

Wim Hofstee -

Hofstee beoordeelt de beoordelaar

□ Henk van Berkel en Desirée Joosten-ten Brinke

We ontmoeten elkaar in de lunchroom op het Centraal Station van Utrecht. Wim Hofstee heeft recentelijk het artikel over Van Naerssen in ons blad gelezen en dat was zijn eerste kennismaking met het tijdschrift.

Wanneer is in uw loopbaan het onderwerp 'examens' interessant geworden?

Hofstee noemt zichzelf een scientist practitioner. "We voelden ons onderwijzers met verstand van toegepaste psychometrie, meer dan collega's uit andere faculteiten, en vonden dus dat we niet te luchthartig moesten

omgaan met het tentamineren van de studenten. Docenten vroegen zich af of hun studenten überhaupt wel iets leerden? Om dat te meten ontwierpen we een toets 'wetenschappelijk Engels' en legden die voor aan respectievelijk 1e jaars- en 2e jaarsstudenten. Het verschil in scores was ruim eenmaal de standaarddeviatie, dus er kon worden geconcludeerd dat er wat werd geleerd." In die tijd heeft Hofstee veel aan het onderwerp 'cesuren' gewerkt. Het gevolg daarvan was het ontstaan van de Hofstee-methode (zie kader), een typische compromis-methode die in de hele wereld bekend is geworden. Deze methode heeft lang stand gehouden, maar Hofstee ziet het gebruik nu langzaam wegzakken. Volgens hem komt dat doordat beoordelaars het werken met modellen met een relatieve component te riskant vinden. De Hofstee-methode, en eigenlijk de hele werkwijze van



Willem K.B. (Wim) Hofstee (1936, Bandung, Indonesië) studeerde psychologie (1954-1961) aan de Rijksuniversiteit Groningen en is daar, met uitzondering van een enkel jaar na de vervulling van de militaire dienstplicht, tot aan zijn emeritaat in 2001 werkzaam gebleven.

Tijdens zijn dienstplicht kwam hij terecht bij de Marine bij de afdeling die verantwoordelijk was voor selectie en onderzoek. Aan het einde van zijn dienstdtijd begon hij parttime bij de Rijksuniversiteit Groningen als wetenschappelijk ambtenaar. In 1963-'64 bracht hij een jaar door bij Educational Testing Service in Princeton. Hij startte rond 1965 met zijn promotieonderzoek op basis van data die hij tijdens zijn dienstplicht had kunnen verzamelen.

Naast vele interne bestuursfuncties aan de Groningse universiteit, als voorzitter van de subfaculteit, had hij zeer vele externe functies, als lid of voorzitter van onder andere het Nederlands Instituut voor Psychologen (N.I.P.), de Commissie Selectievraagstukken, de Beoordelingscommissie ZWO-SVO, de Commissie Programma-evaluatie en COTAN. Hoewel de inhoud van zijn leerstoel primair was gericht op de persoonlijkheidspsychologie, neemt zijn werk op het gebied van beoordelen, examineren en selecteren in Nederland een vooraanstaande plaats in. Hij begeleidde zestig promovendi.

Zie verder zijn uitgebreid cv in zijn wetenschaps-autobiografisch essay *It Ain't Necessarily So*: <http://www.adng.nl/data/docs/biografie/autobiografie%20w.k.b.%20hofstee.pdf>.



Protocollisering is het einde van de beschaving

De methode Hofstee van cesuurbepaling

De compromis methode houdt het midden tussen een relatieve en absolute methode van cijfer geven. Een docent die deze methode hanteert, stelt dat in principe een student bijvoorbeeld zestig procent van de maximale score dient te behalen. Maar indien dan blijkt dat meer dan negentig procent is geslaagd of meer dan vijftig procent is gezakt, past de docent de grens aan. Bijvoorbeeld: als minder dan vijftig procent van de studenten de zestig procent norm haalt, verlaagt de docent de norm naar een punt tussen de zestig procent score en de score waarbij vijftig procent van de studenten hoger scoren. Als meer dan negentig procent slaagt bij de zestig procent norm, wordt de nieuwe zak/slaaggrens het punt tussen de zestig procent en het score waarbij negentig procent slaagt. De genoemde getallen zijn uiteraard arbitrair. Al naar gelang de situatie kunnen docenten ze aanpassen. Maar het principe blijft gelijk: wanneer onverwacht veel of weinig studenten slagen voor de toets, wordt de oorspronkelijke norm zodanig aangepast dat de zak- en slaagpercentages binnen van te voren bepaalde grenzen blijven. Op die wijze wordt een compromis gevonden tussen enerzijds het noodzakelijk geacht kennisniveau en anderzijds een aanvaardbaar zak- en slaagpercentage. De compromis methode leidt er toe dat het aantal gezakten en geslaagden op de verschillende toetsmomenten geen al te grote fluctuaties vertoont. Als dat wel het geval is, is er alle reden te veronderstellen dat onvolkomenheden van de toets hiervan de oorzaak zijn. Daarom is de compromis methode, of een variant daarop, een verdienstelijke methode bij het zoeken naar een rationele oplossing voor een moeilijk oplosbaar probleem. De stappen om de compromis methode te hanteren:

