

[stuk ingebracht op CRWO studiedag 1975]

Ben Wilbrink

6 november '75.

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

Combineren van tentamencijfers. Enkele stellingen.

Stelling 1. Er is slechts één aanpak, zowel bij het combineren van cijfers als bij cesuurbepaling voor afzonderlijke vakken, die rationeel juist is: een besliskundige.

Stelling 2. Het vaak geuite bezwaar tegen 1) dat ongelijksoortige uitkomsten toch moeilijk of niet met elkaar in waarde (utiliteit) vergeleken kunnen worden, gaat niet op.

Stelling 3. Ook wanneer heel strikte minimum eisen gesteld worden m.b.t. twee of meer nogal verschillende vaardigheden (studievakken) is de beste strategie vaak om te lage scores voor de ene door hogere scores voor andere te laten compenseren.

Stelling 4. Afgezien van het onder 3) bedoelde verband tussen betrouwbaarheid en compensatiemogelijkheden, is het combineren van cijfers géén meetprobleem.

Stelling 5. De vraag compenseren of niet kan niet beantwoord worden door oplossing van het cesuurprobleem voor afzonderlijke vakken.

Stelling 6. Die cesuurbepaling is overigens ook niet uit te voeren door naar nauwkeuriger metingen (in het grensgebied) te streven, zoals bij alle pogingen tot nu toe eigenlijk gebeurd

is.

Stelling 7. Het laten overdoen van onvoldoende gemaakte studieonderdelen berust op de onjuiste veronderstelling dat er geen betere alternatieven zijn.

Stelling 8. Het verlies in termen van levensplezier, onderwijsmiddelen, en maatschappelijke kosten voortvloeiend uit de onder 7) genoemde slechte gewoonte, maakt bijna ieder alternatief beter.

2

Stelling 1. Er is slechts één aanpak, zowel bij het combineren van cijfers als bij cesuurbepaling voor afzonderlijke vakken, die rationeel juist is: een besliskundige.

Toelichting: Beoordelingen worden gemaakt om bij onderwijsbeslissingen op basis van relevante informatie te kunnen werken. Waar het dan ook om gaat, is niet de kwaliteit van de beoordelingen, maar de kwaliteit van de beslissingen. Het argument hiervoor is lang geleden reeds door Cronbach en Gleser op uitgebreide wijze gegeven.

Wat we te doen hebben, is inventariseren welke beslissingen op basis van individuele cijfers, of combinaties van cijfers, genomen worden.

De vraag hoe je cijfers voor verschillende vakken combineert, kunnen we pas aanpakken wanneer er een behoorlijk inzicht bestaat in hoe de beslissingsprocedure er uit ziet: immers, de vraag hoe je cijfers combineert is ondergeschikt aan de vraag naar de beste beslissingsprocedure.

Heel summier beschreven is de besliskundige aanpak:

- expliciet stellen wie voor de beslissing verantwoordelijk is, of er belang bij heeft. De analyse zal geheel andere resultaten op leveren al naar gelang er op de belangen van de individuele student gelet wordt, dan wel die van de onderwijsinstelling, of bredere maatschappelijke belangen.

- inventariseren van alle in aanmerking komende

beslissingsalternatieven (waarbij best wat fantasie gebruikt kan worden).

- bij ieder beslissingsalternatief opsommen welke uitkomsten daarbij zoal tot de mogelijkheden behoren.

- bij iedere uitkomst schatting maken van de kans op die uitkomst.

- voor iedere uitkomst de utiliteit schatten in relatie tot de andere mogelijke uitkomsten van alle beslissingsalternatieven.

- als laatste stap: uitrekenen welke de beste beslissing is.

Bij de keuze van utiliteiten kan al direct rekening gehouden worden met ethische eisen, onderwijskundige uitgangspunten, onderwijsvisie, kwaliteitseisen t.a.v. te gebruiken informatie (Standards), óf men kan een en ander later inbrengen. Psychologische virtuositeiten kunnen uitgeleefd worden bij het zoeken naar relevante informatiebronnen en meetinstrumenten.

