

Onderzoek beoordeeld

Door de redactie is mij gevraagd als voorzitter van de internationale onderzoeksbeoordelingscommissie pedagogische en onderwijskundige wetenschappen mijn persoonlijke indrukken te geven over het beoordelingsproces en de resultaten daarvan. Omdat mij uit reacties van faculteiten op het concept-rapport is gebleken dat er ondanks de gedegen vóór-informatie door de VSNU in het Protocol 1994 nogal wat misverstanden heersten omtrent die onderzoeksbeoordeling, lijkt het mij noodzakelijk aan mijn persoonlijke indrukken enige feitelijke informatie over het beoordelingsproces te doen voorafgaan.

Vooraf aan de werkwijze van de commissie en aan de wijze waarop de 'ratings' tot stand kwamen wil ik enige aandacht schenken, omdat daarover grote onduidelijkheid bleek te bestaan.

Met de commissie acht ik mij echter niet geroepen om op verzoek van enkele faculteiten het verschil aan te geven tussen tellingen ('scores', bijv. van productiecijfers) en intersubjectieve 'peer ratings' (bijv. van de productie). De commissie meent dit verschil bekend te mogen veronderstellen.

Wie trouwens meent dat het merendeel van de hier ten overvloede gegeven informatie in het rapport valt te lezen, moet ik daarin gelijk geven, al is het mij niet ontgaan dat niet iedereen tot de goede lezers blijkt te behoren. In een enkel geval verstrekt mijn informatie, denk ik, wat meer inzicht dan het 'final report'.

De werkwijze van de commissie was als volgt. Het eerste beoordelingsmoment was de schriftelijke beoordeling volgens een te voren vastgesteld 'assessment form' van de wetenschappelijke rapporten van de programma's der verschillende faculteiten door de leden van de commissie. Hun was door de voorzitter gevraagd de rapporten in hun geheel te bestuderen, maar in het bijzonder aandacht te schenken aan de programma's die hun als eerste of tweede reviewer door de voorzitter waren toegewezen. De voorzitter zelf trad slechts als

tweede reviewer op. Uiteraard werd bij deze toewijzing zoveel mogelijk rekening gehouden met de speciale expertise van de reviewers.

In de maartssessie van de commissie in Amersfoort speelde zich een eerste discussie af tussen de eerste en de tweede reviewers naar aanleiding van hun huiswerk.

Nadat zij overeenkomst hadden bereikt over hun ratings aan de hand van het Protocol 1994 had een plenaire discussie plaats in de gehele commissie. Na een stevige uitwisseling van standpunten en argumentatie daaromtrent leidde deze tot een door de voltallige commissie gedeelde voorlopige rating.

In de sessie van eind april, eveneens te Amersfoort, werden alle programmaleiders, al of niet van anderen vergezeld, geïnterviewd naar aanleiding van door de commissie toegezonden vragen en opmerkingen. Daarna werden door de commissie de voorlopige ratings opnieuw bezien en indien nodig bijgesteld in een soortgelijke procedure als in maart, eerst door de beide reviewers, daarna in een discussie in de voltallige commissie.

De ratings werden nu door de commissie als definitief beschouwd, eventueel nog bij te stellen na toezending van het concept-rapport aan de faculteiten en de reacties daarop. In een enkel geval heeft die bijstelling plaats gevonden.

De overige dagen werden door de commissie gebruikt om een eerste 'draft' van het rapport op te stellen en te becommentariëren.

De leden gingen naar huis met de opdracht een meer definitief concept aan te leveren, waarbij ook aan voorzitter en secretaris taken werden toegewezen.

De reacties van de faculteiten op het verzonden concept-rapport werden door de leden bestudeerd, waarna het rapport definitief werd bijgewerkt en vastgesteld. De brieven aan de faculteiten werden na overleg beantwoord.

Hoewel het concept-rapport direct na het verstrijken van de in het Protocol genoemde tijdsbestek werd toegezonden viel dit ongelukkigwijze samen met het begin van de academische vakantie, zodat aan de faculteiten tot half september de tijd moest worden gegund om te reageren. Pas daarna kon de commissie die antwoorden bestuderen.