1. Bepaal welk percentage van de maximale toetsscore (de norm) in principe de grens vormt tussen zakken en slagen;
2. Bepaal welk percentage gezakten aanleiding geeft om de norm naar beneden aan te passen;
3. Bepaal welk percentage geslaagden aanleiding geeft de norm naar boven aan te passen;
4. Bepaal het percentage gezakten c.q. geslaagden op basis van de oorspronkelijke norm;
5. Indien het voorgaande percentage niet binnen het gebied valt dat volgt uit stappen 2 en 3, kies dan een nieuwe norm zodanig dat het percentage gezakten c.q. geslaagden wel binnen dit gebied valt.

Uit: Berkel, H.J.M. van (1999). *Zicht op Toetsen*. Assen: Van Gorcum.

Hofstee, is gebaseerd op het sluiten van compromissen. Hij vraagt zich voortdurend af: Weet je dat wel zeker? Maar is dat nu echt zo? De titel van een essay dat hij recentelijk heeft geschreven en dat een terugblik geeft op zijn werkzame leven, heeft niet voor niets de titel *It Ain't Necessarily So*. Hij heeft een hekel aan de *homo economicus* die zoekt naar absolute waarden en slechts uit is op winstmaximalisatie. Om er aan toe te voegen dat mensen met dit soort gedrag met het omvallen van de Icesave bank en de DSB wel het deksel op hun neus hebben gekregen. Eigen schuld, dikke bult, voegt Hofstee er nog net niet aan toe, want dat past niet bij zijn karakter. Er zijn geen absolute waarheden.

Een ander belangrijk onderwerp uit zijn carrière is het lotingsysteem bij de medische opleidingen. Diep in zijn hart is hij een voorstander van *straight* loten, dus zonder rekening te houden met de hoogte van de eindexamencijfers. Omdat zoiets politiek niet haalbaar blijkt, kan hij zich neerleggen bij het compromis zoals we dat nu

kennen: gewogen loting. "Empirische wetenschappelijke argumenten waren ondergeschikt aan de maatschappelijk politieke argumenten," zo stelt hij. Een min of meer vergelijkbare discussie vindt op dit moment plaats over het bindend studieadvies. Hij vindt dat we dat niet moeten invoeren, gezien tal van psychometrische argumenten. "Maar," zegt Hofstee berustend, "ook daar zullen de politieke argumenten het gaan winnen van de wetenschappelijke argumenten."

Waar bent u nu mee bezig?

"Op dit moment heb ik net een artikel gepubliceerd in *Netherlands Journal of Psychology*.¹ Dit vind ik het mooiste artikel dat ik tot nu toe geschreven heb." Het handelt over het begrip intersubjectiviteit. Met dit artikel heeft hij het gevoel weer wat verder te zijn in het denken over beoordelen. Maar daarmee is het nog niet klaar. Eigenlijk zou Hofstee graag een boek schrijven waarin hij zijn beoordelingsfilosofie zou kunnen neerleggen. De kern van

Regressie naar het midden is een fact of life



het beoordelingsvraagstuk is volgens hem een kwestie van intersubjectiviteit en van weddenschappen. Bij dat laatste gaat het erom dat de beoordelaars samen de strijd aangaan om vooraf goed te kunnen schatten hoe iets gaat worden beoordeeld. Helaas, geeft Hofstee aan, gaan veel beoordelaars op zoek naar objectieve indicatoren. Indicatoren kunnen slechts een richting aangeven. "Protocollisering is het einde van de beschaving." Het gaat er bij het beoordelen om dat de beoordeling representatief is voor het beoordelingsforum (dat is de groep van beoordelaars). Daarmee loop je het risico dat er regressie naar het midden optreedt, maar "regressie naar het midden is een *fact of life*", aldus Hofstee.

