De kwaliteit van de medische opleidingen getoetst?

A.J.J.A. Scherpbier, R.P. Zwierstra

Samenvatting

Inleiding: Informatie over opleidingen, zoals de resultaten van het Elsevier/NIPO-onderzoek, moet betrouwbaar en valide zijn om studenten echt te helpen bij het kiezen van een opleiding. In dit artikel worden de Elsevier/NIPO-enquêtes over geneeskunde uit 1994-2001 onder de loep genomen.

Gegevensverzameling studenten: Een representatieve steekproef van 6000 studenten van 22 opleidingen is geënkquêteerd. Dat betekent ongeveer 34 studenten per geneeskundefaculteit. Aan de studenten zijn 32 stellingen voorgelegd. Geen hiervan betreft de co-assistent-schappen en het vaardigheidsonderwijs.

Gegevensanalyse studenten: De gemiddelde studentenoordelen over de faculteiten verschillen weinig (6.9-7.3). Tussen 1994 en 2001 stijgen de gemiddelde scores en vermindert de spreiding. Er is een lichte aanwijzing voor een samenhang tussen de hogere beoordelingen en recente curriculumherzieningen.

Gegevensverzameling hoogleraren: De hoogleraren werd gevraagd welke universiteit – niet hun eigen – de beste geneeskundeopleiding biedt. De respons is laag.

Gegevensanalyse hoogleraren: Groningen staat op de eerste en Maastricht op de tweede plaats; de rangorde van de overige faculteiten varieert. De oordelen zijn gebaseerd op wetenschappelijke publicaties, kennis van de andere opleidingen en studentenoordelen.

Discussie: Gezien de kleine steekproeven is er twijfel mogelijk over de betrouwbaarheid en de gevonden significantie. Het hooglarenoordeel lijkt weinig valide. De meesten weten vrij weinig van de andere opleidingen en wetenschappelijke kwaliteit is iets anders dan onderwijskwaliteit. De studentenoordelen moeten voorzichtig geïnterpreteerd worden. De verschillen zijn gering en een belangrijk deel van de opleiding, de co-assistent-schappen, is niet beoordeeld. Voor kwaliteitsverbetering van het medisch onderwijs zijn naast tevredenheids-onderzoek, zoals de Elsevier/NIPO-enquête, andere systemen voor kwaliteitszorg nodig, zoals externe en interne visitatie en accreditatie. (Scherpbier AJJA, Zwierstra RP. De kwaliteit van de medische opleidingen getoetst? Tijdschrift voor Medisch Onderwijs 2002; 21(1):2-8.)

Inleiding

De kwaliteit van het hoger onderwijs wordt de afgelopen jaren regelmatig onderzocht. Er is een systeem van visitaties waarbij een panel van deskundigen de zelfstudie bestudeert, de opleiding bezoekt en met docenten en studenten spreekt.^{1 2} De opleiding maakt van tevoren een zelfstudie op basis van een vragenlijst van de visitatiecommissie. In de toekomst wordt het systeem van

visitaties vervangen door of onderdeel van een systeem van accreditatie.³ Ook wordt sinds 1994 door Elsevier/NIPO een onderzoek uitgevoerd waarvan de resultaten worden gepubliceerd in de Elsevier.⁴⁻⁹ Tot slot is er ook nog een 'Keuzegids hoger onderwijs'.¹⁰ Het Elsevier/NIPO-onderzoek en de keuzegids zijn bedoeld om vwo-leerlingen en andere toekomstige studenten te helpen bij het kiezen van een opleiding.

Wij vroegen ons af hoe betrouwbaar en valide de gegevens van de Elsevier/NIPO-enquête zijn. In dit artikel beperken we ons tot de informatie over geneeskunde. Om de vraag over de betrouwbaarheid te beantwoorden wordt beschreven hoe de gegevens van studenten en hogleraren verzameld zijn. De vraag over de validiteit wordt beantwoord door de gegevens over de afgelopen jaren te analyseren.

