

### **Principes van beoordeling: Methodiek en ethiek van selectie, examinering en evaluatie.**

Swets & Zeitlinger, Lisse, 2000, ISBN 90 265 1575 8.

*Willem K.B. Hofstee*

‘Deze verhandeling is bedoeld voor wie te maken heeft met beoordelingen, dus voor vrijwel iedereen.’ Met de openingszin van zijn boek maakt Hofstee direct zijn intenties duidelijk: de lezer informeren over de principes van beoordelen in alle mogelijke disciplines. Hofstee gebruikt zijn kennis opgedaan in meer dan dertig jaar onderzoek naar beoordelingspraktijken voor het voetlicht te plaatsen. Aan bod komen de manco’s van onderwijsvisitaties, jurering bij turnwedstrijden, personeelsselecties, persoonlijkheidsbeoordelingen, studietoetsen, zelfbeoordeling, beoordelingen van subsidieaanvragen, onderwijsprogramma’s, beleidsmaatregelen en kunstuitvoeringen en nog meer.

‘Principes van beoordeling’ is een grondige en eigenzinnige studie geworden. Willem Hofstee becommentarieert de ideeën over beoordelingen van wat hij noemt romantici en bureaucraten. Romantici zijn zij die bijvoorbeeld per definitie voor zelfbeoordeling zijn volgens het principe dat de beoordeelde zelf het best weet wat zijn kwaliteiten zijn. Daar tegenover plaatst hij de bureaucratie, waarin de beoordelaar zoveel mogelijk aan banden is gelegd en de beoordeling tot een soort observatie wordt gereduceerd. Bijvoorbeeld door het voorschrijven van gedetailleerde standaard criterialijsten. Beoordelen blijft in de ogen van Hofstee echter mensenwerk. Hij noemt het een ambt. Volledige bureaucratisering is een bedreiging voor de professionaliteit van het beoordelaarsambt. Een persoon moet op eigen merites worden beoordeeld. Een standaardlijst laat daar geen ruimte voor. Dat de beoordelaar een ambtsdrager is, houdt echter ook in, dat een beoordeling niet willekeurig mag gebeuren. Richtlijnen voor zuiver beoordelen spruiten voort uit compromissen tussen de romantiek en bureaucratie van het

beoordelen. Bij zijn betoog gebruikt Hofstee zo veel mogelijk eigen argumentatie. Van de 101 literatuurverwijzingen bestaat bijna een kwart (25) uit verwijzing naar eigen werk. Aan sommige van de overige 76 verwijzingen zal hij op een of andere manier ook hebben bijgedragen: ik telde 8 ‘Groningse’ proefschriften en een niet precies vast te stellen aantal andere ‘Groningse’ publicaties (Hofstee is hoogleraar aan de Rijksuniversiteit Groningen).

#### *Het beoordelingsscript*

In het eerste hoofdstuk presenteert Hofstee een model – het beoordelingsscript – met behulp waarvan hij de gemeenschappelijke principes, die ten grondslag liggen aan beoordelingen binnen verschillende contexten, blootlegt. Met behulp van dit abstracte model worden in de overige elf hoofdstukken methodiek en ethiek van het beoordelen uit de doeken gedaan. De schrijver vermeldt dat Scriven in 1994 in een artikel in *Studies in Educational Evaluation* (titel: *Evaluation as a discipline*, blz 157-166) voor een vergelijkbare integrale benadering kiest, maar deze zijn op het moment dat ‘Principes van beoordeling’ verschijnt (nog) niet uitgewerkt. Mocht Scriven ooit tot deze uitwerking komen, dan betwijfel ik of het veel overeenkomsten zal hebben met de inhoud van ‘Principes van beoordelen’. Zoals ik hierboven reeds opmerkte behandelt Willem Hofstee de beoordelingsproblematiek op een zeer particuliere en erudiete wijze.