Stelling 2. Het vaak geuite bezwaar tegen 1) dat ongelijksoortige uitkomsten toch moeilijk of niet met elkaar in waarde (utiliteit) vergeleken kunnen worden, gaat niet op.

Toelichting: Uitkomsten die (desnoods met enig kunst en vliegwerk) in guldens geëvalueerd kunnen worden, zijn zeldzaam. De moeilijkheden die zich voordoen wanneer uitkomsten niet op eenvoudige wijze in dollars of guldens kunnen worden gevalueerd zijn door Cronbach en Gleser, en door velen na hen, zó breed geëtaleerd dat men niet heeft opgemerkt dat elders wel degelijk oplossingen zijn gegeven (Raiffa 1968, en Lindley 1971 geven inzichtelijke besprekingen). De methode komt er op neer dat een utiliteitsschaal geconstrueerd word (van nul naar één), door:

- uit de lijst van diverse mogelijke uitkomsten bij de opgesomde beslissingsalternatieven de beste of meest gewenste uitkomst de utiliteit 1 toe te kennen;

- idem de minst gewenste uitkomst de utiliteit 0;

- iedere andere uitkomst als utiliteit toe te kennen de kans p op de beste uitkomst (gekoppeld aan de kans $1-p$ op de minst gewenste uitkomst) waarvoor men deze uitkomst nog juist zou

willen inruilen.

De werkwijze betekent even wennen, maar vraagt van de beslissingnemer slechts explicitering van ideeën die anders op impliciete en oncontroleerbare wijze zijn beslissing bepaald zouden hebben.

De plaatsruimte ontbreekt voor een volledige uiteenzetting. Het idee is kort gezegd een maat voor utiliteit te construeren met dezelfde eigenschappen als waarschijnlijkheden, zodat verdere berekeningen eenvoudig mogelijk zijn.

Bij het toekennen van utiliteiten volgens deze procedure kan men zich voorstellen dat iedere afzonderlijke uitkomst 'vertaald' wordt tot een uitkomst in termen van een lot met een kans p op de best mogelijke uitkomst, en een kans $1-p$ op de slechtste uitkomst. Lindley 1971 gebruikt een ander beeld, en nog weer een andere mogelijkheid zou zijn de uitkomsten te vervangen door weddenschappen die men in plaats van die uitkomst zou willen aangaan: weddenschappen met als alternatieven de slechtste, resp. beste uitkomst.

4

Stelling 3. Ook wanneer heel strikte minimum eisen gesteld worden m.b.t. twee of meer nogal verschillende vaardigheden (studievakken) is de beste strategie vaak om te lage scores voor de ene door hogere scores voor andere te laten compenseren.

Toelichting: Een stelling waar weinig plezier aan te beleven valt: de bewijsvoering is door Lord (1962, 1963) gegeven. Desondanks geef ik de stelling, omdat het maar al te makkelijk is te vergeten dat we bij het filosoferen over de inrichting van ons onderwijs, en van examenprocedures, meestal spreken over vaardigheden, inzichten en kennis van studenten in platonische zin. Bijvoorbeeld, waar het gaat om minimumeisen doen we meestal alsof het gaat om ware scores op toetsen die ook nog verondersteld worden de bedoelde vaardigheden valide te meten. In de wat rommelige praktijk van alle dag

hebben we echter te maken met toetsen van enigszins twijfelachtige validiteit, en zeer lage betrouwbaarheid. Het is duidelijk dat uit lage betrouwbaarheid van de gemaakte beoordelingen consequenties voortvloeien t.a.v. de mate waarin we vast kunnen houden aan de idealistische eisenformulering waarvan we in eerste instantie meestal uitgaan.

Met name wanneer met multiple cutting scores gewerkt wordt (conjunctief gecombineerde cijfers) geldt dat alleen al op basis van de gebrekkige betrouwbaarheid van deze beoordelingen er nogal wat compensatorisch water in de conjunctieve wijn gedaan moet worden.

