

## 2 Een vergelijkende studie van zelf-beoordeling en docent-beoordeling bij het afnemen van de anamnese

### Samenvatting

In deze pilot-study onderzochten wij de beoordeling van door studenten afgenomen anamneses door die studenten zelf (self assessment) en door hun docenten met drie hypothesen. De eerste hypothese hield in dat de zelf-beoordelingen en de docent-beoordelingen positief met elkaar zouden correleren, waarbij de tweede hypothese was dat de zelf-beoordeling hierbij strenger zou zijn dan de beoordeling door de docent. De derde hypothese hield in dat de docent-beoordelingen positief met elkaar zouden correleren. Het bleek echter dat de correlatie tussen zelf-beoordeling en docent-beoordeling over het algemeen gering was. De docent-beoordelingen bleken ook slecht met elkaar te correleren. De studenten beoordeelden zichzelf inderdaad lager dan hun docenten deden. De studenten kwamen relatief het best overeen met Docent 1, wiens aandeel in de cursus aanzienlijk groter was dan dat van Docent 2. Wij menen hieruit te mogen concluderen dat het noodzakelijk is dat docenten vooraf trainen om tot interrater-agreement te komen, omdat zij slecht met elkaar blijken te correleren, onafhankelijk van het feit of zij nauw samenwerken. Ook denken wij uit de resultaten te kunnen opmaken, dat de studenten veel waarde hechten aan de doelstelling van hun docent, aangezien zij het best correleerden met de bij de cursus meest betrokken docent. Nader onderzoek lijkt wenselijk om uit te wijzen of een docent-beoordeling nog wel nodig is als de studenten de doelstelling van het betreffende onderwijs duidelijk krijgen uitgelegd door hun docent.

### Inleiding

Medische studenten worden beoordeeld op hun vaardigheden en kennis door docenten die, hoewel bekwaam in hun vak, geen speciale opleiding hebben gevolgd in het beoor-

delen. De vraag dringt zich dan op in hoeverre men waarde kan toekennen aan zo'n beoordeling. Wordt deze beoordeling niet erg beïnvloed door allerlei factoren, zoals persoonlijke waarden en ideeën van de beoordelaar, en is aldus het wel of niet slagen van de student afhankelijk van de toevallig toegewezen docent? Uit talloze artikelen is gebleken, dat de correlatie tussen beoordelaars zeer gering is<sup>1</sup>.

Men kan zich zelfs afvragen of een docent-beoordeling noodzakelijk is: zijn studenten niet uitstekend in staat zichzelf te beoordelen? Aan de waarde van zelf-beoordeling (self assessment) wordt vandaag de dag nog weinig aandacht besteed.

Sommige voorgaande studies wijzen uit dat studenten zichzelf goed kunnen beoordelen<sup>2 3</sup>. Wel blijkt dat zij zichzelf vaak lager beoordelen dan hun docenten doen<sup>2 4-6</sup>, hoewel het verschil geringer wordt naarmate zij met hun studie vorderen<sup>5 7</sup>.

Deze resultaten worden echter weer tegengesproken door een ander onderzoek<sup>8</sup>, waarbij zelf-beoordeling en docent-beoordeling juist slecht overeen kwamen en er ook geen verbetering te zien was in het vermogen tot zelf-beoordeling. Dit resultaat is opmerkelijk gezien het feit dat het een cursus betrof welke een scherp omschreven vakgebied omvatte, waardoor men zou verwachten dat er juist een hoge correlatie was gevonden.

Wij hebben een pilot-study gedaan van twee soorten beoordelingen van de anamnese, namelijk die door studenten van een door henzelf afgenomen anamnese en de beoordeling van deze anamnese door docenten. Verder was het de opzet de docent-beoordelingen onderling te vergelijken.

De studie werd verricht tijdens de introductie-cursus die medische studenten te Utrecht op de overgang van de preklinische naar de klinische fase aan het begin van het vijfde studiejaar volgen. Ze oefenen

dan onder andere in het afnemen van de anamnese met behulp van simulatie-patiënten. Tijdens het afnemen van de anamnese wordt deze beoordeeld door gedragswetenschappers en artsen die nauw samenwerken maar tevoren nooit hebben geoefend om tot interrater-agreement te komen. De anamnese wordt vastgelegd op video, zodat de student naderhand het gesprek zelf kan bekijken. Voor een meer gedetailleerde inhoud van deze cursus zie Oppelaar, Bleys en Gerritsma (1983)<sup>9</sup>.

