

## O N D E R W I J S

## VERNIEUWING VAN EEN PRACTICUM VOOR TANDHEELKUNDIGE HANDVAARDIGHEDEN

## III. FUNCTIONEREN VAN HET INDIVIDUELE STUDIE SYSTEEM (2)

R. W. R. STEURES  
TH. J. M. TROMP

*Uit de vakgroep Cariologie, Endodontologie en  
Pedodontologie van de Universiteit van Amsterdam.  
Uit het Centrum voor Onderzoek van het  
Wetenschappelijk Onderwijs van de Universiteit van Amsterdam.*

Trefwoorden: Onderwijs – Evaluatie

## 1. Inleiding

In het voorgaande artikel (Steures en Tromp, 1980) werd aandacht besteed aan het functioneren van het als Individueel Studie Systeem (I.S.S.) opgezette preklinische practicum aan de Subfaculteit Tandheelkunde van de Universiteit van Amsterdam. Aan de orde kwamen de gehanteerde onderwijsdoelstellingen, de voorkennis, de geprogrammeerde instructies, de beoordeling van de studieprestaties en bovendien het studietempo en de studievertraging. Tevens werd een aantal maatregelen voorgesteld ter oplossing van gerezen problemen en ter voorkoming van minder gewenste situaties.

In dit artikel zal worden ingegaan op onderzoek naar de mening van studenten over het gevolgde onderwijs, de positie van docenten in een IS-systeem en de beoordeling van studieprestaties door docenten zowel als studenten.

## 2. De mening van de betrokkenen

Om een goed inzicht te krijgen in het functioneren van een bepaald deel van het onderwijs is het noodzakelijk om de mening van de docenten en de studenten te kennen. Zij zijn immers de direct betrokkenen en hebben ideeën over dat onderwijs en over eventuele verbeteringen daarvan. Het practicum wordt op gezette tijden bezocht door de werkgroep Onderwijsontwikkeling en er worden oriënterende gesprekken gevoerd met docenten en studenten. Op het practicum is een 'klachtenboek' aanwezig, bestaande uit de volledige reeks van geprogrammeerde instructies, waarin het commentaar van studenten en docenten met betrekking tot het functioneren van het practicum wordt genoteerd. Dergelijk commentaar heeft er bijvoorbeeld toe geleid dat de beoordelingsschema's van de verschillende instructies beter op elkaar werden afgestemd.

Met het verzamelde commentaar als uitgangspunt worden evaluatiebijeenkomsten met docenten en studenten belegd.

Deze hebben onder andere tot gevolg gehad dat de toetsituatie werd aangepast (zie Steures en Tromp, 1980). Verder wordt het commentaar benut voor het maken van enquêtevragen (waarover hieronder meer) en voor de herziening van het schriftelijke en audiovisuele instructiemateriaal.

## 2.1. De enquêtering van studenten

De enquêtes die werden gehouden in het eerste respectievelijk vijfde jaar dat het vernieuwde practicum functioneerde worden in het navolgende besproken en vergeleken. (De enquêtes die in de tussentijd werden afgenomen blijven hier buiten beschouwing.) Op deze wijze wordt een indruk verkregen van veranderingen in de meningen van de studenten.

De vragen van de enquêtes zijn van het type met vijf antwoordmogelijkheden, lopende van 'sterk mee eens' tot 'sterk mee oneens'.

De enquêtevragen hebben betrekking op:

1. de werkcolleges en intreetoetsen;
2. het schriftelijke instructiemateriaal;
3. het audiovisuele instructiemateriaal;
4. de begeleiding door de docenten;
5. de beoordeling van de studieprestaties;
6. algemeen.

De student kon nog aanvullend commentaar geven als antwoord op een open vraag aan het einde van elke enquête.

De enquêtes zijn te omvangrijk om hier in hun geheel te bespreken. Daarom zullen alleen die vragen worden besproken welke beide enquêtes gemeenschappelijk hebben. Voor een beter overzicht zijn de antwoorden op de categorieën 'sterk mee eens' en 'mee eens' respectievelijk 'sterk mee oneens' en 'mee oneens' samengenomen.

De eerste enquête werd afgenomen begin 1974, korte tijd nadat de eerste veranderingen in het practicum met betrekking tot het gedeelte Conserverende Tandheelkunde waren doorgevoerd. De student werd ge-

## Samenvatting:

In dit artikel, het laatste in een reeks van drie, wordt ingegaan op het in de praktijk functioneren van het als Individueel Studie Systeem opgezette preklinische practicum aan de Subfaculteit Tandheelkunde van de Universiteit van Amsterdam.