Beoordelen wordt afgekaderd tot die gevallen waarin mensen, op gezag van anderen, de kwaliteit(en) van iets van iemand vaststellen (blz 14). Hofstee omschrijft beoordelen als een specifieke manier waarop mensen met elkaar omgaan, wat hij vertaalt naar het volgende beoordelingsscript. Een *B* (één of meer beoordelaars) doet een uitspraak over de *kwaliteit* van *iets* (een prestatie, product, aspect) van *P* (een persoon of groep), op *gezag* van een beoordelingsautoriteit *A*. Men kan tegen de beoordeling in beroep gaan bij een hogere instantie *H*. Indien we het beoor-

delingscript doorlopen voor deze recensie: is *B* de schrijver van deze recensie, beoordeelt *B* de kwaliteit van iets het boek 'Principes van beoordeling' van *P* Hofstee, op gezag van een beoordelingsautoriteit *A*, in deze casus de redactie van dit tijdschrift. Deze casus illustreert een ergernis van velen die beoordeeld worden, namelijk het ontbreken van de *H*, de hogere instantie die de beoordelaar eventueel kan terugfluiten. Tenzij Hofstee of de uitgever vindt dat hen met de inhoud van deze recensie smaad wordt aangedaan. In dit uitzonderlijke geval fungeert de rechter als *H* de hogere instantie.

Thema's die behandeld worden zijn kwaliteit, de feilbaarheid van de beoordelaar, meervoudige beoordeling, zelfbeoordelingen, constructie van vragenlijsten, verwerking en rapportage van vragenlijsten, het experiment, de test en de toets, beoordelingsondersteuning, cesuurbepaling en ongelijke beoordeling. De auteur gaat hierbij voortdurend misconcepties, misstanden, modieuze grillen en sofisterij te lijf met vaak psychometrisch onderbouwde argumentatie. De lezer kan genieten van vele scherpzinnigheden. Bijvoorbeeld op blz 51 over misconcepties over onderwijskwaliteit: "...de marges zijn vrij miniem (van leerprestaties tussen scholen: recensent)...en dat 'profiel' en 'missie' voornamelijk het karakter hebben van retorische franje. Het voorziet ontegenzeggelijk in een behoefte als hetzelfde product in verschillende verpakkingen wordt aangeboden: sommige mensen zijn dol op kiezen." Of op bladzijde 133 over meerkeuzevragen: "de multiplechoicetest wordt wel verdedigd als toonbeeld van objectiviteit, in plaats van als verlegenheidsoplossing voor problemen in het massaal onderwijs."

Aan het momenteel binnen de onderwijskunde populaire fenomeen van zelfbeoordeling besteedt Hofstee een heel hoofdstuk. Hierin beargumenteert de schrijver:

(a) waarom zelfbeoordelingen niet bestaan. Zelfbeoordelingen zijn structureel paradoxaal (iemand die zich zelf al bescheiden beoordeelt, is met deze positieve zelfwaardering feitelijk onbescheiden) en zelfbeoordelingen zijn in feite een spel of onderhandeling tussen de beoordelingsautoriteit en de zelfbeoordelaar;

(b) waarom een aantal veel gebruikte argumenten tegen zelfbeoordeling niet steekhoudend is. Zoals het bezwaar van psychologen, dat sociaal wenselijk gedrag leidt tot 'social desirability bias';

(c) dat er geen bezwaar is om een student of sollicitant te laten deelnemen in de beoordeling als mede-beoordelaar, mits de verschillende oordelen worden gewogen. In dit geval heeft de zelfbeoordeling zelfs positieve kanten.

#### *Een parel*

In zijn voorwoord geeft Hofstee zijn pretenties prijs, namelijk een omvattende en representatieve bijdrage leveren aan een beter gevoel voor beoordelen. In mijn ogen is de schrijver hier met glans in geslaagd. Onderwijskundige literatuur over beoordelen vanuit de psychometrische benaderingswijze is vaak weinig toegankelijk voor lezers die niet over een grondige psychometrische achtergrond en/of interesse beschikken. Daarnaast gaan dergelijke publicaties vaak voorbij aan de ethische en psychologische kant van beoordelen. De literatuur vanuit de 'alternatieve assessment stroming' heeft vaak veel weg van pamfletten tegen de onmenselijkheid van de gesloten vraagvormen. De termen beoordelen en toetsen zijn taboe en worden vervangen door meer eufemistische termen als assessment en evalueren. Hofstee combineert zoals hij dat noemt blikken onder de motorkap, oftewel psychometrische methodologie met nuchtere soms zelfs humoristische argumentatie uit de ethiek, wat een boek met veel diepgang heeft opgeleverd. Enig effectbejag is Hofstee niet vreemd. Het boek wemelt van de gecursiveerde woorden of delen van woorden. Bijvoorbeeld de tweede regel van het boek: 'Het accent ligt daarbij op degene die *verantwoordelijk* is ....dus op beoordelaars.'