Daar kan een toegevoegd worden dat dit met name ook geldt waar het gaat om vakken die geheel verschillend zijn, d.i. juist ook in die situaties waar bijv. Van Naerssen (1968) zou zeggen dat je maar liever niet moet compenseren omdat de vakken factorieel zo ongelijk (niet 'homogeen') zijn.

5

Stelling 4. Afgezien van het onder 3) bedoelde verband tussen betrouwbaarheid en compensatiemogelijkheden, is het combineren van cijfers géén meetprobleem.

Toelichting: De stelling volgt direct uit stelling 1. De besliskundige analyse zal er geheel anders uitzien, al naar gelang wie hem maakt (de student, de instelling, overheidsinstanties). Voor optimalisering van de beslissingsprocedure zullen verschillende partijen dan ook gedeeltelijk verschillende informatie willen gebruiken, of dezelfde informatie op andere wijze inschatten. Dat heeft dan weer gevolgen voor de beste wijze waarop die informatie ingewonnen kan worden, met name ook op de te hanteren toetsingsprocedures. Het meetprobleem is slechts een afgeleid probleem (aan de informatiekant).

Stelling 5. De vraag compenseren of niet kan

niet beantwoord worden door oplossing van het cesuurprobleem voor afzonderlijke vakken.

Toelichting: De oplossing van het cesuurprobleem voor afzonderlijke vakken is juist afhankelijk van de wijze waarop cijfers gecombineerd worden in examenregelingen. Een omgekeerde aanpak is principieel onjuist, en mondt op zijn best uit in lapwerk dat voor de betrokkenen nog wel aanvaardbaar is, maar overigens een ondoorzichtige onderwijsstructuur oplevert. En dat kan niet alleen voor de studenten, maar ook voor 'de onderwijsinstelling' ongunstig zijn.

Stelling 6. Die cesurbepaling is overigens ook niet uit te voeren door naar nauwkeuriger metingen (in het grensgebied) te streven, zoals bij alle pogingen tot nu toe eigenlijk gebeurd is.

Toelichting: Een aantal verschillende doelen die men met het cijfergeven nastreeft moeten begripsmatig toch wel goed van elkaar gescheiden worden. Zak-slaagbeslissingen hebben in ons huidige w.o. bestel een onverhuld selectieve werking (ook zo bedoeld, inderdaad): welnu, als men dat wil handhaven, dan moet die cesurbepaling ook op dat doel expliciet afgestemd worden. En dan heeft een dergelijke cesurbepaling nog maar heel weinig te maken met andere functies van het cijfergeefspel, zoals feedback naar student en docent, motivatie, controle op tijdbesteding, e.d. Met name moeten we niet a priori aan elkaar gelijk stellen: onvoldoende=nonmastery, zakken=overdoen=brevet van onvermogen.

6

Stelling 7. Het laten overdoen van onvoldoende gemaakte studieonderdelen berust op de onjuiste veronderstelling dat er geen betere alternatieven zijn.

Toelichting: Een aanzienlijk deel van de cijfercombinatieproblematiek gaat schuil achter het gebruik om onvoldoende gemaakte studieonderdelen over te laten doen. Het geven van de mogelijkheid om onvoldoende cijfers door cijfers voor

andere vakken te compenseren, is een alternatief voor die verspillende traditie. Het is daarom interessant eens na te gaan of er enige rationele grond ontdekt kan worden voor het gebruik onvoldoende vakken te laten overdoen.

Uit de toelichting bij stelling 1. en 2. weten we dat dit probleem zich ook wel zal eigenen voor een dergelijke besliskundige aanpak. Doen we dat, dan ontdekken we alras waar ook Cronbach en Gleser al op wezen: dat we bij dit gebruik impliciet veronderstellen dat voor sommige studenten (met lage scores) verdere behandeling A (nl. overdoen) hogere utiliteit heeft dan verdere behandeling B (gewoon verder gaan met de studie bijv.). Het laten overdoen van onvoldoende gemaakte studieonderdelen veronderstelt dat er een disordinale interactie bestaat tussen studiestudenten zoals bij het tentamen gemeten, en de beide alternatieven overdoen, dan wel verder gaan met de studie.