De eerste hypothese was nu, dat de zelf-beoordeling en de docent-beoordeling positief met elkaar correleren en de tweede hypothese was, dat de zelf-beoordeling hierbij strenger zou zijn dan de docent-beoordeling. De derde hypothese hield in, dat de docenten positief met elkaar zouden correleren en wel in hoge mate.

Dit onderzoek zou mogelijk duidelijk kunnen maken, dat studenten uitstekend in staat zijn zichzelf te beoordelen, waardoor docent-beoordeling overbodig zou worden. Dit heeft als voordeel dat de student de opleiding onder minder stress en daardoor wellicht met betere resultaten volgt. Tevens wordt de docent ontlast.

### Methode

Elf studenten die allen een anamnese hadden afgenomen, kregen elk een vragenlijst (zie bijlage) uitgereikt. De anamnese werd afgenomen bij simulatie-patiënten wier klachten willekeurig varieerden van zojuist hersteld coma diabeticum tot psychogene buikklachten. Wij hebben dus geen onderscheid gemaakt in de aard van de klachten. Met de vragenlijst, bestaande uit 23 observatie-criteria in vier categorieën ingedeeld, konden zij zichzelf dan achteraf met behulp van een zevenpuntsschaal beoordelen, voordat zij enige feedback hadden ontvangen van hun groep medestudenten en de aanwezige docenten. De video van deze anamnese werd later apart door twee docenten beoordeeld met dezelfde vragenlijsten, nadat zij de anamnese in zijn geheel hadden gezien. De anonimiteit van de studenten werd gewaarborgd doordat een secretaresse de lijs-

ten inzamelde en na het verenigen van de bij elkaar behorende lijsten de namen van de lijsten verwijderde.

De verschillen tussen de beoordelingen van de studenten en beide docenten werden getoetst met Student's t-test of Kolmogorov-Smirnov<sup>10</sup>; de correlaties werden berekend volgens de Pearson-product-moment-correlatie. Voor de berekeningen werd SPSS gebruikt.

### Resultaten

In deze pilot-study gingen we uit van drie hypothesen. Hypothese 1: de zelf-beoordeling van de anamnese door de student en de docent-beoordeling correleren positief. In tabel 2 zien we over het geheel genomen een positieve correlatie tussen Docent 1 en student. Deze positieve correlatie-coëfficiënten zijn significant bij observatiecriteria die o.a. betrekking hebben op de inhoudelijke aspecten van de vragen die in de anamnese gesteld worden.

Slechts driemaal is er tussen Docent 1 en student een, weliswaar niet-significante, negatieve correlatie, nl. de observatiecriteria betreffende 13: het aanmoedigen van de patiënt door de student, zoveel mogelijk in eigen woorden te vertellen, 15: het ingaan van de student op gebaar, uitspraak of aanleiding van de patiënt die van belang kan zijn voor diens probleem, als derde observatiecriterium 16: het hanteren van het juiste evenwicht tussen initiatief nemen en geven door de student.

Verder blijkt uit tabel 2 dat Docent 2 en student minder correleren dan Docent 1 en student. Ook correleren Docent 2 en student vaker negatief (nl. zeven maal) dan Docent 1 en student. Opvallend hierbij is de negatieve correlatie tussen Docent 1 en student bij het observatiecriterium 16 over het handhaven van het juiste evenwicht tussen initiatief nemen en geven door de student, welk een positieve correlatie vertoont bij Docent 2 en student.