'Principes van beoordelen' is niet alleen interessant voor beoordelingsspecialisten, maar ook voor hen die beoordelen of beoordeeld worden. En dat is bijna iedereen, zoals Hofstee terecht opmerkt. Of hij in de praktijk alle door hem genoemde doelgroepen zal bereiken is echter de vraag. Sommige delen van het boek vergen de nodige psychometrische voorkennis. Bij een deel van de leden van de doelgroepen zal deze kennis ontbreken. In-

dien zo'n lezer dergelijke passages echter overslaat of niet begrijpt, raakt hij de rode draad in het betoog kwijt en zal hij waarschijnlijk afhaken. In 'Principes van beoordeling' staan zeer toegankelijke passages en moeilijke tekstdelen door elkaar. (Dit is kenmerkend voor de stijl van Hofstee.) Het boek zou een meer evenwichtige moeilijkheidsgraad en grotere toegankelijkheid verkrijgen, indien de psychometrische uitwerkingen in appart te lezen voetnoten of tekstfragmenten waren geplaatst. En een vertaling naar de beoordelingsproblematiek in de hoofdtekst was opgenomen.

Ondanks de bovengenoemde onevenwichtigheid is 'Principes van beoordeling' van een vergelijkbaar niveau als de belangrijkste werken van de nestor van de Nederlandse beoordelings- en methodieliteratuur, A.D. de Groot. Ik verwacht dan ook dat het van invloed zal zijn op het denken van veel onderwijswetenschappers, psychologen en andere beoordelingsdeskundigen. Toepassing van en verwijzing naar Hofstee's beoordelingsscenario is bijvoorbeeld al te vinden in papers over competentiegericht beoordelen.

Met de unieke opzet en degelijke uitvoering van dit boek heeft Hofstee een parel toegevoegd aan de literatuur over beoordelen.

*Erik Driessen*

*Faculteit Geneeskunde*

*Capaciteitsgroep Onderwijsontwikkeling &*

*Onderwijsresearch*

*Universiteit Maastricht*

### **Making sense of word problems**

Swets & Zeitlinger, Lisse, 2000 xxvii+203 pagina's.  
ISBN 90 265 1628 2.

*L. Verschaffel, Br. Greer, & E. De Corte*

Rekenvraagstukken (redactieproblemen) vormen nog steeds een belangrijk struikelblok in het rekenonderwijs in de bovenbouw van de basisschool en zelfs vaak ook nog in de eerste fase van het secundair onderwijs. Niet alleen in Nederland, maar over vrijwel de gehele wereld. Net als vele anderen hebben Verschaffel, Greer & De Corte zich al verscheidene jaren over deze problematiek geboogen. In dit boek presenteren zij op syste-

matische wijze hun onderzoek over de afgelopen decennia. Het is beslist geen compilatie van vroegere artikelen geworden, maar een nieuw gloedvol betoog aan de hand van eerder onderzoek van de auteurs en vele anderen. Het is overigens juist een van de charmes van dit boek dat de auteurs geprobeerd hebben om in een strakke, systematische argumentatie het probleem van de rekenvraagstukken in het onderwijs nader te analyseren. Ze hebben de lezer daarbij niet overspoeld met data, statistische parafernalia of gesofisticeerde technieken, die de lezer al gauw het gevoel kunnen geven dat er iets te verbergen valt in plaats van op te helderen. Nee, de auteurs van dit boek kiezen ervoor om vooral op kracht van een heldere redenering hun stellingen te onderbouwen. Het resultaat mag er zijn.

De auteurs nemen in dit boek hun vertrekpunt in het bekende gegeven dat leerlingen bij het oplossen van rekenvraagstukken vaak hun wereldkennis buiten haakjes zetten: ze geven een antwoord dat in zeker opzicht thematisch correct kan zijn, maar dat als antwoord voor het gegeven probleem eigenlijk nergens op slaat. Het overbekende 'Hoe oud is de kapitein? (Een kapitein heeft vijf schapen en drie geiten aan boord, hoe oud is de kapitein?) is daar een voorbeeld van, maar ook wat serieuzer opgaven als 'we gieten water van 40 °C en van 80 °C bij elkaar, hoe warm wordt het water dan?' of het inmiddels overbekende 'busprobleem' ('Een school met 300 leerlingen gaat op schoolreis, ze bestellen bussen waarin per bus 40 leerlingen kunnen. Hoeveel bussen moet de school bestellen?'). Het is geen zeldzaamheid om bij dit soort opgaven antwoorden te krijgen als 'Het water wordt 120 °C', of "Ze hebben 71/2 bus nodig". Hoewel die kinderen in hun dagelijks leven best weten dat halve bussen niet verhuurd worden en ook niet praktisch zijn in het vervoer, geven ze toch zo'n *onrealistisch* antwoord, omdat dat nu eenmaal de uitkomst is van 300 : 40. Idem dito met het water. In het rekenonderwijs staat voor vele kinderen dit soort wereldkennis buitenspel.