Besproken wordt achtereenvolgens onderzoek naar de meningen van studenten en docenten, de waarde van de beoordeling van eigen werk door studenten en de training van docenten in beoordelen.

vraagd zijn mening te geven over de 14 instructies van dit gedeelte. De enquête bevatte tevens vragen over de practicumgedeelten die nog volgens de oude opzet functioneerden.

De gegevens staan vermeld in tabel I. Zeventenzeventig van de 92 studenten vulden de enquête in. Onder het voorbehoud van het beperkt aantal vragen met betrekking tot de vergelijking van de nieuwe en de oude onderwijsopzet, kan het resultaat als een ondersteuning van de nieuwe onderwijsopzet worden gezien, mede gelet op het feit, dat 84 procent van de studenten de nieuwe opzet positief waardeerden.

In tabel II staan de resultaten met betrekking tot de gemeenschappelijke vragen van de eerste enquête en de tweede, die eind 1978 werd gehouden.

Voor wat betreft twee belangrijke zaken is een duidelijke verbetering te constateren: er is sprake van een aanzienlijke toename van de waardering voor het instructiemateriaal en voor de samenwerking tussen docenten en studenten. Daartegenover staat een afname van de waardering voor (in volgorde van grootte): het practicum als geheel, de intreetoetsen, de beoordelingscriteria en de video-instructies. Een deel van de hogere waardering van de studenten voor de nieuwe onderwijsopzet moet waarschijnlijk worden toegeschreven aan het zogeheten Hawthorne-effect. Dit effect houdt in, dat de toegenomen waardering voor een vernieuwing voor een deel het gevolg is van de invoering van de vernieuwing en niet zozeer van de inhoud van de vernieuwing. Dit geldt voor tabel I (de nieuwe opzet) en tabel II (de eerste enquête). Afgaande op de lage score voor de tweede vraag zou men zeggen dat de studenten teleurgesteld zijn in de mogelijkheden van de geprogrammeerde instructies om juist de moeilijke aspecten van de te leren handelingen te verduidelijken. De blijkers de tweede enquête toegenomen waardering van het instructiemateriaal kan op rekening worden geschreven van de verbeteringen die daarin in de loop der jaren zijn aangebracht. Dat de vraag naar de

Tabel I. De eerste enquête: een vergelijking van de oude en de nieuwe practicumopzet.

Vraag	'(sterk) mee eens' (in procenten van het aantal antwoorden)	
	Oude opzet	Nieuwe opzet
1. Het instructiemateriaal is van tevoren bestudeerd.	19	65
2. Het resultaat van de beoordelingen is inzichtelijk.	26	58
3. De leerstof wordt beheerst.	28	61

Tabel II. Gemeenschappelijke vragen van beide enquêtes.

Vraag	'(sterk) mee eens' (in procenten van het aantal antwoorden)	
	1e enquête (jaar 1)	2e enquête (jaar 5)
1. Door de intreetoetsen wordt meer van de instructies geprofiteerd.	73	45
2. Het instructiemateriaal is voldoende m.b.t. de moeilijke punten.	17	58*)
3. Er is profijt van de video-programma's.	46	38*)
4. De beoordelingscriteria zijn duidelijk.	69	50*)
5. De samenwerking tussen studenten en docenten is bevredigend.	40	71
6. Het practicum als geheel wordt positief gewaardeerd.	84	50

\*) Samengestelde percentages; voor een toelichting, zie paragraaf 2.1. van de tekst.

Tabel III. Uitsplitsing naar instructie van de samengestelde percentages uit tabel II.

Vraag	Instructie	'(sterk) mee eens' (in procenten van het aantal antwoorden)					
		Foramen coecum	Klasse V	Klasse I-prep.	Klasse I-rest.	Klasse II-prep.	Klasse II-rest.
1. Het instructiemateriaal is voldoende m.b.t. de moeilijke punten.		68	74	41	56	44	48
2. Er is profijt van de video-programma's.		53	55	26	29	37	30
3. De beoordelingscriteria zijn duidelijk.		70	67	33	42	37	50

samenwerking tussen de studenten en docenten aanvankelijk een lage score oplevert, wordt mogelijk veroorzaakt door het feit, dat de docenten op dat moment minder goed waren ingespeeld op de hen toebedachte rol in de nieuwe opzet van het practicum (zie Steures en Tromp, 1980).

De in tabel II met een asterisk gemerkte percentages hebben betrekking op de 'gemiddelde' instructie. In de tweede enquête werd de student echter om zijn mening gevraagd over zes (conserverende) instructies, namelijk klasse V-, foramen coecum-, klasse I-preparatie en -restauratie en klasse II-preparatie en -restauratie. Uitsplitsing naar deze instructies (tabel III) levert een specifiek beeld op, waarin de fors uiteenlopende waardering van de onderscheiden instructies opvalt.