Gegevensverzameling over de mening van de studenten

In de enquête zijn 6000 studenten vanaf het tweede studiejaar in 2001 telefonisch ondervraagd. Er werden 22 opleidingen beoordeeld. Er werden geen gegevens verstrekt over het aantal studenten dat niet mee heeft gewerkt. Stel dat iedereen heeft meegewerkt en dat de aantallen studenten gelijk verdeeld zijn over de opleidingen, dan zijn er 272 studenten per opleiding benaderd. Als dit aantal gelijk verdeeld is over de 8 faculteiten geneeskunde, betekent dat 34 studenten per faculteit. Als alleen studenten in jaar 2, 3 en 4 zijn benaderd en de verdeling over de jaren gelijk was, is dat 11 studenten per studiejaar. Als er ook studenten uit jaar 5 en 6 zijn benaderd, dan gaat het om ongeveer 7 studenten per studiejaar. Dat is een (te) kleine steekproef om conclusies te kunnen trekken. De Informatie Beheergroep (IBG) leverde de representatieve steekproef voor het onderzoek. De IBG beschikt niet over gegevens van toetsresultaten, hetgeen betekent dat er niet geselecteerd kon worden op de representativiteit van de respondenten qua studieprestaties binnen de studie-jaren. Als de steekproef groot genoeg zou zijn geweest, was dat natuurlijk geen probleem. Er zijn dus twijfels over de betrouwbaarheid van de informatie.

De studenten kregen 32 stellingen voor-gelegd, waarover ze op een tienpunts-

schaal een oordeel moesten geven. De betekenis van de tienpuntschaal varieerde. Bij sommige stellingen liep de schaal van “zeer onvoldoende tot uitstekend”, bij andere van “volledig oneens tot volledig eens” of van “zeer slecht tot zeer goed”. Een aantal stellingen is samengevoegd. Het is opvallend dat er geen vragen over de co-assistentenschappen zijn gesteld, terwijl die toch ongeveer 30% van het totale curriculum uitmaken. Ook over het vaardigheidsonderwijs en het praktijkonderwijs in de eerste jaren van de studie worden geen vragen gesteld. De validiteit van de informatie wordt hierdoor bedreigd.

Analyses van de gegevens van de studenten

In tabel 1 is het gemiddelde aangegeven per onderdeel van de enquête en het gemiddelde per universiteit. Maastricht is dit jaar eerste, Groningen en Nijmegen delen de tweede plaats.

Opvallend is dat de gemiddelde cijfers niet zo heel ver uit elkaar liggen (6.9-7.3). In de originele tabel in Elsevier zijn de verschillen op 95%-niveau aangegeven voor alle items en de totalen. Het is niet duidelijk of er gecorrigeerd is voor de kans op toevallige significantie bij veel significantieberekeningen (bijvoorbeeld via de Bonferroni-correctie). Daardoor is er twijfel over de betrouwbaarheid van de uitspraken over de verschillen tussen de faculteiten. Het kan immers ook toeval zijn. Gezien de geringe verschillen tussen de cijfers moet er wel een redelijke respons zijn geweest – en wellicht ook grotere aantallen dan in het bovenstaande wordt verondersteld – om überhaupt significante verschillen te kunnen vinden, los van de vraag of de gevonden significantie relevant is. Per onderdeel van de enquête is in tabel 1 weergegeven of de beoordeling van de faculteit positief of negatief

Tabel 1. Waardering van studenten voor het onderwijs uit de Elsevier/NIPO-enquête.⁹ Gemiddelde score per rubriek en per faculteit. Als de score significant verschilt van de gemiddelde score is het cijfer vetgedrukt en is de richting van het verschil aangegeven met ▲ of ▼. In de kolom Totaal staat de gemiddelde score per rubriek voor de acht faculteiten samen. In de rij Totaal staan de gemiddelde scores per faculteit voor de zes rubrieken samen.

	Leiden	RUG	UU	EUR	UM	UvA	VU	KUN	Totaal
Faciliteiten	7.3 ▼	7.5	7.5	7.5	7.9 ▲	7.4 ▼	7.6	7.7 ▲	7.6
Inrichting opleiding	6.8	6.8	6.7	6.5 ▼	7.5 ▲	6.8	6.7	7.1 ▲	6.9
Docenten	7.2	7.5 ▲	7.0	7.0 ▼	7.2	7.2	7.1	7.4 ▲	7.2
Onderwijs	6.9	7.6 ▲	7.0	6.7 ▼	7.3 ▲	6.8 ▼	6.8 ▼	7.3 ▲	7.0
Toetsing	7.1	7.2	7.0	7.0	7.1	7.2	7.2	7.5 ▲	7.2
Organisatie/communicatie	6.2 ▼	6.7	6.5	6.4	7.3 ▲	6.4	6.5	6.4	6.6
Totaal	6.9 ▼	7.3 ▲	7.0	6.9 ▼	7.4 ▲	7.0	7.0	7.3 ▲	7.1
Rangorde	4	2	3	4	1	3	3	2	

verschilt van het gemiddelde per rubriek. Opvallend is dat in Nijmegen op vijf onderdelen van de enquête een significant positief effect wordt gezien ten opzichte van het gemiddelde. Daarna volgt Maastricht met vier en Groningen met twee. Als het aantal significant positieve verschillen geteld was, kwam Nijmegen op de eerste plaats. Omdat ervoor gekozen is de resultaten op alle stellingen te middelen, wat op zich ook betrouwbaarder is, komt er een andere rangorde uit. In figuur 1 zijn de gemiddelde totale beoordelingen uitgezet van alle jaren dat de enquête is afgenomen.⁴⁻⁹