Op enkele reacties van faculteiten wordt in het volgend deel van deze informatie of in mijn persoonlijk commentaar teruggekomen.

Omdat ondanks de beschrijving in het Protocol 1994 en in het concept-rapport misverstanden bleven bestaan omtrent de wijze waarop ratings tot stand kwamen, wil ik daarop nog summier ingaan.

Als criteria van *kwaliteit* golden voor de commissie: de key-publications, de publikaties in vooral Engelstalige, buitenlandse high standard tijdschriften en andere gerefereerde buitenlandse publikaties (boeken), de publikaties in andere, vooral Engelstalige tijdschriften, de publikaties in gerefereerde Nederlandse en eventueel Franse en Duitse tijdschriften en boeken.

Aangezien er een hoge correlatie blijkt te bestaan tussen het aantal artikelen in buitenlandse tijdschriften in het algemeen en het aantal artikelen in buitenlandse 'high impact' tijdschriften is een onderscheid tussen 'high impact' en 'low impact' tijdschriften, op de manier waarop dat door de commissie psychologie is gehanteerd, niet toegepast.

Verder werden in de kwaliteitsbeoordeling betrokken: (inter)nationale awards, lidmaatschappen van internationale wetenschappelijke gezelschappen, redacteurschappen van internationale en van gerefereerde Nederlandsstalige tijdschriften. Een enkele programmaleider was van mening dat de commissie te veel waarde hechtte aan de internationale publikaties. De commissie vindt dat het internationale karakter van wetenschapsbeoefening dat rechtvaardigt.

Wetenschapsparadigma's speelden alleen een rol wanneer de commissie die als gedateerd beschouwde.

Eén faculteit maakte daar in haar reactie bezwaar tegen, omdat zij dit geen aspect van kwaliteit van publikaties achtte. De commissie had daarvoor wel een zeker begrip om een zo dadelijk uiteen te zetten reden, maar zij hield toch vast aan haar standpunt. Als aspect van wetenschappelijke kwaliteit had zij immers als eerste punt genoemd "the originality of approaches and ideas in tackling scientific problems", evenals andere commissies, bijv. die van "Electric Engineering" dat ook hebben gedaan. Tot die 'original approaches' rekende zij ook het paradigma. Het 'zekere begrip' van de commissie sproot voort uit de gedachte dat men met een zekere relativering over deze kwestie behoort te denken. Want al is bijv. in de onder-

wijskunde het behavioristische paradigma volstrekt gedateerd, het thans in zwang zijnde constructivistisch paradigma is in wezen reeds aan te wijzen in de 'discovery learning'-benadering van Piaget en Bruner, maar toch kennelijk niet gedateerd geacht. Die relativering leidde tot opname van een appendix van betrokken faculteit.

De beoordeling van de *productiviteit* heeft volgens de commissie een kwalitatieve en een kwantitatieve, numerieke kant. De commissie sprak dus niet louter op basis van numerieke gegevens een oordeel uit. Om die reden lette zij ook minder op het volume dan op de kwaliteit van het produkt, al werd een artikel van 2 of 3 pagina's veelal niet hoog gewaardeerd. De kwaliteit van de produktie beoordeelde de commissie op basis van de '1- and 2-rule', een internationaal erkende regel, waarin het cijfer 1 aangeeft het aantal publikaties in gerefereerde internationale tijdschriften of handboeken, en het cijfer 2 het aantal andere internationale publikaties per fte per jaar. In deze zin is toch uitdrukking gegeven aan het onderscheid tussen high impact publikaties en andere.