Een verklaring hiervoor is niet gemakkelijk te geven. De genoemde enquêtes worden na elk onderwijsblok (bestaande uit twee of drie instructies) afgenomen. De enquêtes zijn kort en voor elk blok hetzelfde, zodat de gegevens systematisch met elkaar kunnen worden vergeleken en de ernst van de opgespoorde problemen kan worden bepaald. In de praktijk blijkt het moeilijk te zijn om alle studenten van het belang van het invullen van de enquêtes te overtuigen. Niet iedere student levert de enquêtes in en door de verschillen in studietempo komen deze druppelsgewijs binnen. Hierdoor kunnen de verkregen resultaten slechts als globale indicaties worden gehanteerd.

De op basis van de resultaten doorgevoerde maatregelen blijken niet altijd het gewenste effect te sorteren. Dit was bijvoorbeeld het geval met de verwijzing naar de bibliotheek voor het bekijken van de video-instructies. De studenten zagen de verwijzing als een onplezierige onderbreking van het practicum en namen niet de moeite om voor het zien van een programma naar de bibliotheek te gaan. Daarom werden de mogelijkheden om deze instructies op de eigen plaats in de practicumruimte te bekijken uitgebreid.

Wanneer de antwoorden van studenten op een enquêtevraag in een minder gunstige richting wijzen, dan wordt geprobeerd om door overleg met de docenten en studenten de problemen op het spoor te komen en oplossingen te vinden. Voor dit doel worden de antwoorden op de open vraag van de enquêtes gerubriceerd. In tabel IV staan bij wijze van illustratie enkele antwoorden. Het is verleidelijk om de oplossing voor problemen zoals in deze tabel vermeld, te zoeken in hun tegendeel; bijvoorbeeld, méér toetsgelegenheden, minder wachttijd bij de beoordeling van werkstukken, etc. Deze aanpak zal waarschijnlijk nog wel opgaan voor elk probleem afzonderlijk, maar is om redenen van mankracht en organisatie niet uitvoerbaar voor alle problemen te zamen.

Tabel IV. Voorbeelden van antwoorden op de open vraag van de tweede enquête.

1. Er zijn te weinig toetsgelegenheden.
2. Er zijn discrepanties in de beoordeling van toets- en oefenwerkstukken.
3. De loop van het practicum wordt teveel en onverwacht onderbroken.
4. Er is een lange wachttijd bij de beoordeling van werkstukken.
5. De taak van de docenten dient op bepaalde punten duidelijker te worden gemaakt.
6. De instructies zijn niet wetenschappelijk genoeg.
7. Er wordt van docenten een actievere begeleiding gevraagd.
8. Organisatorische zaken en eventuele veranderingen daarin dienen tijdig te worden bekend gemaakt.

## 2.2. De mening van docenten

Met betrekking tot de mening van de docenten over het nieuwe onderwijs werden geen gegevens door middel van enquête verzameld. Volstaan moet hier dan ook worden met enkele algemene opmerkingen over de opvattingen van docenten die werkzaam zijn binnen een IS-systeem. Van der Meer en Plomp (1977) stellen een indeling van IS-systemen voor, die berust op de wijze waarop de samenwerking tussen de betrokken docenten is geregeld:

1. éénmanssystemen;
2. meermanssystemen;
3. systemen met externe medewerkers;
4. systemen met gescheiden voorbereiders en uitvoerders.

Het preklinische practicum behoort tot de vierde categorie. Uit de literatuur blijkt, dat docenten die in een systeem van deze categorie uitsluitend met uitvoerende taken zijn belast, minder voldoening ervaren dan wanneer zij conventioneel onderwijs verzorgen: men voelt zich meer 'administrateur' dan docent. Het is de vraag of deze bevinding van toepassing is op het gedeelte van het practicum dat in het tweede cursusjaar is gesitueerd. Het merendeel van de docenten bestaat daar namelijk uit studenten (in de functie van student-assistent), die niet eerder ervaring hebben opgedaan met het geven van onderwijs en die bovendien na een betrekkelijk korte periode door anderen worden opgevolgd. Voor het derdejaars-gedeelte van het practicum, met afgestudeerde docenten, is waarschijnlijk wél van een geringere arbeidssatisfactie sprake. De oplossing voor het tevredenheidsprobleem moet in het algemeen worden gezocht in het afzwakken van de strikte scheiding tussen de ontwikkeling en de uitvoering van het onderwijs. Indien de uitvoerders een deel van de voor-

bereiding van het onderwijs voor hun rekening kunnen nemen (het maken van de leerstof e.d.), dan kunnen zij meer hun stempel drukken op het onderwijsgebeuren. (Hierdoor zou het preklinische practicum verhuizen van de vierde naar de tweede categorie van IS-systemen.)