In deel 1 van het boek onderwerpen de auteurs dit fenomeen aan een nader onderzoek. Ze laten zien dat het geen lokaal verschijnsel is, maar hardnekkig voorkomt in verschillen-

de landen. Om de aard van dit probleem te achterhalen doen de auteurs verschillende deelonderzoeken om te zien wat de invloed van de oriëntatie van de leerlingen is, of wat de invloed van de aard van de opgave is. In dat onderzoek blijkt dat de prestaties van de leerlingen niet uitzonderlijk verbeteren als ze erop alert gemaakt worden dat in sommige gevallen de uitkomsten merkwaardig zijn. Iets betere, maar toch evenmin echt opzienbarende resultaten werden verkregen in onderzoek waarin de opgaven zelf werden aangeboden in meer realistische, levensechte situaties. Maar ook in dit geval gaf zo'n 50% van de leerlingen nog onrealistische antwoorden.

De auteurs redeneren dat er iets anders aan de hand moet zijn. Hun hypothese is dat het opvallende gedrag van de leerlingen veroorzaakt wordt door de manier waarop de leerlingen de regels van het rekenonderwijs hebben leren hanteren. In hun ervaringen in het rekenonderwijs en onder invloed van de regels die ze daarin gewoonlijk geacht worden te volgen, bouwen leerlingen schier onwrikbare routines en denkbeelden op over wat de bedoeling van rekenopgaven is: het vinden van het enige juiste antwoord door het rechttoe rechtaan en correct uitvoeren van de ene of andere rekenoperatie. Dit zogenaemde *didactisch contract* zit er in onze onderwijscultuur diep ingehamerd, en niet alleen bij leerlingen, ook bij leerkrachten, zo wijst het onderzoek van de auteurs uit. Kennelijk is het in het onderwijs vooral zo dat we leerlingen leren om schoolse opgaven op te lossen (zelfs al levert dit non-realistische antwoorden op!), niet om de geleerde wiskundige vaardigheden als een zinvol hulpmiddel te gebruiken bij het oplossen van betekenisvolle problemen. De ene generatie na de andere staan leerlingen dus bloot aan wat de auteurs noemen een 'verarmd en stereotiep dieet van standaard rekenvraagstukken'. En het zijn volgens hen beslist niet alleen maar de lessen en rekenboekjes zelf die dit dieet aandragen, ook de wijze van testen van leerlingen op hun rekenvaardigheden is hier waarschijnlijk voor een belangrijk deel debet aan.

Om deze hypothese van het didactisch contract nader te onderzoeken analyseren de auteurs in deel 2 van het boek eerst het tradi-

tionele rekenonderwijs op een aantal punten om na te gaan of er mogelijke oorzaken te vinden zijn voor dit typische didactische contract. In een onderzoek onder aankomende leraren konden zij bijvoorbeeld aantonen dat het genoemde didactische contract al stevig in hun kijk op rekenonderwijs (en dus ook in hun beoordelingen van leerlingantwoorden) verankerd is. In een vervolg op die studie voeren Verschaffel & De Corte twee interventiestudies uit met de bedoeling de aard van het didactisch contract daadwerkelijk te veranderen. Met name de tweede studie is grondig, zorgvuldig en voorbeeldig: de studie ging uit van realistische opdrachten, gaf de leerkrachten een uitvoerige begeleiding en verschaft de leerlingen een heuristisch systeem voor het aanpakken van de opgaven. Bovendien werd in de lessen een klimaat gecreëerd waarin de leerlingen via gezamenlijk discussie (over mogelijke oplossingen maar ook over de regels van het mathematische spel) tot een gezamenlijke oplossing konden komen. Het onderzoek behelsde 20 lessen in het aanpakken van rekenvraagstukken in 4 experimentele klassen naast 7 controle klassen. Een grootse opzet dus. Het voert te ver alle resultaten hier te bespreken, sommigen meer overtuigend dan andere. In ieder geval mag gezegd worden dat de onderzoekers verschillen vinden tussen de experimentele en de controle groepen die erop wijzen dat de gekozen onderwijsaanpak een positief en duurzaam effect heeft op de manier waarop leerlingen 'word problems' oplossen. De auteurs noemen de resultaten "promising" (p. 107), maar ze wijzen er meteen ook op dat "some of the findings warn against over-optimistic conclusions". In hoeverre de hypothese van het didactisch contract hiermee ook bevestigd is, blijft onduidelijk, omdat het niet goed zichtbaar is geworden of er inderdaad een nieuwe klassencultuur is ontstaan (de auteurs wijzen daar zelf ook op, p. 107). Niettemin draagt het onderzoek bij aan de verdere versteviging van het idee dat een radikaal andere aanpak van 'word problems', waarbij meer van levensechte problemen is uitgegaan en waarin discussie mogelijk is over alternatieve aanpakken, tot positieve resultaten leidt. Zoals de auteurs ook laten zien wijzen deze resultaten in dezelfde richting als die verkre-

