

[In de beide TOR-artikelen van 1980 heb ik een aantal van deze ideeën scherper gemaakt, en er een model bij geleverd. bw 2025]

[stellingen voor CRWO-studiedag 13 december 1974 in Groningen]

Universiteit van Amsterdam
cowo Centrum voor Onderzoek van
het Wetenschappelijk Onderwijs
Voetboogstraat 1
Postadres: Spui 21 1
Tel.: 525 2835

Amsterdam 13 december 1974

Ben Wilbrink

Stellingen die de discussie levendig kunnen houden:

- a. In het algemeen is het aan te raden om studenten aan volgende programma onderdelen te laten werken in plaats van onvoldoende tentamens te laten herhalen.
- b. De docent of de faculteit die herhaling van onvoldoende gemaakte tentamens eist zal het nut ervan moeten aantonen tegenover alternatieve tijdbestedingsmogelijkheden voor de student.
- c. Oplossingen voor het probleem waar de grens zakkenslagen te trekken, zijn beter realiseerbaar bij adoptie van bijv. het Amerikaanse grade point systeem.

Toelichting

In de volgende bladzijden een uiteenzetting over het hoe en waarom van deze stellingen. Daarbij zijn een aantal veronderstellingen nodig, die er nogal grof en misschien wat onwaarschijnlijk uitzien. Het gaat er in de discussie niet om,

verfijningen in deze veronderstellingen aan te brengen, of andere bewoordingen te vinden. Wat aan de orde is, is de aannemelijkheid van het 'model' zoals geschetst, hoe ruw ook.

Cijfers tussen haakjes verwijzen naar de bijlage met aantekeningen en literatuur.

Inleiding

Als onderwerp van discussie is de cesuurproblematiek gekozen: op welke wijze kan, mag of moet de grens zakken/slagen bepaald worden. Het probleem kan vanuit een groot aantal verschillende optieken benaderd worden: waarschijnlijk bent u bekend met de kernitem methode van De Groot, en de methode Wijnen als concurrent daarvan (1).

Een andere ingang is via uitwerking van de begrippen transparantie, billijkheid en doelmatigheid, door De Groot bepleit als noodzakelijke aanvulling op de langzamerhand wat knellende psychometrische begrippen betrouwbaarheid en validiteit (2). Dat alles laten we zoals het is, om de aandacht te richten op een merkwaardig soort consequentie die het laten zakken vaak heeft, nl. dat het studieonderdeel later alsnog voldoende gemaakt moet worden. We maken die analyse omdat bij het trekken van de grens zakkenslagen de consequenties die bepaalde beslissingen voor de betrokkenen hebben van belang zijn, en al naar hun aard tot andere cesuren kunnen leiden.

Veronderstellingen en terminologie

Voor de overzichtelijkheid zullen we veronderstellen:

- dat een onvoldoende gemaakt studieonderdeel overgemaakt wordt;
- dat ieder studieonderdeel en ook herhaling van studieonderdelen, evenveel tijd kost (zeg 2 maanden); (3)
- dat we een criterium hebben dat als maat kan dienen voor wat de student in een bepaalde periode aan nieuwe kennis en inzichten

vergaard heeft.

Een treatment is de studieactiviteit die de student toegewezen krijgt. In ons model zijn er twee treatments mogelijk: een nieuw studieonderdeel doen, of een al eerder gedaan onderdeel herhalen.

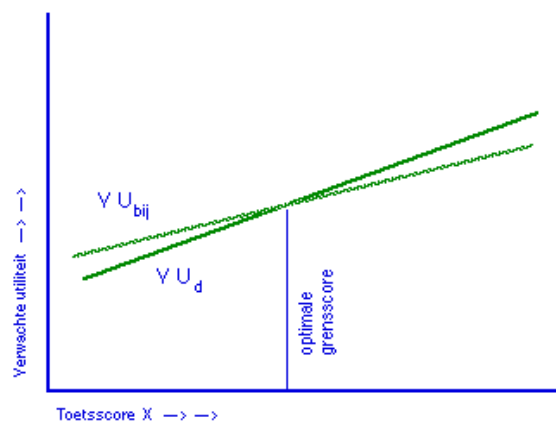
Excuus voor de term treatment i.p.v. behandeling o.i.d.; straks zal dit ongemak nog van pas komen.

Een payoff functie is het verband tussen de studieprestaties waarvoor we de grens zakkenslagen zouden willen bepalen, en het criterium dat we gekozen hebben als index voor wat de student in de erop volgende treatment 'wijzer' wordt (aan kennis, inzicht en vaardigheden).

De grote impliciete veronderstelling.