Persoonlijkheid blijft een terugkerend thema tijdens het interview. Bij het beoordelen van persoonlijkheid blijft er weinig over van individuele verschillen. Het gebruik van klassieke statistiek blaast die verschillen op. Het is daarom goed om niet naar de groep te kijken, maar om individuen te beoordelen. En dan zal blijken dat de meeste mensen 'ok' zijn; sommigen niet en die verdienen volgens Hofstee alle aandacht.

Wat drijft u?

"Niks. Ik heb geen programma voor mijn leven gemaakt. Ik heb mijn hele leven gereageerd op absolute posities en ben groot voorstander van het relativisme. Maar dat verhindert niet om buitengewoon enthousiast te zijn over iets wat plotseling invalt. Waar Hofstee slecht tegen kan is als mensen zich gaan gedragen naar een model. Als dat teveel gebeurt, gaat het mis. "Modellen zijn er om de gedachten te bepalen, niet om het handelen te dicteren."

Wat is uw visie op examineren? Waarom hebben we examens en moeten we ze wel hebben?

"Goede vraag. Als we het over examens hebben, dan hebben we het over beoordelen en beoordelingsinstrumenten. Hierbij gaat het om vragen als 'wat is een goede beoordeling?'. Vaak wordt dan geredeneerd dat beoordeling het primaire proces is, maar daarmee is Hofstee het niet eens. Beoordelen staat in dienst van het onderwijs. Examens kunnen niet anders dan de bedoeling hebben om de kwaliteit van het onderwijs te stimuleren. Toetsen en examens zijn idealiter "congruent" met de onderwijsdoelstelling, zoals Messick² dat benadrukte: ze zouden gewenst leergedrag moeten uitlokken.

Eén aantekening wil hij hierbij wel maken. Het is al heel wat als examens de kwaliteit van het onderwijs niet teveel in de weg zitten. Het is een noodzakelijk kwaad, waar we goed mee moeten omgaan. Beoordelingsvrij onderwijs is namelijk onrealistisch. Er is pas sprake van onderwijs als er een beoordelingsmoment in past. Als studenten niet het doel hebben om een diploma te behalen, is er geen sprake van onderwijs. Anderzijds is het gebruik van nieuwe examenvormen om een krachtige leeromgeving te vormen, ook onrealistisch. Het is onvermijdelijk dat leerlingen werken voor hun examens en het is een naïeve gedachte dat met de invoering van alternatieve toetsvormen congruentie te bereiken valt. Een oplossing voor dit soort dilemma's is het werken vanuit een speelse opvatting. "Laten we doen alsof examens belangrijk zijn, dan zorg je dat het niet uit de hand loopt."

Hoe denkt u dan over competentiebeoordelingen?

Hiervoor verwijst Hofstee naar de recensie van zijn hand uit 2003.³ Daarin geeft hij aan: "Ik ben geen liefhebber van beoordeling van collectieve prestaties of van bijdragen aan het groepsproces, en van (mede)beoordeling door onbevoegden," en "Het is nog niet bewezen dat werken met nieuwe assessmentvormen in de praktijk beter werkt [...] Authentiek toetsen is een paradox; voor zover het iets betekent, houdt het in dat de student iets op tafel legt dat waarde in zichzelf heeft, zodat expliciete toetsing achterwege kan blijven." Hij zegt het niet zo direct, maar de hausse aan vakpublicaties over het toetsen van competenties kan hij moeilijk plaatsen.

Wat is het hoogtepunt uit uw carrière?

"Ik zie niet echt een hoogtepunt, en trots is een woord waarmee ik niet goed uit de voeten kan. Ik ben tevreden over hoe mijn carrière loopt of is gelopen."

¹Hofstee, W.K.B. (2009). Promoting intersubjectivity: a recursive-betting model of evaluative judgements, *Netherlands Journal of Psychology*, 65 (1), 2-9.

²Samuel J. Messick (1931-1988), 'Leader In Educational Testing Field', was werkzaam bij Educational Testing Service in Princeton, N.J.; hij is onder meer auteur van 'Assessment in Higher Education: Issues of Access, Quality, Student Development, and Public Policy' (1999).