gen naar aanleiding van de bekende Jasper series (Cognition and Technology Group Vanderbilt).

In het derde deel van het boek proberen de auteurs in theoretische zin de balans op te maken van het onderzoeksprogramma. Met het oog daarop gaan zij in op 'word problems' vanuit een perspectief van 'modeling'. Zij beweren daarin dat het oplossen van 'word problems' gezien kan worden als 'exercises in mathematical modeling' (p. 138). Hier blijkt echter dat de theoretische oriëntatie van de auteurs toch niet krachtig genoeg is om de opbrengsten van hun onderzoeken overtuigend tot een theoretisch geheel te smeden. In de uitleg van het 'modeling'-concept blijft dit begrip toch te veel een molair geheel dat niet microgenetisch geanalyseerd is. De relatie met representeren bijvoorbeeld is hier essentieel, maar deze blijft onuitgewerkt. De auteurs presenteren verder bijvoorbeeld (p. 168) een schema met een 'elaborated view of the modeling process', maar merkwaardig genoeg is 'model' een onderdeel van dit proces. Het lijkt er op dat datgene wat verklaard moet worden, zelf al een onderdeel van de verklaring is. Op meerdere punten in het boek laat nauwlettende lezing zien, dat het modeling-begrip toch niet helder gedefinieerd is.

Opmerkelijk in hun analyse is voorts dat nauwelijks een conceptueel verband wordt gelegd tussen modeling en de eerder besproken interessante hypothese van het didactisch contract. Het interessante van het didactische contract was dat dit een mogelijkheid biedt om de invloed van de cultuurhistorie op de ontwikkeling van de activiteiten in de rekenlessen te begrijpen. Dat element is nu jammerlijk verloren gegaan in deze uitleg van modeling die toch eerder als een individuele activiteit (in een sociale context weliswaar) lijkt te worden gezien (als we bijvoorbeeld naar de schema's 7.2 en 9.1 kijken), dan een sociale activiteit waarin de ander (c.q. de cultuur) vanaf het begin *in* het proces van modeling meedoet. Het gaat te ver deze discussie hier te voeren, maar gezegd moet worden dat het didactisch contract en modeling hier toch te veel als losse begrippen naast elkaar blijven. Men kan zich zelfs afvragen of het onderwijs zoals in het experimentele pro-

gramma ontwikkeld is, niet gewoon een vervanging van het klassieke didactische contract is voor een 'modeling'-contract: waar vroeger gecijferd moest worden, moet nu 'gemodeld' worden. Als dat zo is, is het de vraag wat een leerling hier mee op schiet en of dit wel een stevige basis is om de 'sense-making' voor leerlingen in het onderwijs terug te brengen.

Wat de uitkomst van deze discussies ook mogen zijn, het boek van Verschaffel, Greer en De Corte is er een waar de onderwijspsychologie als discipline trots op mag zijn. Het brengt niet alleen via zorgvuldige analyse en redenering enkele onderwijsproblemen dicht bij een oplossing, maar genereert tevens nieuwe problemen. Zij laten daarmee overtuigend zien dat 'word problems' niet alleen voor leerlingen en leerkrachten een nieuwe betekenis moeten krijgen, maar ook voor onderzoekers. Welke dat is en vooral in welke theoretische taal die het beste beschreven kan worden, is nog een open probleem voor toekomstige discussie.

*Bert van Oers*  
*Afdeling Onderwijspedagogiek*  
*Vrije Universiteit*