## 3. De beoordeling door studenten van eigen werk\*

### 3.1. Aanleiding tot het onderzoek

In het vorige artikel (Steures en Tromp, 1980) werd ingegaan op de betekenis die in het preklinische practicum wordt gehecht aan de beoordeling door de studenten van hun eigen werkstukken. Wil men dit oordeel over de eigen studieprestaties niet alleen als een 'aardigheidje' beschouwen,

\* Deze paragraaf is gebaseerd op Tromp (1979).

maar ook als belangrijk voor een goede studievoortgang, dan moet men weten in welke mate op dat oordeel kan worden vertrouwd.

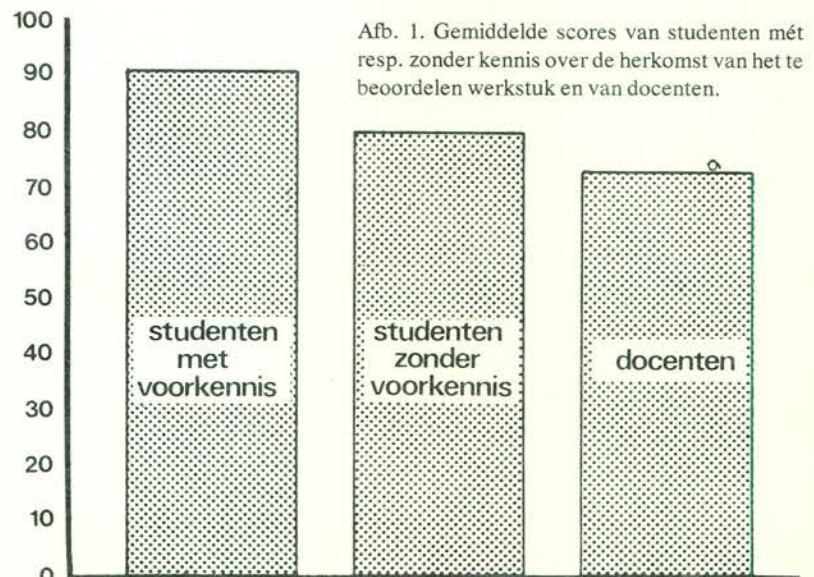
Nu is het gebruikelijk dat studenten hun oordeel geven in een situatie waarin zij weten dat het om het eigen werk gaat. De vraag is in hoeverre die kennis van invloed is op de beoordeling. De aanleiding tot het hier te bespreken onderzoek wordt gevormd door de waarneming, dat studenten over het algemeen het eigen werk overschatten. Deze bevinding is wellicht begrijpelijk, wanneer men bedenkt, dat elk oordeel van een student wordt gevolgd door het oordeel van minstens één docent (over hetzelfde werkstuk). In een dergelijke situatie is het vanuit de student gezien niet onverstandig om het eigen werk wat positiever te beoordelen: 'nee' heeft hij, 'ja' kan hij krijgen.

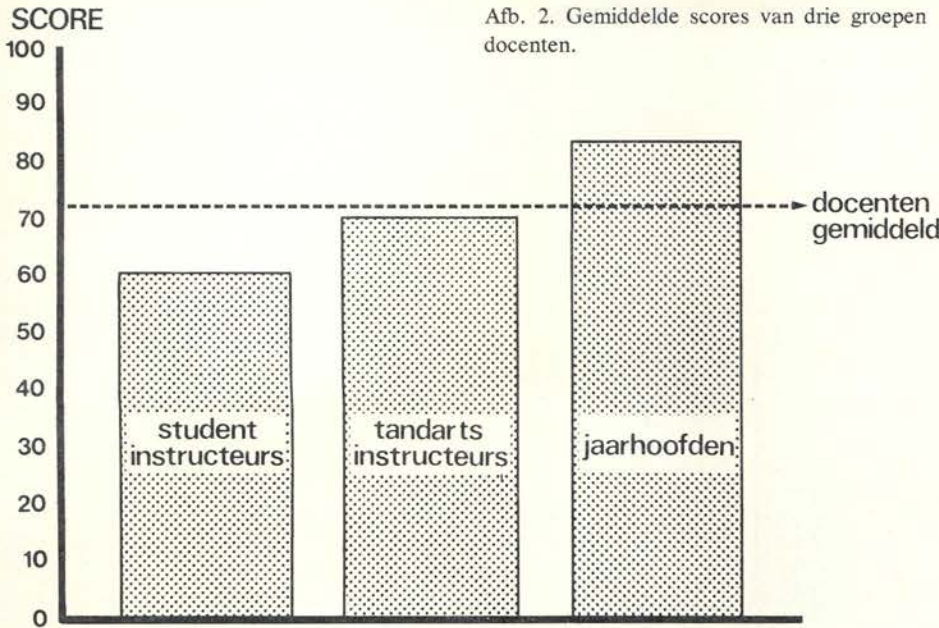
De docenten verkeren in een andere situatie dan de studenten; daarom mag worden verwacht dat hun oordeel gemiddeld minder positief is.