Opvallend is dat de spreiding in 1994 veel groter was dan in 2001. De verleiding is groot om dit te verklaren door alle acti-

viteiten die er in het medisch onderwijs plaatsvinden. Als gekeken wordt naar andere studierichtingen blijkt dat er vergelijkbare fenomenen zijn. Bij biologie varieerden de oordelen in 1994 tussen 6.7 en 7.2 en in 2001 tussen 7.1 en 7.3. Bij psychologie varieerden de oordelen in 1994 tussen 6.6 en 6.8. en in 2001 tussen 6.8 en 7.1. Het lijkt erop dat de genoemde opleidingen ook beter zijn geworden of dat de studenten in de loop der jaren milder zijn geworden in hun oordeel. Bij de vergelijkingen tussen 1994 en 2001 valt wel op dat het gemiddelde rendement bij geneeskunde duidelijk gestegen is. In 2001 is het gemiddelde rendement van het artsexamen (6 jaar opleiding) zelfs hoger dan van vierjarige opleidingen na zes jaar. Een fac-

Figuur 1. Elsevier/NIPO-enquete 1994-2001 per faculteit.⁴⁻⁹

tor die invloed gehad zou kunnen hebben op de kwaliteit van de opleidingen tussen 1994 en 2001, de visitaties, moet ook genoemd worden

De afgelopen jaren hebben er ook curriculumherzieningen bij de faculteiten geneeskunde plaatsgevonden. Het zou logisch zijn dat deze beoogde verbeteringen in figuur 1 terug te vinden zijn. De validiteit van de metingen zou hierdoor ondersteund worden. Daarbij is het van belang om nog eens te vermelden dat studenten vanaf het tweede jaar worden geënquêteerd. Het duurt dus na de start van de curriculumherziening een paar jaar voordat een eventueel effect zichtbaar wordt. In Maastricht hebben tot 2001 geen grote wijzigingen plaatsgevonden. De beoordelingen zijn stabiel. In Groningen is in 1993 gestart met een nieuw curriculum. Vanaf 1996 is er een lichte stijging van de beoordelingen. Bij de UvA en de VU heeft de curriculumherziening plaatsgevonden voor de eerste meting. De beoordelingen van de UvA vertonen een stabiel beeld. Bij de VU is er een lichte stijging te zien de afgelopen jaren. Dit zou veroorzaakt kun-

nen zijn doordat een aantal aanpassingen van het curriculum aan de VU wat later hebben plaatsgevonden. In Rotterdam zijn over de jaren de beoordelingen stabiel. Recent is daar gestart met een herziening. De curriculumherziening in Nijmegen is gestart in 1995/1996 en vanaf 1999 is er een heel duidelijke stijging. De herzieningen in Utrecht en Leiden zijn een paar jaar geleden gestart. Er is een lichte stijging bij beide faculteiten, hoewel de invoering van de herziening eigenlijk te kort geleden is om al veel effect te zien. Kortom, er zijn lichte aanwijzingen dat de enquête gevoelig is voor veranderingen.

Gegevensverzameling hoogleraren

De vraag: “welke universiteit biedt in Nederland de beste opleiding op een vakgebied” staat centraal in de schriftelijke enquête onder de hoogleraren (N=3600).⁹ De hoogleraren mochten de opleiding aan hun eigen universiteit niet beoordelen. Na het versturen van de enquête zijn degenen die de enquête niet terugstuurden nog eens per e-mail benaderd. De respons was 37%. Hiervan was 22% afkomstig uit de

sector 'gezondheid'. Hoeveel hoogleraren een oordeel over de geneeskundeopleidingen hebben gegeven, wordt niet exact vermeld. Het wordt niet duidelijk of er een analyse van de non-responsgroep is gedaan. In ieder geval betekent een zo lage respons in het algemeen dat de gegevens niet betrouwbaar genoeg zijn om te publiceren. De enquête bevatte ook vragen over de bronnen waarop de hoogleraren hun oordeel baseren, hun functie, nevenwerkzaamheden, et cetera.