Voor wat betreft de eerste hypothese, dat de zelfbeoordeling van de anamnese door de student en de docent-beoordeling positief zouden correleren, zien we dus

tabel 1 Beoordeling van de anamnese door de student zelf en twee docenten; significantie van de verschillen in beoordeling

observatie-criteria in trefwoorden	student x ± sd	Doc 1 x ± sd	Doc 2 x ± sd	stud-Doc 1 significantie	stud-Doc2	Doc1-Doc2 verschil
1.hand	6,3+0,7	7,0+0	5,5+2,2	<0,05 <sup>*</sup>	0,34	0,06
2.voorstellen	6,0+1,2	7,0+0	6,0+1,5	<0,05 <sup>*</sup>	0,88	0,08
3.positie juco	5,5+1,3	6,3+0,7	6,0+1,4	0,09	0,38	0,56
4.juiste tractus-anamn.	4,4+1,9	5,7+1,9	4,3+0,9	0,03 <sup>*</sup>	0,88	<0,05 <sup>*</sup>
5.gedetaill. hoofdklacht	3,8+1,8	4,4+2,1	3,7+1,1	0,49	0,85	0,32
6.open vragen	3,8+1,1	4,2+2,0	6,0+1,0	0,56	0,01 <sup>*</sup>	0,08
7.suggestieve vragen	3,8+1,3	4,5+1,6	5,5+1,6	0,07 <sup>*</sup>	0,02 <sup>*</sup>	1,16
8.meerdere vr. tegelijk	4,8+1,5	6,3+1,3	6,0+1,4	<0,01 <sup>*</sup>	0,13	0,66
9.samenvatten	3,4+1,6	4,4+2,1	2,2+2,1	0,22	0,12	0,01 <sup>*</sup>
10.overgangen	3,7+1,6	4,8+2,1	5,0+1,5	0,12	0,10	0,83
11.lich.houding	4,8+1,3	5,5+1,4	5,5+0,8	0,24	0,17	1,00
12.moeilijke termen	5,6+1,2	5,4+2,0	6,8+0,4	0,73	0,01 <sup>*</sup>	0,15
13.aanmoed. patiënt	4,2+1,4	4,8+2,0	5,6+0,9	0,41	0,02 <sup>*</sup>	0,31
14.onderbrekingen	5,2+1,3	7,0+0	6,7+0,9	<0,01 <sup>*</sup>	0,01 <sup>*</sup>	0,34
15.ingaan op patiënt	4,6+1,6	4,1+1,9	3,9+1,1	0,49	0,25	0,65
16.initiatief	3,5+1,2	5,0+2,0	5,5+0,9	0,06	0,01 <sup>*</sup>	0,47
17.stiltes	4,2+1,5	4,5+2,0	4,1+0,7	0,66	0,86	0,38
18.pat. in waarde laten	5,1+0,9	6,5+1,2	6,5+0,7	<0,01 <sup>*</sup>	0,01 <sup>*</sup>	0,85
19.ingaan op angst,enz.	3,8+1,3	2,8+2,2	3,0+0,8	0,09	0,14	0,45
20.advies onvold. grond	5,5+1,2	7,0+0	6,6+1,2	<0,01 <sup>*</sup>	0,01 <sup>*</sup>	0,34
21.gerustst. onvold. gr.	5,5+1,4	7,0+0	6,7+0,9	<0,01 <sup>*</sup>	0,01 <sup>*</sup>	0,34
22.gelegenh. vragen st.	3,2+2,4	2,7+2,8	2,5+2,5	0,34	0,12	0,08
23.uitleg verdere proced.	4,1+1,6	5,3+2,2	2,9+1,9	0,07	0,17	0,01 <sup>*</sup>

