

Boekbespreking

Prof. dr. W.K.B. Hofstee
(w.k.b.hofstee@ppsw.rug.nl)
is verbonden aan het Heymans
Instituut, Rijksuniversiteit
Groningen.

Boekbespreking 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.***
- **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.***

Afgelopen zomer laaide in Vlaamse en Nederlandse kranten een discussie op naar aanleiding van de herijking van de Wechsler Intelligence Scale for Children. Op vragen als 'wat is een eiland' gaven Vlaamse kinderen ofwel het goede antwoord ofwel geen antwoord, maar de Nederlandse kinderen kletsten een eind weg en kregen vaak halve credit voor de klokjes die ze hadden horen luiden. Commentatoren brachten dat in verband met verschillen in volksaard in het algemeen en in schoolsystemen in het bijzonder, doorgaans ten nadele van Nederland. (Mijn sympathieën liggen omgekeerd: naar mijn overtuiging bestaat kennis niet uit files in een of ander random access memory, maar uit halfwaarheden die bij gelegenheid, in gesprek met andere mensen tot bloei kunnen komen, of zoiets).

Tegen de achtergrond van zulke stereotypen is het opmerkelijk dat de meest radicale van deze twee teksten, die van Dochy e.a. (verder aan te duiden met *Assessment*) een zwaartepunt in Vlaanderen heeft: qua auteursamenstelling betreft het hier vooral een Leuvens-Maastrichtse productie. Bij Van Berkel en Bax (verder aan te duiden als *Toetsen*) liggen de zwaartepunten in Amsterdam en, opnieuw, Maastricht. *Toetsen* bevat, zij het telkens in korter bestek, zowel de 'klassieke' als de 'moderne' didaxometrie. Het boek wordt gepresenteerd als uitbreiding van Van Berkel en Bax's *Beoordelen in het Onderwijs* van 1993, waarin dat laatste onderwerp nog niet werd behandeld. De teneur is eveneens verschillend: voor de auteurs van *Assessment* heeft de klassieke toets wel zo'n beetje afgedaan, terwijl in *Toetsen* over moderne vormen wordt opgemerkt: 'De meeste hiervan moeten hun status nog bewijzen'. (Je vraagt je af wat dat woord 'status' hier doet, in plaats van gewoon 'waarde', maar daar is waarschijnlijk over nagedacht, al is het resultaat taalkundig niet om over te juichen).

Wat zijn de moderne topics? De meest prominente is de OverAll-toets: M. Segers is de enige auteur met een hoofdstuk in beide teksten. De OverAll-toets (de eigenzinnige binnenhoofdletter accentueert het merksnaamkarakter) is, op de keper beschouwd, een open-boektoets bestaande uit stelling- en essayvragen. De vragen zijn uiteraard nieuw voor de tentaminandus, en de tijd om ze te beantwoorden is beperkt. Studenten hebben er problemen mee. Da's logisch, want goede voorbereiding is weliswaar een noodzakelijke, maar geen voldoende voorwaarde om slimme vragen te kunnen beantwoor-

den: je moet tevens en vooral met een behoorlijk IQ zijn geboren. De bevindingen doen me denken aan de problemen waar we in de begintijd van de multiplechoicetest tegenaan liepen: kun je ook inzicht- in plaats van alleen kennisvragen construeren. Dat lukte uitstekend, maar het bleek dat inzichtvragen onveranderlijk een abominabel lage p-waarde hadden. Exit de inzichtvraag. Ook de OverAll-toets geeft blijkbaar geen eenvoudige oplossing voor veruit het meest prangende probleem van de hedendaagse docent: hoe krijgen we de horden gammastudenten in vredesnaam zo ver dat we ze met goed fatsoen een voldoende kunnen geven.

Gemeenschappelijke onderwerpen in *Toetsen* en *Assessment* zijn voorts Portfolio's (Driessen & Van Tartwijk, resp. Janssens, Boes & Wante), en beoordeling van groepsprestaties, peer assessment en dergelijke (hoofdstukken van Van Berkel, Cluitmans en Bax & Perrenet, respectievelijk Sluismans & Dierick). Wat die laatste onderwerpen betreft: uw recensent is geen liefhebber van beoordeling van collectieve prestaties of van bijdragen aan het groepsproces, en van (mede)beoordeling door onbevoegden. Maar wie deze weerstanden niet deelt, kan in beide teksten een heel scala aan overwegingen en handgrepen aantreffen.

Gemeenschappelijk op een abstracter niveau is een theorema dat momenteel in vakkringen vrijwel algemeen aanvaard lijkt, en bij alle auteurs in een of andere variant is terug te vinden. Het luidt dat de beoordeling tot doel heeft het leerproces te bevorderen, of misschien liever gezegd zo weinig mogelijk te verstoren, en dat die beoordeling daarom 'congruent' (*Toetsen*, p. 18) moet zijn aan de onderwijsdoelstelling. Bijvoorbeeld, als je onderwijs gericht is op interdisciplinaire en interpersoonlijke probleemoplossende competenties, dan ligt het niet voor de hand studenten uitsluitend met gesloten antwoordtoetsen op monodisciplinaire feitenkennis te gaan toetsen, want dat leidt hun aandacht af.