Analyse van de gegevens van de hoogleraren

In figuur 2 is de rangorde over de vier jaar dat de enquête onder de hoogleraren is afgenomen weergegeven.⁶⁻⁹ Groningen staat steeds op de eerste plaats en Maastricht op de tweede plaats. De rangorde van de andere faculteiten wisselt over de jaren. De oordelen van de hoogleraren zijn in 2001 voor 77% gebaseerd op de kwaliteit van de wetenschappelijke publicaties, voor 74% op feitelijke kennis van de opleidingen en slechts voor 38% op het oordeel van studenten over het onderwijs.⁹ In een grote faculteit, zoals geneeskunde, met veel hoogleraren is het de

vraag in hoeverre hoogleraren een goed beeld hebben van de eigen opleiding. Een goed beeld hebben van andere opleidingen is nog veel moeilijker. Dat het oordeel voor 74% gebaseerd wordt op feitelijke kennis van de opleiding, is een zeer opmerkelijk gegeven. Een verklaring zou kunnen zijn dat er een selectiebias bestaat in die zin dat alleen die hoogleraren die veel van curricula weten de enquête hebben ingevuld. Deze verklaring wordt ondermijnd door de vaker geconstateerde verschillen tussen het 'curriculum op papier' en het 'curriculum in de praktijk' en 'wat studenten echt leren'.¹¹ Docenten weten vaak niet goed hoe het curriculum echt functioneert. Om een goed beeld te krijgen van hoe het gaat, zijn dan ook studentoordelen onmisbaar. De oordelen van studenten over andere opleidingen zijn in het algemeen niet bekend. Interessant is dat de kwaliteit van de wetenschappelijke publicaties een grote rol speelt bij het oordeel. Kennelijk wordt verondersteld dat er een relatie is tussen deze kwaliteit en de opleiding. Voor een wetenschapsstage en een aio-plaats is dat wel voorstelbaar, maar voor de kwaliteit van de opleiding kunnen wij ons geen verband voorstellen.

Figuur 2. Oordeel hoogleraren.⁶⁻⁹

De validiteit van het oordeel van de hoogleraren moet gezien het bovenstaande betwijfeld worden.

Discussie

Het is goed dat aankomende studenten informatie krijgen over de sterke en zwakke kanten van de opleidingen. Hetzelfde geldt voor docenten en opleidingsbesturen. De informatie kan voor studenten een goed hulpmiddel zijn bij het kiezen tussen opleidingen. Informatie over de universiteit als geheel is daarbij voor hen niet van belang. Het gaat om de specifieke opleidingen. Om als basis te dienen voor verantwoorde keuzes moet de informatie wel voldoende betrouwbaar en valide zijn. Dat geldt ook voor docenten en opleidingsbesturen die het curriculum willen verbeteren.

Het oordeel van de hoogleraren is ons inziens door de lage respons te onbetrouwbaar en kan dan ook beter niet gepubliceerd worden. Daarnaast wordt de validiteit betwijfeld.

Het oordeel van de studenten lijkt gebaseerd op een te kleine steekproef om betrouwbaar te kunnen zijn. Er ontbreekt informatie over een groot deel van het curriculum, namelijk de co-assistenten, het vaardigheidsonderwijs en het praktijkonderwijs, hetgeen de validiteit van de oordelen vermindert. De verschillen tussen de faculteiten zijn klein. De relevantie van de gevonden significante verschillen is niet helemaal duidelijk en er is twijfel over de significantie op zich. Onze conclusie is dan ook dat het studentenoordeel vooralsnog met voorzichtigheid en relatief geïnterpreteerd moet worden. Bij het trekken van conclusies is het ook belangrijk om te benadrukken dat het onderzoek onder studenten in feite een tevredenheidsonderzoek is. De conclusie dat de geneeskundeopleiding in Maastricht 'de beste' is, zou vervangen

moeten worden door de conclusie dat de Maastrichtse studenten het 'meest tevreden' zijn over de eigen opleiding.