De kwantitatieve wetenschappelijke output (scoring), vooral in de internationale pers speelde eveneens een rol. Vakpublicaties werden niet in de beoordeling betrokken. Uiteraard heeft de commissie niet van elke publikatie aangegeven of zij dit een wetenschappelijke publikatie of een vakpublicatie vond. Dat is ook volstrekt onnodig na wat hiervoor over haar beoordeling van publikaties is opgemerkt. Zij heeft wel voorbeelden gegeven van wat zij als vakpublicaties beschouwde. Als bij wijze van zulk een voorbeeld artikelen in *Onderwijskundig Lexicon*, etc. werden opgevoerd als wetenschappelijke publikaties heeft de commissie die niet als zodanig gewaardeerd. Het voorwoord van dit lexicon, editie II geeft zelf duidelijk aan dat het daar om vakpublicaties gaat.

Het *relevantie*-aspect betrof de bijdrage aan de uitbreiding van wetenschappelijke kennis, maar ook het belang van het programma voor de praktijk van opvoeding en onderwijs en voor het beleid.

Het oordeel over de *viabiliteit* betrof de toekomstige taak die voor het programma gezien werd, uitgaande van de ontwikkeling in de tijd en rekening houdend met de persoonlijke en materiële middelen die ter beschikking stonden.

Kom ik nu tot mijn *persoonlijk commentaar*, dan wijs ik eerst op het door mij duidelijk ervaren belang van een evaluatie van onze universitaire onderzoeksprogramma's door een internationale commissie, die ons deels nog wat provincialistisch pedagogische wereldje een spiegel voorhoudt. Dat het niet voor iedereen prettig is het eigen spiegelbeeld te zien, moet men denk ik voor lief nemen, al mag men best op onderdelen met de commissie van mening verschillen.

Wat de beoordeling voor de commissie niet eenvoudig maakte is de grote diversiteit van subdisciplines: wijsgerige en historische pedagogiek, empirisch pedagogische studies, studies in jeugdkunde en in volwasseneneducatie, orthopedagogiek met een brede mix van leeren gedragsstoornissen van kinderen en jeugdigen, onderwijskunde op micro-, meso- en macroniveau, onderwijstechnologie, statistische en mathematische methodieken, voor welke laatste de commissie externe adviseurs om een oordeel heeft gevraagd.

De discussie over de programma's in al die subdisciplines werd wel in de gehele commissie gevoerd, maar de experts hadden uiteraard een belangrijke invloed in die discussie.

Een punt van zorg lijkt mij de verhouding tussen de fundamentele research en de theoretische basis daarvan en de toegepaste research, welke laatste nogal oververtegenwoordigd was. Dat is wellicht begrijpelijk in wetenschappen die zo sterk op het handelen betrokken zijn als pedagogiek en onderwijskunde, maar dat neemt niet weg dat de basis van fundamentele research ook in deze wetenschappen onmisbaar is en m.i. meer nadruk moet krijgen. In de internationale vergelijking draagt het Nederlands onderzoek wat sterk een toegepast karakter. Was er daarom zo weinig excellent onderzoek te bespeuren?

Wellicht vallen de oordelen van de commissie mede daardoor wat lager uit dan menig onderzoeker blijkens de reacties gehoopt of verwacht had. Het Nederlands onderzoek is volgens internationale maatstaven in het algemeen van een bevredigend niveau ('*satisfactory*'), maar niet meer dan dat.

Een enkele faculteit wist dat zo moeilijk te verwerken dat zij het rekenwerk van de commissie volgens een andere methode dan die van de commissie heeft overgedaan en daardoor tot

andere conclusies is gekomen dan de internationale commissie. Deze heeft van die bewering met belangstelling kennis genomen zonder zich echter van de eigen koers te laten afleiden.

Er waren zelfs onderzoekers die zich, buiten medeweten van de voorzitter van de commissie, hebben gewend tot hun bekende individuele leden van de commissie om die tot andere gedachten te brengen, hetgeen de commissie uiteraard niet bijzonder apprecieerde.

Weer andere researchers vergeleken hun scores en ratings met die van andere universiteiten en suggereerden dat zij in vergelijking met die anderen onrechtvaardig behandeld waren. Uit deze voorbeelden blijkt duidelijk, hoe belangrijk men een evaluatie als deze vindt en hoe moeilijk men het vindt de uitkomsten daarvan te accepteren als die ongunstiger uitvallen dan men verwacht had.