\* p < 0,05

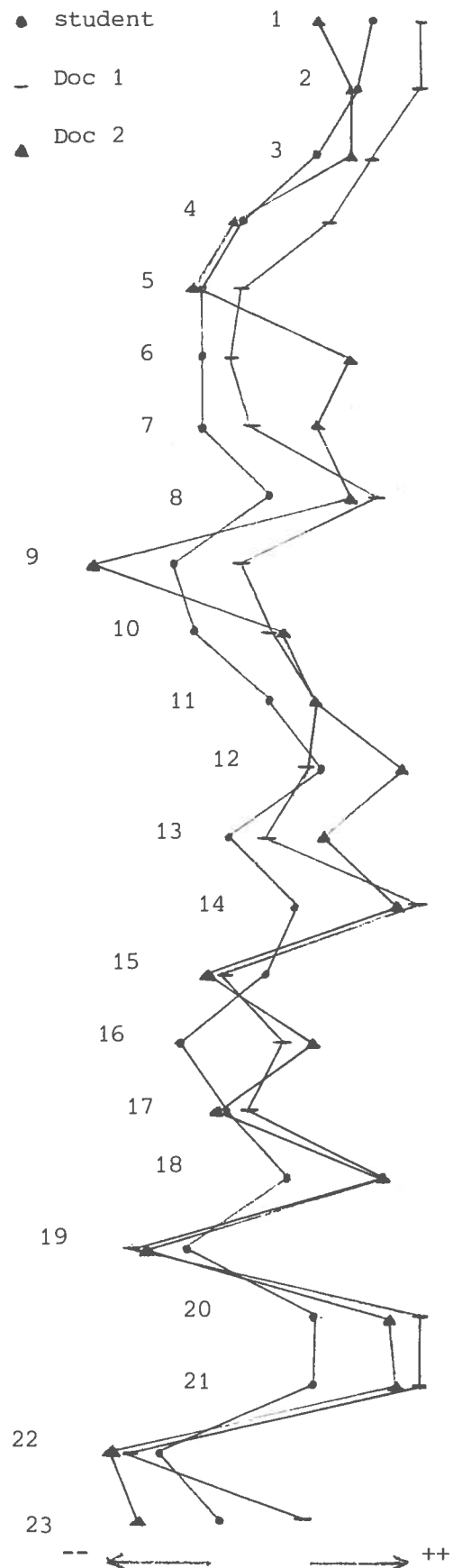
De spreiding van de beoordeling van Docent 1 was bij vijf van de observatie-criteria 0; daardoor kon Student's t-test niet worden toegepast. Significantie van het verschil in beoordeling in die gevallen bepaald met een non-parametrische test (Kolmogorov-Smirnov)

figuur 1 Gemiddelde beoordeling van studenten en docenten

tabel 2 Pearson-correlatie-coëfficiënten

observatie-criteria in trefwoorden	stud/Doc1	stud/Doc2	Doc1/Doc2
1.hand	/-	-0,33	/-
2.voorstellen	/-	-0,20	/-
3.positie juco	0,19	0,19	0
4.juiste tr.anamn.	0,57 <sup>*</sup>	0,05	-0,13
5.gedetaill.hoofdkl.	0,16	0,48	0,31
6.open vragen	0,25	0,21	-0,18
7.suggestieve vragen	0,67 <sup>*</sup>	0,14	0,24
8.meerdere vr. tegelijk	0,64 <sup>*</sup>	-0,28	-0,11
9.samenvatten	0,10	0,27	0,56 <sup>*</sup>
10.overgangen	0,38	-0,17	-0,16
11.lich.houding	0,26	0,18	-0,45
12.moeilijke termen	0,50	0,49	0,09
13.aanmoed. patiënt	-0,02	0,51	0,21
14.onderbrekingen	/-	0,05	/-
15.ingaan op patiënt	-0,02	-0,23	0,64 <sup>*</sup>
16.initiatief	-0,25	0,51	0,21
17.stiltes	0,31	-0,12	0,54 <sup>*</sup>
18.pat. in waarde laten	0,39	0,55 <sup>*</sup>	-0,21
19.ingaan op angst,enz.	0,57 <sup>*</sup>	0,35	0,20
20.advies onvold. grond	/-	0,40	/-
21.gerustst. onvold. gr.	/-	0,60 <sup>*</sup>	/-
22.gelegenh. vragen st.	0,84 <sup>*</sup>	0,84 <sup>*</sup>	0,99 <sup>*</sup>
23.uitleg verdere proc.	0,56 <sup>*</sup>	-0,13	0,27

\*  $p < 0,05$  /- niet te berekenen



dat dit meer opgaat voor Docent 1 en student dan voor Docent 2 en student.