We moeten ervoor oppassen dat deze opiniepeilingen te veel gewicht gaan krijgen, zoals in de Verenigde Staten op dit moment het geval is.¹² De opiniepeiling van de studenten is gemakkelijk te verbeteren door de steekproef te vergroten en de vragenlijst uit te breiden. Die van de hoogleraren kan vervallen. Het is verder een kwestie van de keuze van de 'koppen' in Elsevier of de uitspraken die gedaan worden, ook gebaseerd zijn op goed onderzoek. Dit is een verantwoordelijkheid van de redactie. Om de kwaliteit van het medisch onderwijs te verbeteren zijn naast de informatie uit opiniepeilingen ook andere systemen voor kwaliteitszorg, zoals visitatie en accreditatie van belang. De visitaties kunnen ook intern worden georganiseerd, zoals dat bijvoorbeeld in Groningen met de co-assistenten in geaffilieerde ziekenhuizen gebeurt.¹³

Literatuur

1. Vereniging van Samenwerkende Nederlandse Universiteiten. Visitatierapport geneeskunde en gezondheidswetenschappen. Utrecht: VSNU; 1992.
2. VSNU Visitatiecommissie Geneeskunde en Gezondheidswetenschappen. Onderwijsvisitatie geneeskunde en gezondheidswetenschappen. Utrecht: VSNU; 1997.
3. Naar een keur voor kwaliteit. Utrecht: VSNU; september 2001.
4. Elsevier/NIPO's universiteitentest. Elsevier 25 oktober 1994; 50:28-54.
5. Elsevier/NIPO's onderzoek 1996; universiteitentest. Elsevier 19 oktober 1996; 52:32-49.
6. Elsevier special. Universiteitentest. Elsevier/NIPO onderzoek 1998. Bijlage bij Elsevier nr. 39; 26 september 1998.
7. Elsevier special. De beste studies. Speciale uitgave van Elsevier; oktober 1999.
8. Elsevier special. De beste studies. Speciale uitgave van Elsevier; oktober 2000.
9. De beste hogescholen en universiteiten; onderzoeken NIPO en IOWO 2001. Speciale uitgave van Elsevier; oktober 2001.

10. Steenkamp F, Maljaars W, Blankesteijn E, redactie. Keuzegids hoger onderwijs. Editie 2001-2002. Amsterdam: Uitgeverij Balans; 2001.
11. Coles CR, Grant JG. Curriculum evaluation in medical and health care education. *Med Educ* 1985; 19:405-22.
12. McGaghie WC, Thompson JA. America's best medical schools: a critique of the U.S. News & World report rankings. *Acad Med* 2001; 76(10):985-92.
13. Cohen-Schotanus J, Seton-Mulder RR, Zwierstra RP. Co-schappen beoordeeld. Onderwijsvisite in de geaffilierte ziekenhuizen. *Medisch Contact* 2001; 56(38):1379-81.

De auteurs:

Prof. dr. A.J.J.A. Scherpbier is wetenschappelijk directeur van het Onderwijsinstituut van de Faculteit der Geneeskunde van de Universiteit Maastricht.

Prof. dr. R.P. Zwierstra is directeur van het Onderwijsinstituut van de Faculteit der Medische Wetenschappen van de Rijksuniversiteit Groningen.

Correspondentieadres:

Prof. dr. A.J.J.A. Scherpbier, Onderwijsinstituut, Faculteit der Geneeskunde, Postbus 616, 6200 MD Maastricht, tel. 043-3881815, a.scherpbier@oifdg.unimaas.nl.

Summary

Introduction: *Information about higher education, like the annual Elsevier/NIPO survey, should be reliable and valid to be of real help to students in selecting a university. In this article we scrutinise the Elsevier/NIPO surveys conducted between 1994 through 2001.*

Data collection students: *A representative sample of 6000 students from 22 different university courses was surveyed. That is 34 students per medical school. They responded to 32 statements. There were no statements included on clinical clerkships or clinical skills training.*

Data analysis students: *The mean scores differ little among schools (6.9-7.3). Between 1994 and 2001 student satisfaction increases and the spread of the schools' scores decreases. There is some indication that curriculum revisions may have affected the rise in scores.*

Data collection professors: *The professors were asked which university – apart from their own school – offered the best medical education. The response was low.*

Data analysis professors: *Groningen and Maastricht consistently rank first and second. The ranking of the other schools varies. The professors base their judgments on scientific publications, their own knowledge about schools and student judgments.*

Discussion: *The samples surveyed are too small for reliable judgments. The value of the reported statistical significance is doubtful. The professors' judgments have only limited validity. It is unlikely that the professors have detailed knowledge of other medical curricula and scientific quality does not imply educational quality. The students' judgments should be interpreted cautiously. The differences are small and clinical training was not included in the survey. Improvement of the quality of medical education requires other quality assurance tools, besides the Elsevier/NIPO survey, such as internal and external site visits and accreditation. (Scherpbier AJJA, Zwierstra RP. The quality of medical education assessed? *Dutch Journal of Medical Education* 2002; 21(1):2-8.)*