Dat zij groot belang hechtten aan het oordeel van de commissie blijkt ook uit het feit dat onderzoekers van een der universiteiten niet konden wachten op het definitieve rapport, alvorens zichzelf op de borst te kloppen, en hebben gelekt naar de eigen universitaire pers die hun succes voortijdig publiceerde.

De gegeven voorbeelden vormen uitzonderingen. De meeste faculteiten stelden zich, na op feitelijke onjuistheden in het concept-rapport gewezen te hebben, positief op ten aanzien van het werk van de commissie, ook wanneer de oordelen voor de eigen faculteit minder gunstig waren. Het rapport geeft immers duidelijk aan waar de sterkten en de zwakten van programma's liggen en doet in diverse gevallen suggesties voor verbetering die de viabiliteit van de programma's kunnen versterken en de kwaliteit kunnen verbeteren. Minder in de breedte, meer zich beperken tot in de diepte werken is meermalen het advies.

Ook is aansluiting bij onderzoeksscholen en versterking van de invloed daarvan een advies dat werd verstrekt.

Wat zou de VSNU volgens mijn perceptie kunnen doen ter verbetering van de procedure?

Aan het voorbereidend werk van de VSNU (zie het Protocol 1994) en de ondersteuning van het werk van de commissie kan niet veel worden verbeterd. (Dat het concept-rapport

precies volgens de protocollaire regels bij de faculteiten arriveerde op het moment dat de academische vakantie begon, is in het vervolg door planning te voorkomen).

Dat onze commissie het beter vond met een 10-punts dan met een 5-punts ratingschaal te werken is een kleinigheid en wordt wellicht ook door andere commissies niet gedeeld.

Een andere kleinigheid is dat het voor de commissie gemakkelijk zou zijn, en door de VSNU bevorderd kan worden, dat alle faculteiten internationale en nationale publikaties in aparte lijsten zouden vermelden.

Overweging verdient ook om aio's niet meer mee te tellen bij het wp, daar de meeste aio's niet vóór hun promotie publiceren. Sommige faculteiten hebben daar terecht op aangedrongen.

De verbetering van de procedure moet echter vooral van andere zijden komen. Zo is mij van verscheidene kanten verteld dat leden van de commissie onbekenden voor betrokkenen waren. Nu kan dat het gevolg zijn van het gering aantal internationale contacten van een aantal onderzoekers. De curricula vitae van de gewone leden laten zien dat het om internationaal bekende persoonlijkheden gaat.

Toch zou de vraag gesteld kunnen worden of het samenspel tussen de disciplineoverlegorganen en de KNAW van zodanige aard is geweest dat die organen zich voldoende in de keuze van de KNAW konden herkennen.

Van groter belang is een tweede punt: er valt echt wat te doen van de zijde van de faculteiten. Wie de divergentie van de oordelen van de commissie over verschillende programma's binnen vrijwel elk van de faculteiten overziet, kan niet anders dan tot de conclusie komen dat een aantal programma's niet ter beoordeling ingediend had moeten worden. De faculteiten hebben blijkbaar geen voldoende invloed om op de rem te trappen als zo'n zwak programma toch voor beoordeling wordt voorgedragen.

Nu er onderzoeksscholen bestaan in Nederland, waaraan de universiteiten, de minister, maar ook de commissie grote waarde hechten, zou het dan niet beter zijn dat door die onderzoeksscholen een vóórselectie werd gemaakt voor de indiening van internationaal te beoordelen programma's? Een beroepscommissie kan dan eventueel tevoren een uitspraak doen

als faculteiten problemen hebben met de keuze van de onderzoeksscholen. Maar de internationale beoordeling moet daaraan volgend wel blijven bestaan.

Een laatste persoonlijke ervaring, maar die ook door de meeste leden van de commissie wordt gedeeld is dat het werken in dit internationaal verband inspirerend, aangenaam, maar volgens de aangegeven procedure ook zeer vermoeiend is. Wellicht moet men in het vervolg een jongere, in het buitenland werkende voorzitter voor zulk een commissie benoemen.

A. M. P. Knoers