Het is overigens niet voldoende om het gemiddelde oordeel van de docenten te kennen. Uit voorafgaand onderzoek (o.a. Hoekman, 1975) blijkt, dat er aanzienlijke verschillen zijn in de beoordelingen gedaan door de onderscheiden deelgroepen van docenten, te weten het jaarhoofd, de tandarts-instructeurs en de student-instructeurs. Het jaarhoofd blijkt het meest positief van allen te beoordelen en de tandartsen positiever dan de student-instructeurs. De zelfbeoordeling van studenten ligt ongeveer op het niveau van de tandartsen.

### 3.2. Werkwijze bij het onderzoek

Achternvijftig derdejaarsstudenten maakten (najaar 1978) elk een klasse II-preparatie, die zij beoordeelden op 15 criteria; dit is de zelfbeoordeling met voorkennis. Het





werkstuk diende te worden gemaakt als gold het een toets. Bij eventuele fouten kon men opnieuw beginnen. De score voor een werkstuk bestond uit de som van het aantal criteria waarop het werkstuk als voldoende was beoordeeld. (In het onderhavige geval was de score maximaal 15, minimaal 0.)

Van de genoemde studenten beoordeelden er 50 geruime tijd later voor de tweede keer hetzelfde werkstuk, nu zónder van de herkomst op de hoogte te zijn; dit is de zelfbeoordeling zonder voorkennis.

Het onderzoek werd als volgt bij de studenten geïntroduceerd. De staf van het practicum wil een indruk hebben van het peil van de (in het tweede cursusjaar) opgedane vaardigheden. Het gaat om een gemiddelde indruk. Mochten de werkstukken over het algemeen onvoldoende blijken te zijn, dan zou dit aanleiding kunnen zijn om extra onderwijs in te lassen.

Na de beoordeling door studenten beoordeelden zeven docenten van het derdejaars practicum allen alle werkstukken. Het ging om één student-instructeur, vier tandarts-instructeurs en twee jaarhoofden. De samenstelling van deze groep was ongeveer representatief voor de normale stafbezetting.

### 3.3. Resultaten van het onderzoek

De eerste verwachting luidde, dat de beoordeling, in de wetenschap met een eigen werkstuk te doen te hebben, een hogere score oplevert dan een beoordeling zonder deze kennis. Deze verwachting blijkt juist te zijn (t-toets voor afhankelijke steekproeven:  $t = 5,8$  ( $p = .001$ ); afb. 1).

De tweede verwachting luidde, dat de zónder voorkennis gegeven beoordeling niet verschilt van de gemiddeld door docenten

gegeven beoordeling. Deze verwachting blijkt niet juist te zijn; de studenten beoordelen positiever (t-toets voor afhankelijke steekproeven:  $t = 3,6$  ( $p = .01$ ); afb. 1).

De veronderstellingen zijn getoetst met behulp van gemiddelde scores, maar zij gaan ook op voor het merendeel van de afzonderlijke gevallen. Voor de eerstgenoemde gaat dit in 90 percent van de gevallen op, voor de tweede in ruim 70 percent.

De derde verwachting luidde, dat de jaarhoofden positiever beoordelen dan de tandarts-instructeurs en dat deze positiever beoordelen dan de student-instructeurs. Deze verwachting blijkt juist te zijn (afb. 2). Gebruik werd gemaakt van de procedure van multiële vergelijkingen volgens Dunn (Kirk, 1968).

### 3.4. Implicaties van het onderzoek

De voornaamste vraag van dit onderzoek betreft de 'status' van het oordeel van studenten over hun eigen werk. Het blijkt significant verschil uit te maken of studenten al dan niet weten, dat het werk dat zij beoordelen hun eigen werk is. Studenten zijn kennelijk in staat hun oordeel over het eigen werk aanzienlijk te wijzigen in een meer realistische richting, te weten het niveau van de docenten. Daarom zou voortaan de beoordeling door de student van het eigen werk gedurende de oefenfase van de instructies verplicht moeten zijn. Om het belang van een realistisch eigen oordeel te benadrukken, zou het oordeel van de docent dan moeten vervallen. Voor de gemiddelde student (in termen van studietempo en benodigde hulp) zou dit betekenen, dat deze zo goed als zelfstandig de toetsfase van een instructie bereikt. De vraag kan worden gesteld of deze aanscherping van de opzet van het practicum niet te radicaal

is. Voor een aantal studenten zullen de instructies alleen te weinig zekerheid verschaffen. Een meer reële oplossing lijkt daarom, het docentenoordeel niet zozeer te laten vervallen als wel er van uit te gaan dat beider oordeel in beginsel gelijkwaardig is. Wil deze wijziging effect sorteren, dan is het nodig dat de zin van het verrichten van zelfbeoordelingen in een Individueel Studie Systeem bij voortdurend bij studenten aannemelijk wordt gemaakt. Anders loopt men het risico, dat zulke beoordelingen door sommige studenten slechts als een versiering worden gezien.

