

Boeken

W.K.B. Hofstee

Principes van beoordeling. Methodiek en ethiek van selectie, examinering en evaluatie

Lisse: Swets & Zeitlinger, 1999, 216 p.



Beoordelen is een activiteit die velen moeten uitvoeren en waarbij iedereen af en toe als lijdend voorwerp betrokken is. Goede beoordelingen zijn goed voor (onder andere) de economie, slechte beoordelingen kunnen ellendige consequenties hebben voor individuen, organisaties en de maatschappij. Hoe kunnen de psychologie en haar beoefenaars de kwaliteit van het beoordelen vooruithelpen? Wim Hofstee publiceert over veelsoortige onderwerpen. In zijn boek *Principes van beoordeling* ontspint zich een web van draden waarop veel van die onderwerpen hun plek vinden. Ze hebben allemaal met beoordeling te maken.

Centraal in de behandeling van beoordeling staat het beoordelings-script. Daar beschrijft de auteur de verschillende rollen van de betrokken actoren: de beoordelaar B , de beoordeelde P , de beoordelingsautoriteit A (de institutionele opdrachtgever), de hogere instantie H die een procedurele toetsing kan uitvoeren. De relaties tussen die actoren zijn precair. Wim Hofstee heeft al die rollen zelf gespeeld en hij is 'terzake geëngageerde terzake deskundige' (p. 49) bij uitstek.

Staan op de schouders van WKB zijn lezers van dit boek in staat hun inzicht in het fenomeen beoordeling te verrijken en het ten bate van de eigen praktijk en die van anderen aan te wenden. Omdat beoordeling een rode draad door vele praktijken is, kunnen veel mensen er iets aan hebben. Hofstee behandelt vooral de onderwerpen selectie, examinering en programma-evaluatie.

De schouders van WKB zijn hoog. Inzicht wordt niet zonder stevig klimwerk bereikt. Het onderwerp is gecompliceerd, het netwerk is fijn vertakt en zoals bekend uit ander werk van de auteur houdt hij ervan om verschijnselen die hij als problematisch ervaart, te formaliseren in mathematische modellen.

De eerste drie hoofdstukken van het boek zijn normatief van aard. Het eerste hoofdstuk is een semantische analyse en afpaling van beoordeling. In de andere hoofdstukken zullen vanuit de basisbegrippen van beoordeling voortdurend semantische preciseringen worden aangebracht. In hoofdstuk 2 treden de hoofdrolspelers in het beoordelings-script ten tonele, elk met hun belangen en taken. In de rest van het boek treft men passages aan die de delicate relatie van de relaties tussen die spelers in toenemende mate aan het licht brengen. Aan de orde komt dat beoordelingen waardeoordelen zijn, dat ze feilbaar zijn en aan de normen van de tijd gebonden. Toch hebben beoordelingen tot doel een uitspraak te doen over kwaliteit (hoofdstuk 3). Kwaliteit heeft uiteindelijk maar één dimensie waarop de beoordeelde objecten zijn te rangordenen om een beslissing mogelijk te maken. Beoordeling op kenmerken is de route om tot het ééndimensionale kwaliteitsoordeel te komen.

In hoofdstuk 4 vindt er een overgang naar beoordelingstechniek plaats, daartoe genoodzaakt doordat de feilbaarheid van beoordelaars onverbidde-lijk aan het licht is gebracht. We moeten het doen met beoordelaars die het schoolcijfer 'onvoldoende' (vier plus/ min één) krijgen als het gaat om hun onderlinge overeenstemming. Zelfs experts scoren onvoldoende. De vraag is: hoe krijg je met feilbare beoordelaars toch de kwaliteit te pakken?

Hoofdstuk 5 neemt in het boek de centrale positie in. Het geeft de oplossing van het probleem van het feilen van individuele beoordelaars. Dit probleem is op te heffen door meervoudige beoordeling. Dankzij de bijdragen van Spearman en Brown ligt modelmatig vast dat betrouwbaarheid van beoordelingen toeneemt met het aantal beoordelaars.