In tabel 1, welke de beoordeling van de anamnese door student en twee docenten en de significantie van de verschillen in beoordeling weergeeft, zien we dat t.a.v. hypothese 2: de zelf-beoordeling van de student is lager/strenger dan die van de docent, dit over het algemeen inderdaad het geval is. Ook hier is er weer een verschil tussen Docent 1 en student enerzijds en Docent 2 en student anderzijds. Dit is nog duidelijker te zien in figuur 1. Docent 1 beoordeelt alleen de observatie-criteria 12 (het gebruik van moeilijke termen/jargon), 15 (student gaat in op gebaar, uitspraak of aanleiding van de patiënt die van belang kan zijn voor diens probleem), 19 (student gaat in op eventuele angsten/zorgen van de patiënt) en 22 (de student geeft de patiënt aan het eind van het gesprek gelegenheid tot vragen stellen) lager dan de student; het verschil is bij deze vier observatie-criteria echter niet significant. Alle andere negentien observatie-criteria beoordeelt hij hoger!

Docent 2 beoordeelt tien observatie-criteria lager, dertien hoger dan de student. Het verschil van de beoordelingen Docent 2/student is wel vaker (nl. negen maal) significant dan bij Docent 1/student (nl. acht maal).

Tenslotte de derde hypothese: docenten onderling vertonen een hoge positieve correlatie. In tabel 2 zien we dat dit over het algemeen niet het geval is. Observatie-criteria 9 (de student vat aan het einde van een bepaalde serie vragen de gegevens samen), 15 (student gaat in op gebaar, uitspraak of aanleiding die van groot belang kan zijn voor diens probleem), 17 (student weet goed om te gaan met stiltes) en 22 (student geeft de patiënt aan het eind van het gesprek gelegenheid tot vragen stellen) zijn de enige vier met een significante positieve correlatie. Bij zes observatie-criteria bestaat er een negatieve correlatie tussen beide docenten.

In tabel 1 zien we dat Docent 1 over

het algemeen hoger beoordeelt dan Docent 2. Alleen de observatie-criteria 11 (student laat in zijn/haar lichaamshouding (oogcontact e.d.) betrokkenheid met de patiënt merken) en 18 (student laat patiënt in zijn/haar waarde) werden door beide docenten gemiddeld hetzelfde beoordeeld.

Opvallend is verder de hoge positieve correlatie bij zowel Docent 1/student, Docent 2/student als Docent 1/Docent 2 t.a.v. observatie-criterium 22: de student geeft de patiënt aan het eind van het gesprek gelegenheid vragen te stellen.

Middelen van de Pearson-correlatie-coëfficiënten<sup>x</sup> Docent 1/student, Docent 2/student en Docent 1/Docent 2 geeft respectievelijk 0,34, 0,17 en 0,15.

### Discussie

Beoordelen blijkt over het geheel genomen toch niet zo gemakkelijk te zijn. De drie hypothesen (1: beoordeling van de door een student afgenomen anamnese door de student zelf en docent-beoordeling correleren positief, 2: studenten beoordelen zichzelf strenger daarbij dan de docenten, 3: er is een hoge positieve correlatie tussen docenten) die we in deze pilot-study over de anamnese-beoordeling door student en twee docenten opstelden, lijken op het eerste gezicht alle drie aannemelijk, maar de resultaten geven hierover toch de nodige twijfels.

Docent 1 en student correleren vaker positief dan Docent 2 en student, zoals blijkt uit tabel 2. Bij het middelen van de correlaties blijkt ook een hogere correlatie te bestaan tussen Docent 1 en student dan tussen Docent 2 en student. Bovendien bestaat tussen Docent 2 en student vaker een negatieve correlatie dan bij Docent 1 en student. Wat zou hieraan ten grondslag kunnen liggen?