Het is overigens ook een vraag in hoeverre staat kan worden gemaakt op het docentenoordeel, dat weliswaar gemiddeld gesproken 'strenger' is dan het studentenoordeel (afb. 1), maar dat intern aanzienlijke verschillen vertoont (afb. 2). De studenten mogen dan niet op één lijn zitten met de docenten, de docenten doen dat zelf onderling ook niet. Aan afbeelding 2 valt af te lezen, dat het gemiddelde docentenoordeel in feite alleen redelijk het oordeel van de 'middenmoters' (dat zijn de tandarts-instructeurs) weerspiegelt. In de praktijk overigens vormen dergelijke beoordelingsverschillen tussen docenten nauwelijks een probleem, doordat de studenten zijn toegewezen aan vaste koppels van docenten.

Dat er sprake is van aanzienlijke beoordelingsverschillen tussen de docenten, hoeft niet strijdig te zijn met de in het vorige artikel (Steures en Tromp, 1980) vermelde bevinding, dat docenten middels training dergelijke verschillen weten te verkleinen. De positieve effecten van calibratietrainingen zijn waarschijnlijk tamelijk instructiegebonden en van betrekkelijk korte duur.

## 4. Training van docenten in beoordelen

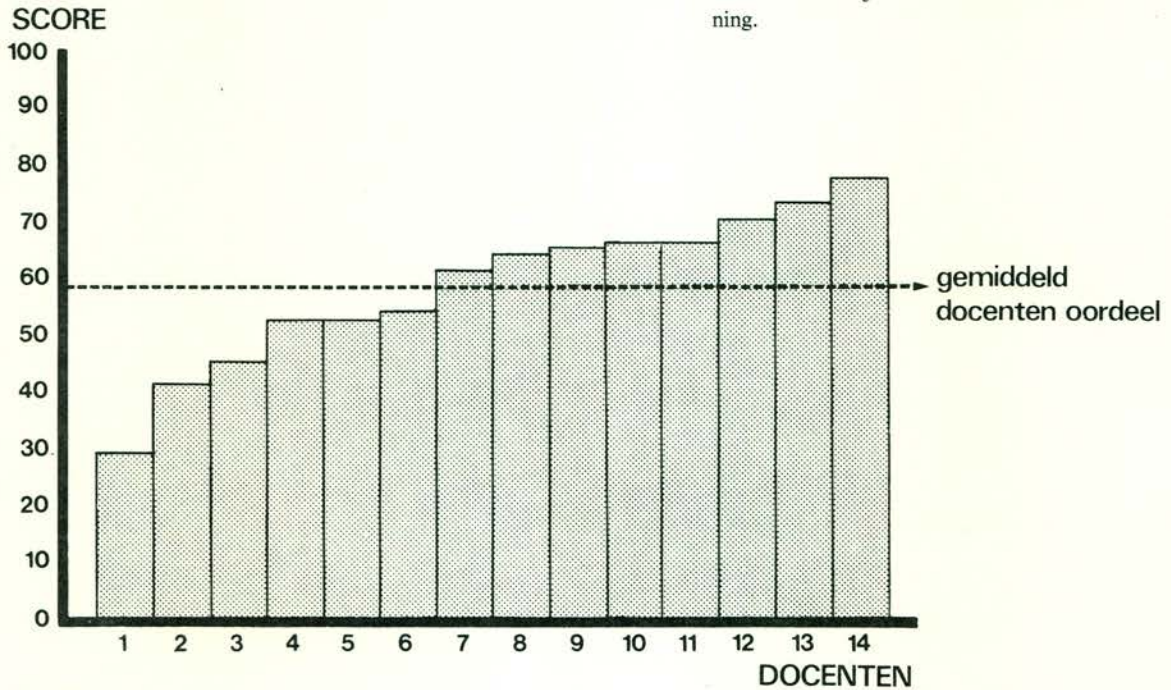
### 4.1. Doel van de training

Regelmatig worden voor de docenten van het preklinische practicum calibratietrainingen gehouden. Het voornaamste doel van een calibratietraining is, het vaststellen van eventuele (forse) verschillen tussen docenten in de beoordeling van de prestaties van studenten en het door middel van bespreking van de mogelijke oorzaken terugdringen van deze verschillen.