Op de schouders van een gigant

Paul van der Maesen

De unieke componenten vallen tegen elkaar weg ten gunste van de gemeenschappelijke elementen. Om te kapitaliseren op het gemeenschappelijke kan men ervoor zorgen dat oordelen van afwijkende beoordelaars minder sterk worden meegewogen dan van beoordelaars die het gemeenschappelijk oordeel goed kunnen voorspellen.

In hoofdstuk 6 schetst Hofstee de problemen die aan zelfbeoordeling (bijvoorbeeld in de vorm van de bekende persoonlijkheidsvragenlijst) kleven. De paradoxaliteit van zelfbeoordelingen ('Ja' zeggen op 'Ik ben bescheiden') komt aan de orde. Doordat in deze methode elke P (beoordeelde) zijn eigen B (beoordelaar) heeft, mag niet van een test worden gesproken. Zelfbeoordelaars zijn in het beoordelings-script ook actoren die kunnen verschillen in hun neiging tot sociale wenselijkheid. De conclusie is dat de zelfbeoordelaar hoogstens een medebeoordelaar kan zijn.

Hoofdstuk 7 gaat over de constructie van vragenlijsten, opgevat als hulpmiddel voor de beoordelaar (B) om 'kwaliteit' op een betrouwbare wijze af te dekken. Methoden om in een vragenlijst een kwaliteitsdomein te bestrijken en om te verifiëren of dat is gelukt, komen aan de orde. Het meest praktisch bruikbare deel betreft dat over het opstellen van vragen. Kwesties als ontkenningen, onzakelijkheden, interne consistentie, aantal schaalpunten passeren de revue.

Hoofdstuk 8 betreft de verwerking en rapportage van vragenlijsten. De auteur analyseert belangrijke kwesties als correctie voor excentrisch schaalgebruik, weging van vragen en absoluut interpreteren van relatieve scores en geeft advies over hoe correct te handelen. De dimensionering van vragenlijsten, waaronder natuurlijk het circumplexmodel, vindt in dit hoofdstuk een plaats.

Hoofdstuk 9 behandelt experimenten, tests en toetsen. In tegenstelling met vragenlijsten gaat het om methoden om de beoordeelde P op de proef te stellen. Evaluatie-experimenten zijn verwaterde experimenten; verschillende bronnen van verwatering worden besproken. Bij de bespreking van tests komen aspecten aan de orde als 'maximum versus typical performance' en institutioneel versus individueel perspectief. Validiteit en effect en hun relatie krijgen aandacht van de auteur.

Hoofdstuk 10 betreft procedures die beoordelaars ondersteunen bij het combineren van allerlei informatie tot een eindoordeel. In het bijzonder gaat het om het afwegen van intuïtieve indrukken en van scores op tests/toetsen en objectieve indicatoren. In de eerste plaats moet het systeem verenigbaar zijn met de wens om kwaliteit als output te krijgen. De auteur noemt (en hekelt, door het hele boek heen) systemen die beoordeelde de gelegenheid bieden (feed-forward) om de uitkomst van het systeem te beïnvloeden waardoor er van kwaliteit weinig overblijft. Een systeem moet open zijn voor verbetering en voor creatieve aanpassing (bijvoorbeeld op grond van unieke, niet voorziene aspecten die een ν meebrengt). De auteur geeft aan de hand van een formeel model inzicht in de parameters die bepalen of in een bepaald geval meer gewicht dient te worden gegeven aan objectieve informatie dan wel aan intuïtieve oordelen.

De kwestie hoe beoordelaars normen, cesuren en standaarden bepalen en zouden moeten bepalen, wordt behandeld in hoofdstuk 11. De auteur laat zien dat beoordelaars hun ervaring (kennis van hoe prestaties in de populatie zijn verdeeld) en ad hoc verkregen informatie (prestaties in de steekproef) mengen bij het bepalen van een cesuur en werkt dit uit in een model. Andere onderwerpen van dit hoofdstuk zijn: verschillen tussen selectiebeoordelaars in strengheid (af testen) en kostenbaten-evaluatie van programma's.

