikt over de gegevens in tabel 1 kan geen konklusies trekken over de mate waarin

een extra uur studie de individuele student een hoger studieresultaat oplevert. Daar zijn geen gegevens over, dus daar kan niets over gezegd worden.

Maar verder zijn de gegevens gedetailleerd genoeg om vragen op verschillende niveaus te kunnen beantwoorden. Voor ieder studieonderdeel afzonderlijk is de correlatie tussen tijdbesteding en studieresultaat respectievelijk: .11, -.26, -.13, -.05, en -.32. Correlaties die er uitzien alsof ze op reële gegevens gebaseerd zijn (vgl. bijv. Van Vliet (1977) die .26 vond).

Het gemiddelde (via z-transformaties) is -.13.

Over alle vijf studieonderdelen gesommeerd (laatste twee kolommen in tabel 1) is de correlatie tussen studietijd en studieresultaat -.26. De onderzoeker die er op uit is positieve correlaties te laten verschijnen waar ze volgens zijn vooropgezette mening zouden moeten verschijnen, kan zijn toevlucht nemen tot de volgende truc uit het handboek 'how to lie with statistics': voor de individuele student is de correlatie tussen studietijd en studie resultaat over de zeven onderdelen respectievelijk .91, .

87, .21, .32, -.056, .42, en .91. Het gemiddelde daarvan (berekend via z-transformaties) is .24.

De onderzoeker die niet tevreden is met een lage correlatie kan per studieonderdeel de gemiddeld benodigde tijd en het gemiddeld behaalde resultaat berekenen, en bemerkt tot zijn verrassing dat de correlatie tussen die gemiddelden over de vijf vakken .85 is

Met andere woorden: over iedere dataset zijn veelal heel wat verschillende correlaties te berekenen, al naar het niveau dat gekozen wordt, en of over personen, dan wel over studieonderdelen naar samenhangen gezocht wordt. Het is dan wel zaak om van tevoren goed afgepaald te hebben wat de eiken onderzoeksvraagstelling is, welke gegevens nodig zijn voor het beantwoorden daarvan, en op welke wijze welke datareductie technieken worden gebruikt.


[Alle verwijzingen behalve Tromp & Wilbrik 1977, en Guilford 1954, zijn naar artikelen in het TOR:]

 Tijdschrift voor onderwijsresearch 1975-76 : een twee-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch, Volume: 1 <https://objects.library.uu.nl/reader/index.php?obj=1874-208331&lan=en#page//12/18/90/121890336681949754930372395362806322489.jpg/mode/1up>

 Tijdschrift voor onderwijsresearch 1977 : een twee-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch, Volume: 2 <https://objects.library.uu.nl/reader/index.php?obj=1874-208332&lan=en#page//26/74/95/26749560703042481697071521276378389898.jpg/mode/1up>

 Tijdschrift voor onderwijsresearch 1978 : een twee-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch, Volume: 3 <https://objects.library.uu.nl/reader/index.php?obj=1874-208519&lan=en#page//98/99/90/98999060937023592333007028343388461078.jpg/mode/1up>

 Tijdschrift voor onderwijsresearch 1979 : een twee-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch, Volume: 4 <https://objects.library.uu.nl/reader/index.php?obj=1874-208520&lan=en#page//42/42/64/42426457430689864543607766942257668794.jpg/mode/1up>

 Tijdschrift voor onderwijsresearch 1980 : een twee-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch, Volume: 5 <https://objects.library.uu.nl/reader/index.php?obj=1874-208333&lan=en#page//10/54/58/105458251576153910545828926083522853606.jpg/mode/1up>