Beide docenten zijn nauw betrokken bij de introductiecursus die de studenten, op de overgang van preklinische naar klinische fase volgen; beide worden ook even bekwaam geacht. Docent 1 heeft echter de cursus opgezet en het programma opgesteld. Voorts is hij vaker bij de cursusonderdelen aanwezig dan Docent 2. Hieruit zou geconcludeerd kun-

<sup>x</sup> De gemiddelden zijn berekend zonder transformatie daar de effecten hiervan verwaarloosbaar zijn (zie Glass e.a.<sup>12</sup>)

nen worden dat Docent 1 een zwaarder stempel op het geheel drukt; hij heeft misschien ook meer invloed op de studenten t.a.v. bepaalde doelstellingen die hij zelf gesteld heeft een aan hen wil overbrengen. De studenten komen meer in aanraking met Docent 1 en zijn wellicht geneigd te voldoen aan zijn doelstellingen/eisen; Docent 2 is minder vaak aanwezig en oefent dus minder invloed uit op de studenten.

Opvallend is wel de significante positieve correlatie bij zowel Docent 1/student, Docent 2/student en Docent 1/Docent 2 bij observatie-criterium 22: aan het eind van het gesprek geeft de student de patiënt gelegenheid vragen te stellen. Kennelijk is dit een duidelijker criterium dan bijvoorbeeld observatie-criterium 1: de student geeft de patiënt een hand!!! Men zou verwachten dat dit criterium door student, Docent 1 en Docent 2 unaniem beoordeeld zou worden. Mogelijk speelt een rol dat observatie-criterium 22 aan het eind van de anamnese beoordeeld moest worden; de beoordeling geschiedde immers altijd pas na afloop van het anamnese-gesprek en mogelijk herinnerden student en beide docenten zich dan meer van het eind dan van het begin van de anamnese.

Wat betreft de tweede hypothese (de student beoordeelt zichzelf strenger dan beide docenten), ook hier is weer een verschil aanwezig tussen Docent 1 en student enerzijds en Docent 2 en student anderzijds. Uit de resultaten blijkt dat Docent 1 de studenten over het algemeen hoger beoordeelt dan zij zichzelf beoordelen. Hij beoordeelt slechts vier observatie-criteria lager dan de student. Hiertoe behoren de observatie-criteria 15 (student gaat in op gebaar, uitspraak of aanleiding van de patiënt die van belang kan zijn voor diens probleem) en 19 (student gaat in op eventuele angsten/zorgen van de patiënt). Deze observatie-criteria zou men meer persoonlijk getint kunnen noemen en dus minder absoluut te beoordelen. Persoonlijke ervaringen en waarden spelen mogelijk een belang-

rijke rol.

Docent 2 beoordeelt de ene keer strenger, 7 de andere keer milder dan de student. Een verklaring voor dit alles zou weer kunnen liggen in het feit dat Docent 1 meer bij de cursus betrokken is dan Docent 2.

De beoordeling van observatie-criterium 1 (de student geeft de patiënt een hand) ligt meer in de lijn van Docent 1, die alle studenten een 7 geeft, dan van Docent 2. Het verschil in beoordeling tussen student en Docent 1 bij dit observatiecriterium is significant. Studenten durven misschien zichzelf niet het hoogste cijfer (een 7) te geven en kiezen daarom voor een iets lagere waardering, die wel min of meer overeenkomt met die van Docent 1. Ook uit andere studies blijkt dat studenten zichzelf lager plegen te beoordelen<sup>2 4-6</sup>. Hetzelfde zou men kunnen stellen voor observatie-criterium 2: de student stelt zich voor.

T.a.v. de eerste en de tweede hypothese zou men dus kunnen concluderen, dat de beoordeling door de docent die het meest betrokken is bij de cursus, en de zelf-beoordeling van de student in zekere mate overeenkomen, zij het dat de student zich over het algemeen lager beoordeelt.

De laatste hypothese stelde een hoge positieve correlatie tussen beide docenten. Ons inziens een belangrijke hypothese: immers, van docent-beoordelingen hangt dikwijls veel af voor de student t.a.v. zijn/haar studieverloop. Bij het bekijken van de resultaten valt allereerst op dat Docent 1 over het algemeen de student hoger beoordeelt dan Docent 2. Mogelijkerwijs vindt Docent 1 eerder, dat de student aan "de" (=zijn) eisen voldoet. Bij zes observatiecriteria werd een negatieve correlatie gevonden! Slechts bij vier observatie-criteria, nl. 9 (de student vat aan het eind van een bepaalde serie vragen de gegevens samen), 15 (de student gaat in op gebaar, uitspraak of aanleiding van de patiënt die van belang kan zijn voor diens probleem), 17 (de student weet goed om te gaan met stiltes) en 22 (student geeft de patiënt aan het eind van het gesprek gelegenheid tot vragen stellen) wordt een significante positieve correlatie gevonden.