Hoofdstuk 12 ten slotte behandelt gelijke behandeling of non-discriminatie van beoordeling met het oog op selectie. Verschillende psychometrische definities van non-discriminatie

worden onder de loep genomen waarna die van onderpredictie van de achtergestelde groep als bruikbare definitie overblijft.

Na dit overzicht van de inhoud van het boek plaats ik graag enkele kanttekeningen.

Ik vind het een monumentaal boek dat van groot belang is voor mensen die met beoordeling bezig zijn. Het boek van Wim Hofstee brengt je tot inzichten waardoor je weet waarom het vak toch zo interessant is.

Het boek is moeilijk, de informatie-dichtheid is groot, het is een kwestie van lezen en herlezen, ook voor mensen die in het vak bezig zijn. Ik vrees dat het boek niet de mensen bereikt die het hard nodig hebben: beoordelaars. Wkb mag gelijk hebben met zijn standpunt dat voor iedere beoordelaar het toegepast-psychometrisch gezichtspunt van beoordeling van centraal belang is, er moet daarvoor nog wel werk verricht worden in de vorm van een gepopulariseerde uiteenzetting van de inzichten van het boek. Daarnaast zou het boek vertaald moeten worden in het Engels omdat het beoordeling op een manier behandelt waarop het bij mijn weten niet eerder is gedaan (of gelukt) in de wetenschappelijke wereld.

Wim Hofstee intrigeert met eigenzinnig taalgebruik (bijvoorbeeld de beperking van de betekenis van *test*, *indicatoren* en *diagnostiek*) en met het verenigen van schijnbaar strijdige begrippen. Concurrentie (wedenschap) en kwaliteit (gemeenschappelijkheid) weet hij, in een soort polder-beoordelingsmodel, met elkaar te verzoenen door de taak van een beoordeling (bij meervoudige beoordeling) op te vatten als een taak om zo goed mogelijk het gemeenschappelijke oordeel te voorstellen.

Het vorige punt illustreert de voortdurende spanning die in dit boek (evenals in eerder werk van Hofstee) voelbaar is tussen de behoefte om als output van de beoordeling kwaliteit te realiseren en de angst om het beoordelingsproces te bureaucratiseren. Een stelregel van wkb die vaak terugkomt in het boek is de speelruimte die aan de beoordelaar gegund moet zijn om buiten het criteri-

um en de objectieve procedures (de bureaucratische mal) te gaan en opvallende (emergente) kwaliteiten van P in beschouwing te nemen, overigens op een manier die zeker niet vrijblijvend is.

Wat ik me afvraag is hoe het gesteld is met de intersubjectiviteit van de beoordeling van deze emergente kwaliteiten. Als die beoordeling in het algemeen onbetrouwbaar blijkt (ik wed dat dit het geval is), zouden beoordelaars er uiterst zorgvuldig mee moeten omgaan. De remedie tegen onbetrouwbaarheid, meervoudige beoordeling, is in de meeste contexten (selectie, examinering, personeelsbeoordeling) een onhaalbare luxe.

Zoals in zijn andere publicaties neemt de auteur slecht doordachte of hype-achtige begrippen en rituelen op de hak en prikt erdoorheen, zoals functioneringsgesprek, formatieve toetsen in het onderwijs, beoordeling van docenten door studenten, beoordelen op sollicitatiegedrag in plaats van op geschiktheid, bindend studieadvies, oneigenlijk gebruik van 360-gradenbeoordeling. Dit maakt dat er tussen alle serieuze beschouwingen door ook gelachen kan worden.

De uitgever heeft auteur en lezers geen goede dienst bewezen door veel spel fouten te maken of te laten staan. Met name in formules (bladzijde 70-71) is dit storend. Deze omissie wordt niet gecompenseerd door het feit dat de uitgever één bedoelde fout heeft laten staan.

Dr. P. van der Maesen de Sombreff is psycholoog bij Van der Maesen Advies voor Personeelsmanagement.