Aha, zeggen we dan, hier wordt dus eigenlijk zoiets als een Aptitude Treatment Interaction, een ATI, verondersteld. En zijn dergelijke ATI's in het onderwijs niet behoorlijk zeldzaam (Berliner & Cahen, 1973); zijn ze niet erg moeilijk met opzet in de onderwijsstructuur in te brengen (Hills, 1971)? Is die veronderstelling dan wel reëel? Wordt het niet tijd om die veronderstelling eens op haar juistheid te toetsen (ofwel op theoretische gronden te weerleggen, ofwel maar liever direct op andere onderwijsprocedures over te gaan)?

Mogelijke alternatieven:

- compenseren van alle cijfers (al dan niet binnen een studiepuntenstelsel);
- de herhaling van het tentamen onmiddellijk aansluitend aan de eerste tentamenafname geven, of althans binnen zeer korte tijd (zeg een week);
- gewoon door laten gaan met de studie, met de aantekening dat onderdeel zus of zo onvoldoende was (dossier diploma of zoiets);
- in plaats van dit vak alsnog een ander vak extra doen;
- weliswaar hetzelfde vak nog eens bestuderen, maar dan geheel andere of nieuwe onderwerpen daaruit;
- de student uit de alternatieven laten kiezen, inclusief het

alternatief hetzelfde nog eens over te doen.

Stelling 8. Het verlies in termen van levensplezier, onderwijsmiddelen, en maatschappelijke kosten voortvloeiend uit de onder 7) genoemde slechte gewoonte, maakt bijna ieder alternatief beter.

Toelichting: Ik geef deze stelling, omdat het mijn sterke overtuiging is, en niet omdat ik een demonstratie ervan aan de hand van het nodige cijfermateriaal zou kunnen geven. Desondanks is het niet onwaarschijnlijk dat bij verder uitzoeken van deze zaak inderdaad zou blijken dat de verliezen die aan deze onderwijstraditie toegeschreven moeten worden, fantastisch hoog zijn. Een besliskundige aanpak van beoordelings- en selectieproblemen in het wetenschappelijk onderwijs lijkt niet louter een spelletje van academisch belang te zijn.

8

Biblia:

L.J. Cronbach en G.C. Gleser: "Psychological tests and personnel decisions". London, University of Illinois Press, 1965 (1957).

H. Raiffa: "Decision analysis. Introductory lectures on choices under uncertainty". London, AddisonWesley, 1968.

D. Lindley: "Making decisions". London, WileyInterscience, 1971.

F.M. Lord: "Cutting scores and errors of measurement" Psychometrika, 1962, 27, 19-30.

F.M. Lord: "Cutting scores and errors of measurement a second case". Educational and Psychological Measurement, 1963, 23, 63-68.

R.F. van Naerssen: "Van score tot beslissing: slagen of zakken".
In: Onderzoek van Wetenschappelijk Onderwijs. Tweede
Nationaal Congres, Utrecht, 1968.

D.A. Berliner & L.S. Cahen: "TraitTreatment Interaction and
Learning". In: F.N. Kerlinger (ed.): "Review of Research in
Education" Itasca, Illinois, Peacock Publ., 1973.

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

Enkele stukken van eigen hand, waarin onderdelen van de in de
stellingen aangesproken problematiek verder uitgewerkt
werden (bij mij verkrijgbaar, tel.: 020 525 2819):

Het laten overdoen van onvoldoende studieonderdelen.
Diskussiestuk, 13 december 1974.

Selectie voor numerus fixus is irrationeel. NRC/Handelsblad,
Cultureel Supplement, 27 september 1974.

Overzicht van de selectieproblematiek bij numerus
fixusregelingen. Concept rapport, 8 april 1975.

Het nemen van beslissingen: de beginselen van het
op rationele wijze nemen van beslissingen wanneer
meerdere alternatieven voorhanden zijn. Interne nota, 12
september 1975.