³Hofstee, W.K.B. (2003). Boekbesprekingen van Henk van Berkel en Anneke Bax (red.), Toetsen in het hoger onderwijs. Houten/Diegem: Bohn Stafleu Van Loghum, 2002, 324 pagina's, ISBN 90-313-3639-4 en van Filip Dochy, Ludo Heylen en Herman van de Mosselaer (red.), Assessment in onderwijs: nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs. Utrecht: Lemma, 2002, 272 pagina's, *Tijdschrift voor Hoger Onderwijs*, 27, 57-61.

⁴Ben Wilbrink (*1944), van 1968-1999 werkzaam in onderwijsresearch aan de Universiteit van Amsterdam, eerst in het onderzoekcentrum COWO, later in het SCO-Kohnstamm Instituut; specialisatie: toetsen, examens, selectie (loten voor numerus-fixusopleidingen), aansluiting onderwijs-arbeidsmarkt. Zie: <http://www.benwilbrink.nl/>.



Men moet individuen beoordelen, niet de groep

Is er wel eens iets mis gegaan?

"Het is wel een keer voor gekomen dat ik aan iets heb meegewerkt en ik achteraf niet blij was met het resultaat. Het ging om het vaststellen van de criteria voor onderzoekskwaliteit, in opdracht van de Koninklijke Nederlandse Akademie van Wetenschappen. Het doel was de alfa- en gamma-richtingen enigszins in bescherming te nemen omdat er dreigde dat zij met dezelfde criteria beoordeeld zouden gaan worden als de bèta-richtingen. Dat zou dan zeer ongunstig uitpakken. Ons voorstel destijds was om op min of meer losse gronden, maar wel verdeelbaar, te komen tot een beoordeling van onderzoeksprestaties. Vervolgens gingen de bureaucraten er mee aan de haal met als gevolg waanzinnig gedetailleerde checklists van hoe je de kwaliteit van onderzoek meet. Zoiets mist de kern. Ik had er niet aan moeten beginnen. We hebben de bureaucraten destijds op verkeerde gedachten gebracht."

Hebt u leerlingen/opvolgers?

"Als iemand iets deed wat ik al gedaan had, dan vroeg ik of die niet iets anders kon bedenken. Met zo'n houding krijg je natuurlijk niet direct een 'eigen school'. Ik heb wel mensen aan het denken gezet en dat zie ik wel als een mooie verdienste. Er zijn wel veel rolmodellen, maar dat is dan altijd wederkerig. Ben Wilbrink⁴ is bijvoorbeeld iemand met wie ik graag in gesprek ga. Ook onder de studenten zitten soms wel personen die als een soort rolmodel optreden." Zo gebruikt hij nog steeds een uitspraak

van een vroegere student: het doel van onderwijs moet zijn dat een student leert zich een oordeel te vormen. Daar vindt Hofstee een grote mate van waarheid in zitten. Zo'n uitspraak neem je je hele leven mee. Nee, leerlingen of opvolgers heeft hij niet en hij citeert Sartre, "Ik ben ongeschikt om bevelen te geven." Hier is hij te bescheiden. Van zijn zestig promovendi heeft hij er ongetwijfeld een aantal gevormd.

Wat is de toekomst van examinering?

"Ik houd mijn hart vast voor de verdere bureaucratisering van het onderwijs, waarin examinering meer een afrekenmechanisme wordt. Het zou mooier zijn als ieder antwoord op iedere examenvraag van iedere leerling opnieuw beoordeeld zou worden. Dat we er niet bij voorbaat van uitgaan dat slechts één antwoord het juiste antwoord is. Laten we niet vergeten dat examens een bedreiging vormen van het primaire proces. Maar ik ben ook niet zo naïef om het standpunt in te nemen dat we ze daarom maar in de prullenbak moeten gooien. Ook hier ben ik realistisch of relativistisch. Van absolutisme moet ik niets hebben. Ik sta aan de zijlijn en becommentarieer de beoordeelaar. Dat is ook nodig. Als je beoordeelaars hun gang laat gaan, neigen ze naar het absolutisme. En dat kunnen ze niet waarmaken."

□ De heer dr. H.J.M. van Berkel en mw. dr. D. Joosten-ten Brinke zijn respectievelijke hoofdredacteur en redacteur van *EXAMENS*. E-mail: desiree.joosten-tenbrinke@ou.nl.

Ook de EVC procedures van



zijn beoordeeld door
de Stichting Examenkamer

Examen
kamer
kwaliteit door onafhankelijkheid