In ons globale model, evenals bij de traditionele onderwijspraktijk, is de niet uitgesproken veronderstelling dat de payoff functie voor de ene treatment de payoff functie voor de andere kruist (4):

[Ik leen hier de figuur 5 uit TOR 1980b om de ontbrekende figuur te vervangen. bw 2025]



Figuur 1

Gesteld dat deze situatie regel is, dan zou de cesuur zó gelegd kunnen worden dat diegenen het studieonderdeel moeten overdoen voor wie de verwachte payoff onder die treatment hoger is dan voor 'niet overdoen'.

Nu behelst de eerste stelling (blz. 1) dat in het algemeen de in figuur 1. geschetste interactie niet reëel is, en dat in plaats daarvan de beide payoff functies niet van elkaar te onderscheiden zijn, of elkaar niet kruisen (5):

Immers, in dergelijke gevallen moeten we alle studenten dezelfde treatment geven, wat er op neer komt dat we niemand vragen het studieonderdeel over te doen.

Zijn er uit de literatuur aanwijzingen te krijgen over de redelijkheid van bijv. de verwachting dat er een interactie (eerste figuur) bestaat?

Onderzoek naar Aptitude Treatment interacties

Een belangrijk onderwerp van onderwijsresearch is het zoeken naar Aptitude Treatment Interacties, ATI's, zoals in de eerste figuur geschetst. 'Aptitude' staat voor eigenschappen, kenmerken, kwaliteiten van studenten, en voor onze probleemstelling geldt dat we de score voor het studieonderdeel behaald, opvatten als een maat voor de aptitude: 'beheersing van de stof'. Veel ATI-onderzoek wordt gedaan met persoonlijkheidsvariabelen, intelligentie e.d. als 'aptitudes', en verschillende vormen van instructie als 'treatments' (6).

In deze ATI-research wordt gezicht naar interacties. Let wel, er wordt niet gezocht naar situaties waarin dergelijke interacties ontbreken, want als regel ontbreken ze inderdaad. Je mag dan ook stellen dat degenen die bij de inrichting van zijn onderwijs een bepaalde interactie veronderstelt, ook de bewijslast voor die veronderstelling draagt. Een interactie is een bijzondere situatie waarvan maar aannemelijk gemaakt moet worden dat ze inderdaad bestaat. Zoals we op de vorige blz. gezien hebben zijn het laten overdoen van studieonderdelen, en meer algemeen het geven van bijspiijkercursussen, maatregelen die op de impliciete veronderstelling van het bestaan van interacties gebaseerd zijn. Dit

als toelichting op de tweede stelling.

Hoe bijzonder het voorkomen van ATI's is wordt gedemonstreerd door de uitblijvende resultaten van onderzoek ernaar (als er al eens een ATI wordt gevonden dan is er bij replicatie van het onderzoek vaak weinig meer van over). Berliner & Cahen (6) geven als mogelijke reden dat de verschillen tussen de treatments vaak niet belangrijk zijn. Als dat zo zou zijn, zou je misschien voor onze situatie wat meer hoop op het bestaan van een ATI mogen hebben, gezien het forse verschil tussen het moeten herkauwen van oude stof en het verdergaan met nieuwe onderwerpen. Een andere verklaring die deze auteurs aandragen is dat in het onderwijs de meeste variantie in de criteriumscores (payoff) verklaard wordt door intelligentie. Wanneer het inderdaad zo is dat studieresultaten voornamelijk bepaald worden door een combinatie van intelligentie, ijver en motivatie, dan is er weinig hoop op het vinden van een ATI, omdat méér ijver, méér intelligentie en méér motivatie vrijwel altijd betere studieresultaten opleveren (en geen interacties) (7).

Vanuit de hoek van ATI-onderzoek wordt de veronderstelling van een interactie tussen laten overdoen of niet van studieonderdelen vooralsnog niet aannemelijk.

Bijspijkeren, homogene groepen, diagnostisch toetsgebruik

Een aantal onderwijspraktijken die alle impliciet ATI's veronderstellen, worden op zeer ontnuchterende wijze besproken door Hills (8). De conclusies die Hills trekt uit het overzicht dat hij geeft, zijn in overeenstemming met de resultaten van een Nederlands onderzoek (9): weinig of geen significante resultaten, waarbij de weinige wèl significante resultaten van evaluatieonderzoek van bijv. bijspijker cursussen (remedial teaching) nog wel eens in de niet verwachte richting zijn. Van Lieshout (9) geeft als mogelijke verklaring voor zijn resultaten dat het geven van speciale cursussen aan speciaal daarvoor uitgezochte studenten (homogene groepen) de mogelijkheid uitsluit dat studenten elkaar effectief kunnen helpen (wat bij heterogene

groepen het geval zou kunnen zijn). Gezien de zeer intensieve begeleiding die juist de 'zwakke' groepen van Van Lieshout ontvingen, lijkt het een weinig aannemelijke verklaring. Op indirecte wijze geeft Hills een groot aantal redenen waarom resultaten van bijspijkeronderwijs e.d. moeilijk of niet aantoonbaar zijn. Hij doet dat in de laatste bladzijden van zijn artikel waar hij de moeilijkheden opsomt die zich voor zullen doen wanneer men er expliciet op uit is het onderwijsprogramma flexibel in te richten op basis van Aptitude Treatment Interactions. Deze moeilijkheden, die uiteraard gelijk zijn aan de problemen die zich voordoen wanneer men onderwijsmaatregelen treft die impliciet ATI's veronderstellen, zijn volgens Hills:

- niet toereikende onderwijskundige kennis en medewerking van de docenten die het programma moeten ontwerpen en uitvoeren;

- de hoge graad van complexiteit die vereist wordt van de tests of toetsen die gebruikt worden om studenten aan de verschillende treatments toe te wijzen;

- problemen van administratieve aard (zoals zich bijv. in ons onderwijs voordoen wanneer studenten achterkomen op het min of meer rigide programma omdat ze bepaalde studieonderdelen moeten herhalen);

- er is nog geen theorie beschikbaar die als uitgangspunt voor het langs ATI lijnen opzetten van onderwijs kan dienen.

Met andere woorden, stel dat iemand het voorstel doet om speciale bijspijker cursussen te organiseren voor studenten die een studieonderdeel onvoldoende gemaakt hebben, waarbij van de student gevraagd wordt het onderdeel gewoon maar eens over te doen. In het licht van de analyse van Hills zou een dergelijk voorstel weinig reëel zijn, en tot grote problemen leiden.

Een ander, en compensatorisch, cijfersysteem

Wanneer het bestaan van een interactie tussen laten overdoen of niet, niet aannemelijk gemaakt kan worden, dan moeten we dit stukje onderwijsfolklore afschaffen. Hamvraag is dan, welke nieuwe

procedures het opengevallen gat kunnen opvullen.

Een opleiding is meestal bedoeld om de student een bepaalde (minimum) hoeveelheid kennis, inzicht en vaardigheden bij te brengen. Hoe kan dit globale doel toch bereikt worden zonder als eis te stellen dat onvoldoende studieonderdelen overgedaan moeten worden?

Volgens mijn derde stelling is een eerste stap, als voorbereiding op een oplossing, de overgang naar een ander cijfersysteem. Het wordt langzamerhand tijd om ons huidige cijfersysteem met al zijn vijven en zessen (8) en ingebakken tradities, in te ruilen voor een flexibeler systeem. Een vijf puntsschaal, oplopend van 1 tot 4 voor alle voldoende prestaties, en de kwalificatie 'onvoldoende' (zo u wilt met een 0 op de schaal aan te duiden) zou daarvoor overwogen kunnen worden.

In plaats van met cijfers kan misschien beter met letters gewerkt worden. Op dezelfde wijze als bij de oude cijferschaal de neiging bestaat om door een groot deel van de schaal te gebruiken veel studenten een onvoldoende te geven, kan deze vijf puntsschaal niet alleen 0 als onvoldoende gezien worden als een poging om (of overigens weinig rationele, zij het ook niet irrationele wijze) het aantal onvoldoendes dat uitgedeeld wordt kunstmatig te beperken. Inderdaad, gezien de beschouwing op de voorgaande blz., vallen er alleen in uitzonderlijke gevallen onvoldoendes. Examineisen zouden niet meer gesteld kunnen worden in het met goed gevolg afgelegd hebben van de diverse studieonderdelen, maar kunnen uitgedrukt worden in een bepaald cijfergemiddelde dat behaald moet worden (9).

Voor het cesuurprobleem betekent het dat ten aanzien van ieder afzonderlijk studieonderdeel alleen het veel minder ernstige probleem overblijft of de grens tussen de enkele onvoldoende die er valt en de rest, en de grenzen tussen de overige 4 punten op de beoordelingsschaal wel redelijk getrokken worden. Omdat cijfers voor de afzonderlijke studieonderdelen elkaar ruim kunnen compenseren, wordt de invloed van onbetrouwbaarheid op de

uiteindelijke gemiddelde score zeer klein vergeleken met waar we in het traditionele systeem mee moeten werken.

Blijft over het belangrijke probleem welke eisen in termen van te behalen gemiddelde cijfer (zo mogelijk aangevuld met andere criteria) in de opleiding gesteld moeten worden. Dit echter is een min of meer 'normaal' selectieprobleem en valt daarmee grotendeels buiten het kader van de cesuurproblematiek (12).

Aantekeningen

(1) A.D. de Groot, R.P. van Naerssen, e.d.: Studietoetsen, construeren, afnemen, analyseren. Mouton, Den Haag, 1969.