Verwonderlijk is dan toch wel de negatie-

ve correlatie tussen beide docenten bij observatie-criterium 4: gebruik van de tractus-anamnese op het juiste moment en op de juiste manier. Medisch gezien toch een zeer belangrijk criterium, immers, van deze tractus-anamnese hangt voor een zeer groot deel af wat er verder met de patiënt gebeurt. Hoe is deze negatieve correlatie te verklaren?

In de loop van de jaren als praktizerend arts heeft de tractus-anamnese van Docent 2 mogelijk veranderingen ondergaan, zodanig dat zijn tractus-anamnese niet meer overeenstemt met die van Docent 1 en dus, omdat Docent 1 de cursus heeft opgesteld, ook niet overeenkomt met de wijze waarop de studenten de tractus-anamnese tijdens de introductiecursus leren. T.a.v. dit observatie-criterium beoordelen de studenten zichzelf lager dan Docent 1 (waarbij bovendien het verschil significant is) en hoger dan Docent 2; dit pleit ook ervoor dat de studenten zich naar Docent 1 richten. Bovendien bestaat de tractus-anamnese uit verschillende onderdelen waarop een verschillende nadruk kan worden gelegd, mogelijk tot uiting komend in een verschillende beoordeling door beide docenten. Stel dat de student tijdens een examen door beide docenten zou worden beoordeeld op grond van de tractus-anamnese, dan zou een student die bij de ene docent een voldoende zou halen van de andere een onvoldoende krijgen!

Zelfs het geven van een hand door de student aan de patiënt (observatie-criterium 1) wordt door beide docenten verschillend beoordeeld. Blijkbaar is het een hand geven op zich al een "moeilijkheid" in het beoordelen! Docent 1 geeft alle studenten voor dit observatiecriterium een 7; hij beschouwt kennelijk alleen het feit of er al dan niet een hand gegeven wordt. Docent 2 let waarschijnlijk tevens op de manier waarop een hand wordt gegeven; hoe kijkt de student, wat zegt hij bij het geven van de hand?

Zo ook observatie-criterium 14: de student laat de patiënt uitspreken zonder onnodige onderbrekingen. Volgens Docent 1 worden er geen onderbre-

kingen gemaakt (hij geeft alle studenten een 7). Docent 2 (en de student zelf) vinden blijkbaar dat de patiënt soms wel wordt onderbroken. Misschien is het een kwestie van interpretatie: Docent 1 vindt dat de patiënt een vraag voldoende heeft beantwoord en dat verdere uitweiding overbodig is, zodat het van belang en dus geoorloofd is de patiënt een volgende vraag aan te bieden, terwijl Docent 2 (en de student) dit als onderbreken interpreteren. Een soortgelijke verklaring is aan te dragen voor het verschil in beoordelen van Docent 1 en Docent 2 van o.a. observatiecriteria 20 (student geeft advies op onvoldoende gronden) en 21 (student stelt zonder voldoende reden gerust).

Docenten onderling verschillen dus in hun beoordeling, dat blijkt ook uit de lage correlatie die wordt gevonden na het middelen van de Pearson-correlatie-coëfficiënten. Een betere communicatie over de criteria die beoordeeld moeten worden zou een betere overeenstemming in de beoordeling (z.g. interrater agreement) teweeg kunnen brengen. Het feit dat docenten elkaar goed kennen of nauw met elkaar samenwerken is dus onvoldoende garantie. De docenten zouden formeel moeten oefenen om tot deze overeenstemming in beoordelen te komen.

Voorts is uit onze pilot-study naar voren gekomen, dat de student zich min of meer richt naar de meest betrokken docent. Het overbrengen van een duidelijke doelstelling aan de student, door de student goed duidelijk te maken wat van hem/haar verwacht wordt, lijkt dus aangewezen.