In dit opzicht is er wel een accentverschil tussen *Assessment* en *Toetsen*, ten gunste van de laatste. De tekst van Dochy e.a. is overwegend programmatisch, soms op het ronkende af. Er wordt gesproken over voorbijgaande tijdperken en paradigm shifts; Dierick e.a. stellen 'edumetrie' tegenover psychometrie, alsof er voor die laatste een alternatief zou zijn. De stijl van *Assessment* is er een van 'authentieke methoden meten en stimuleren de beoogde competenties, en klassieke toetsen doen dat niet'. Wat daarbij wordt vergeten, is dat dit nog maar de vraag is, namelijk, een empirische vraag. De Eerste Hoofdwet van de onderwijskunde is dat analyses tekort plegen te schieten. Je kunt een maatregel, bijvoorbeeld invoeren van portfolio's, nog zo mooi beredeneren, maar daarmee is niet bewezen dat hij in de praktijk beter werkt. De reden is dat wij de werkelijkheid niet plegen te kunnen overzien – met name niet het onvermoeibaar vernuft waarmee studenten zelf zich onder onze goede bedoelingen weten uit te draaien (bijvoorbeeld door teksten van Internet af te plukken). De stijl van *Toetsen* is hier in doorsnee terughoudender; men spreekt bijvoorbeeld over 'meetpretenties', wat al de broodnodige relativering in zich herbergt.

Het punt is niet zozeer wat deze recensent ervan vindt. Beide teksten fungeren naar ik aanneem voornamelijk als cursusboek voor docenten, met auteurs in de rol van cursus-

leider. Voor zover de cursisten zelf een wetenschappelijke opleiding hebben, zouden ze wel eens behoorlijk kunnen afknappen op de reclametaal die met name in *Assessment* regelmatig de kop opsteekt, en ten onrechte worden bevestigd in hun vooroordeel dat de onderwijskunde geen wetenschap is. De kwaliteit van het onderwijs zou daarmee niet zijn gediend. En voor zover docenten wel door *sweeping* statements worden aangesproken, zou ik ze bij voorbaat liever niet op leerlingen en studenten losgelaten willen zien.

Rest nog de vraag waarom in *Toetsen* wel aandacht aan klassieke toetsvormen wordt besteed, en zelfs het woord ('toetsen') wordt gehandhaafd dat vanaf de tijd van A.D. de Groot onverbreekelijk met 'objectief scorebaar' leek te zijn verbonden. Immers, Van Berkel en Bax onderschrijven het congruentietheorema, en zulke voorstanders van stampwerk zullen ze nu ook weer niet zijn. Het antwoord ligt in het voorafgaande besloten: je kunt met keuzevragen heel wel hogere denkprocessen in het vizier krijgen. Het hoofdstuk van Van Berkel bevat intelligente en praktische aanwijzingen daarvoor. Wel rijst hier een speciaal probleem, vergelijkbaar met dat van 'sociale onveiligheid' (een wijk of weg kan nog zo veilig zijn, maar dat verhindert mensen met een bepaalde persoonlijkheidsstructuur niet zich lekker onveilig te voelen of althans actie tegen 'de overheid' te voeren): er zijn wel aanwijzingen dat ook intelligente keuzetoetsen 'sociaal incongruent' zijn, dus althans bij sommige studenten de indruk wekken dat er alleen voor hoeft te worden gestampt. Aan dat perceptieprobleem wordt in *Toetsen* weinig of geen aandacht geschonken. Misschien moet dat ook niet: ik althans zou er (in beide gevallen) de voorkeur aan geven zulke mensen in hun vet te laten gaar smoren: misconcepties blijven doorgaans in een behoefte te voorzien.

Het lukt in dit bestek niet om aan alle auteurs recht te doen, en evenmin om detailkritiek te leveren. Ik wil tot slot nog wel een wat filosofische opmerking kwijt. Voor veel van de onderhavige auteurs zijn toetsen en beoordelingen hun dagelijkse handel en winkel. Dat stimuleert de gedachte dat die dingen belangrijk zijn, en die gedachte komt in beide teksten om de haverklap terug. Ik wou in de eerste plaats dat uitgangspunt in zijn algemeenheid ter discussie stellen. Mijn ideaal als onderwijzer is beoordelingsvrij onderwijs: de beoordelingsrelatie is ten principale een storend element in de onderwijsrelatie en het leerproces. Het is mij bekend dat dit ideaal in het massaonderwijs niet haalbaar is. Maar een hogere status voor beoordelingen dan die van noodzakelijk kwaad kan er bij mij niet af. Bij wijze van compromis propageer ik dan ook een marginale opvatting van beoordelingen: beste leerlingen en studenten, het moet nu eenmaal, want jullie weten ook wel dat je geen donder uitvoert zonder stok achter de deur, maar laten we er alsjeblijft niet te serieus over doen, want je leert voor jezelf, en er is toch al extrinsieke motivatie genoeg in de wereld.

In de tweede plaats, en toegespitst op 'authentieke' beoordelingvormen: die vinden waarschijnlijk hun oorsprong in diezelfde anarchistische onderwijsfilosofie. In de praktijk fungeren ze zelfs menigmaal als excuus om geen onvoldoenden te hoeven toekennen, hetgeen weer net iets te ver gaat. Maar via een onnavolgbare list der historie zijn ze nu bezig zich juist te ontwikkelen tot een Krachtige Beoordelingsomgeving. Van het nood-

zakelijk kwaad wordt een deugd gemaakt: de student studeert voor de toets, welnu, dan zullen ze ook. Voor zover de beoordeling in zijn opzet slaagt (quod non, want die perfectie is in dit ondermaanse niet bereikbaar), wordt het leerproces er integraal door gedomineerd. Onnodig te zeggen dat dit met authenticiteit in wat voor authentieke betekenis van het woord dan ook niets meer te maken heeft. 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.