W.H.F.W. Wijnen, 1971: Onder of boven de maat. Swets & Zeitlinger, Amsterdam.

(2) A.D. de Groot, 1972: Selectie voor en in het hoger onderwijs. Staatsuitgeverij, Den Haag.

Definitie van het begrip transparantie: "Een examen is transparant als de gegadigde over alle informatie beschikt die hij nodig heeft om de voor hem best mogelijke strategieën voor voorbereiding op en deelname aan het examen te ontwikkelen".

(3) Deze aanname is niet zo gek als ze er op het eerste gezicht uitziet. Een praktikum dat overgedaan moet worden kost vrijwel evenveel tijd als de eerste keer. Een tentamen over moeten doen kan erg veel tijd kosten voor weinig 'transparante' toetssituaties, of bij een ongelukkige onderwijsprogrammering. Van Naerssen, Een model voor tentamens, openbare les, 1970, veronderstelt transparante toetsing en de mogelijkheid voor de student om zijn tijd die niet voor voorbereiding op de herhaling nodig is, nuttig te besteden. Genoemde, en andere, aannamen maken het tentamenmodel van Van Naerssen niet zonder meer vergelijkbaar met de in dit paper geschetste, meer 'traditionele', onderwijssituatie. Zie ook Van Naerssen, 1974: A mathematical

model for the optimal use of criterion referenced tests. Nederlands Tijdschrift voor de Psychologie, 29, 431 446.

(4) L.J. Cronbach & G.C. Gleser, 1965: Psychological tests and personnel decisions. University of Illinois Press, Urbana.

blz.:25: "Even more to the point is the fact that use of tests for educational placement implicitly assumes the existence of such different payoff functions".

(5) In deze inleiding spreken we van interactie slechts dan wanneer de payoff functies elkaar kruisen. In de literatuur wordt dit de disordinale interactie genoemd, ter onderscheiding van de ordinale interactie (waar de payoff functies elkaar bij denkbeeldige verlening buiten het aangegeven schaalbereik zouden kruisen, zoals in de laatste van de drie figuren getekend). De ordinale interactie is meestal niet interessant voor degenen die verschillende treatments in het onderwijs willen hanteren. Voor de statistische methodologie die nodig is om disordinale interacties te kunnen toetsen

Plomp (6), en de literatuurverwijzingen gegeven door Berliner & Cahen (6).

(6) D.A. Berliner & L.S. Cahen, 1973: TraitTreatment interaction and learning. In F.N. Kerlinger (editor): Review of Research in Education. Peacock Publ., Itasca, Illinois.

Tj. Plomp, 1972: Statistische technieken voor ATlonderzoek. Enschede, T.H.Twente, T.W.memorandum, nr. 45.

G. Salomon, 1972: Review of Educational Research, volume 42, no. 3, Heuristic models for the generation of AptitudeTreatment interaction hypotheses.

(7) Er zijn sterke aanwijzingen dat tentamens in veel situaties niet meer meten dan dergelijke variabelen als genoemd zijn. Van Naerssen (1968, congresboek 2e nationaal congres Onderzoek

van Wetenschappelijk Onderwijs) bijv. spreekt over clusters van tentamens die "essentieel dezelfde 'factor' meten: die combinatie van logisch redeneren, verbaliteit, ijver, efficiëntie, etc., die nu eenmaal nodig is om met succes te studeren".

(8) J.R. Hills, 1971: Use of measurement in selection and placement. In: R.L. Thorndike (editor), 1971: Educational Measurement. American Council on Education, Washington D.C.

Degenen die in het hoofdstuk geïnteresseerd zijn, zullen ook belangstelling hebben voor: M.R. Novick & P.H. Jackson, 1970: Bayesian guidance technology, Review of Educational Research, volume 40, 459-494.

(9) G.M.G.M. van Lieshout, 1974: Met verschillende cursusversies aansluiten op verschillen in beginniveau? Memoreeks Onderwijsresearch, jaargang 2.

(10) A.D. de Groot, 1966: Vijven en zessen. Wolters, Groningen.

(11) In vele gevallen zullen de eisen niet gedicteerd hoeven te worden door de (staf van de) betrokken opleiding, maar gesteld worden (als één van de toelatingseisen) door instellingen van vervolgonderwijs (zoals in het Amerikaanse systeem de graduate schools).

(12) Mogelijk is dit ook de achtergrond van het ontbreken van Amerikaanse literatuur over het voor ons onderwijs toch zo typerende cesuurprobleem. In een stelsel waar met Grade Point Averages e.d. gewerkt wordt zijn de cesuurproblemen a.h.w. ingeruild voor selectie, plaatsings en adviseringsproblemen, zie Hills (8).

[map publicaties, veronderstellend dat dit stuk als paper op dat CRWO-congres in Groningen is behandeld.]