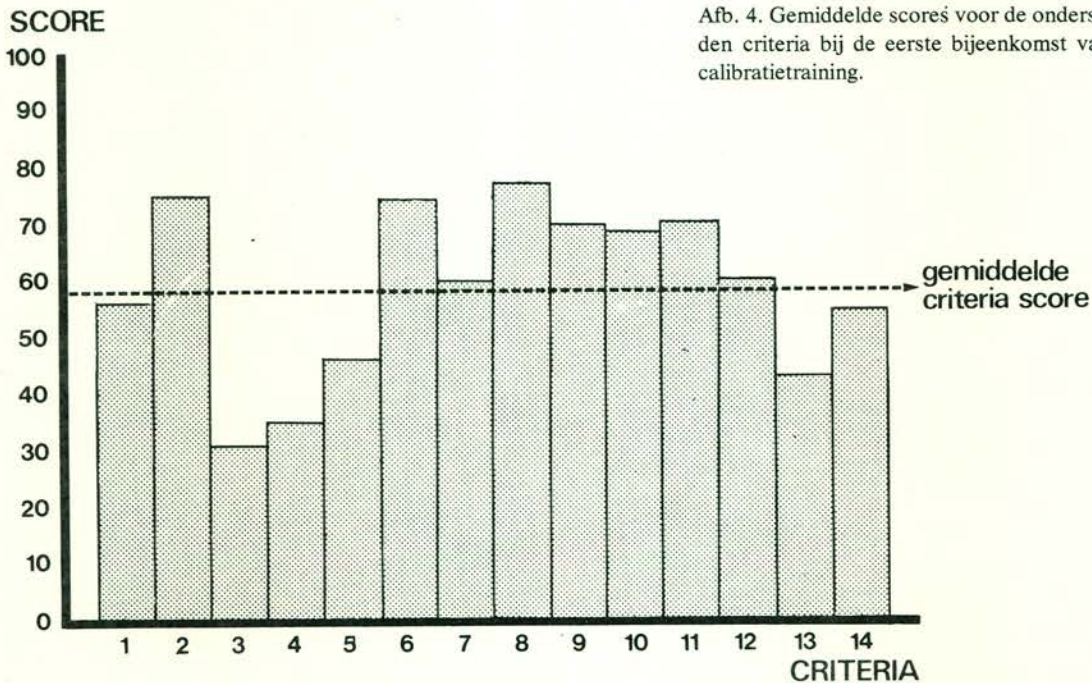
In de voorgaande artikelen (Steures en Tromp, 1978; 1980) kwamen de criteria, waarmee de werkstukken op hun kwaliteiten worden beoordeeld, reeds ter sprake. Deze beoordelingscriteria kunnen om de volgende redenen objectief worden genoemd:

1. de formulering is waar mogelijk in kwantitatieve termen gesteld (millimeters, graden e.d.);
2. de ruimte die een beoordelaar wordt gelaten is tot een minimum teruggebracht door het gebruik van een twee-

Afb. 3. Gemiddelde scores van de docenten tijdens de eerste bijeenkomst van de calibratietraining.



Afb. 4. Gemiddelde scores voor de onderscheiden criteria bij de eerste bijeenkomst van de calibratietraining.



puntsschaal ('voldoende' - 'onvoldoende').

De genoemde objectiviteit is natuurlijk niet absoluut. Docenten vormen een heterogene groep, door een uiteenlopende opleiding en praktijkervaring, en persoonlijkheid. Daardoor zal het altijd noodzakelijk blijven om de beoordelingen regelmatig onder de loop te nemen.

4.2. Gang van zaken bij de training

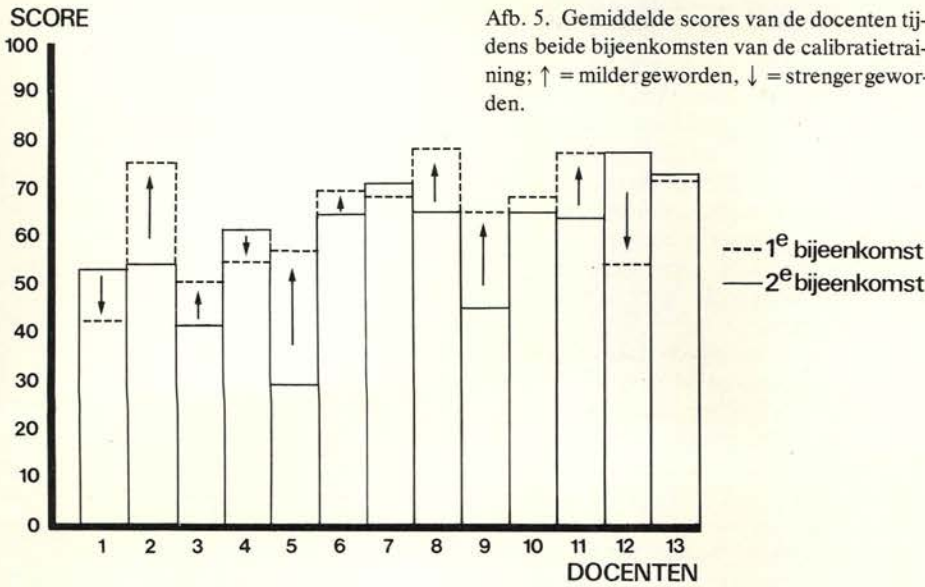
Voor alle trainingen geldt in de eerste plaats, dat de werkstukken van studenten