Mogelijk zou een verder onderzoek over het zichzelf beoordelen van de student, met en zonder overbrenging van duidelijke doelstellingen, aan het licht kunnen brengen of docent-beoordeling overbodig wordt als de studenten weten wat van hen verlangd wordt. De docenten zouden dan meer een sturende dan een beoordelende functie kunnen krijgen.

#### Dankbetuiging

Wij danken mw. C.M. Thijs voor het helpen verzamelen van de gegevens, I. Delawi voor het beoordelen van video-banden, J.G.M. Gerritsma voor zijn bijdrage aan het op-

stellen van de vragenlijst, J.A. Smal voor zijn hulp bij de verwerking van de gegevens in de computer en F.C. Bleys voor zijn functie als onze wetenschappelijke stagebegeleider.

## Literatuur

1. Maatsch J L, Huang R, Downing S M, e.a. Predictive validity of medical specialty examinations. East Lansing: OMERAD, 1983
2. Morton J B, Macbeth W A A G. Correlations between staff, peer and self assessment of fourth-year students in surgery. Med Educ 1977;11:167
3. Hammond K R, Kern F. Teaching comprehensive medical care. Cambridge, Mass.: Harvard University Press, 1959
4. Linn B S, Arostegui M, Zeppa R. Performance rating scale for peer and self-assessment. Br J Med Educ 1975; 9:98
5. Cochran S B, Spears M C. Student self-assessment and instructors ratings: a comparison. J Am Diet Assoc 1980; 76:253
6. Kegel-Flom P D. Predicting supervisor, peer and self ratings of intern performance. J Med Educ 1975; 50:812
7. Kennell J H, Tempio C R, Wile M Z. Self-evaluation by first-year medical students in a clinical science programme. Br J Med Educ 1973; 7:230
8. Sclabassi S E, Woelfel S K. Development of self-assessment skills in medical students. Med Educ 1984; 18:226
9. Oppelaar L, Bleys F C, Gerritsma J.G. M. Use of simulation techniques in an intermediate course linking up preclinical and clinical studies. Med Teacher 1983; 5:96
10. Siegel S. Non-parametric statistics for the behavioral sciences. New York, etc: McGraw-Hill, 1956
11. Bleys F C, Gerritsma J G M, Netjes I. Skills development by medical students and the influence of prior experience: a study using evaluation by students and self assessment. Med Educ; geacc. voor publikatie
12. Glass G V, Macgaw B, Smith M. Meta-analysis in social research. Beverly Hills, etc.: Sage, 1981

## Summary

We did a pilot-study comparing student self assessment and expert assessment of history-taking. Based on prior research data we had formulated three hypotheses: 1. self assessment and expert assessment correlate positively; 2. self assessment results in lower ratings than expert assessment; 3. individual ratings by experts correlate positively in a high degree. Only hypothesis no. 2 was proved to be true. The results relating to hypothesis no. 1 were highly dependent on the expert chosen, the correlation being higher with the expert who has the greatest influence on the courses at large. There was virtually not any correlation between the expert ratings, although they know one another very well and collaborate closely in teaching. We conclude that every examiner ought to undergo training in order to obtain inter-rater agreement, and that it is extremely important to instruct students very clearly about course objectives. If the latter condition is fulfilled, examinations may prove to be superfluous.

P.S. De bedoeling was oorspronkelijk drie artsen bij dit onderzoek te betrekken en één statisticus. Tevens hadden we van zestien video's gebruik willen maken. Dit bleek echter niet haalbaar. Slechts twee artsen verklaarden zich bereid hun tijd hieraan te besteden. Ook bij het opstellen van de vragenlijst rezen enige problemen in verband met de wijze van vraagstelling en schaalverdeling, zodat er een socioloog aan te pas moest komen. Verder nam het verzamelen van de gegevens meer tijd in beslag dan verwacht. Het liep namelijk wel eens mis met het bewaren van de video's. Deze bleken dan abusievelijk gebruik te zijn voor nieuwe opnamen, nog voordat beide docenten ze hadden beoordeeld. Uiteindelijk hebben we elf video's kunnen gebruiken, compleet met drie-en-dertig vragenlijsten.