anoniem worden beoordeeld. In de tweede plaats vinden de beoordelingen door de docenten onafhankelijk van elkaar plaats; met andere woorden, zonder dat men van elkaars beoordelingen op de hoogte is. In de derde plaats is het zo, dat de werkstukken aan de docenten worden aangeboden in willekeurige volgorde, terwijl hetzelfde beoordelingsformulier wordt gebruikt als in de instructies.

Een calibratietraining bestond aanvankelijk uit twee bijeenkomsten, met een tussentijd van ongeveer een week. Door deze tussentijd konden de beoordelaars betrekkelijk onbevangen aan een tweede beoor-

deling van hetzelfde materiaal beginnen. Bovendien was er dan tijd voor de verwerking van de gegevens van de eerste ronde. Van de deelnemers werd verwacht, dat zij het instructiemateriaal bij de te beoordelen werkstukken nog eens grondig doornamen.

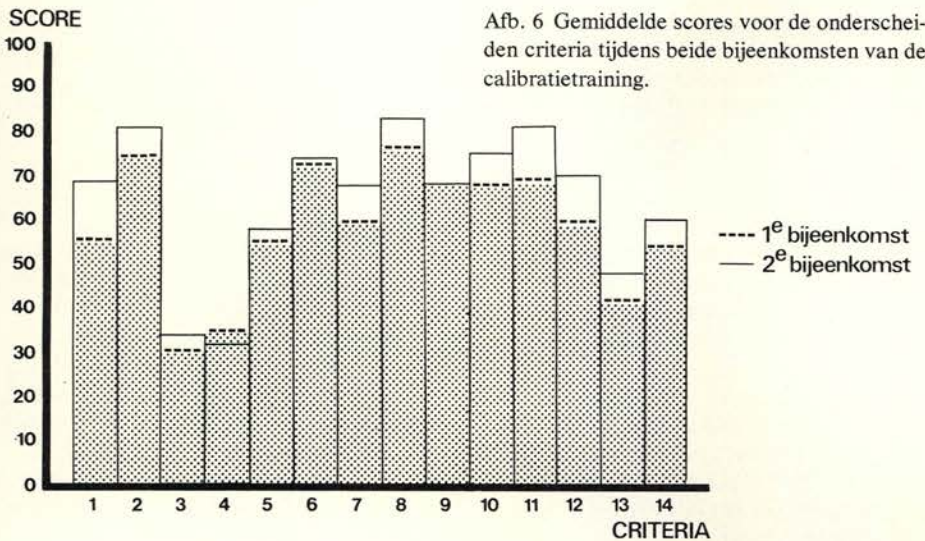
Een calibratietraining was in het algemeen zo opgezet, dat elke docent op beide bijeenkomsten elk werkstuk op alle criteria beoordeelde; de onderzoekopzet is zodoende 'volledig gekruist'. Voor de trainingen werd veelal de klasse II-preparatie,



met 15 criteria, gebruikt. Het aantal werkstukken dat tijdens een calibratietraining werd beoordeeld is om organisatorische redenen tenminste gelijk aan het aantal beoordelaars.

Er wordt sinds enige tijd afgezien van het beleggen van de tweede trainingsbijeenkomst, omdat het positieve effect van de trainingen zich in de loop der jaren heeft bewezen en er zeer veel werk is gemoeid met de volledige bewerking van de gegevens.

Bij een calibratietraining gaat de belangstelling in de eerste plaats uit naar de verschillen in strengheid waarmee de docenten beoordelen. In de tweede plaats zijn van belang de verschillen in de mate waarin de werkstukken aan de respectieve criteria voldoen (de 'moeilijkheid' van de criteria). Tenslotte is belangrijk de interactie tussen beoordelaars en criteria, dit wil zeggen de mate waarin de beoordelaars onderling verschillen met betrekking tot een en hetzelfde criterium.



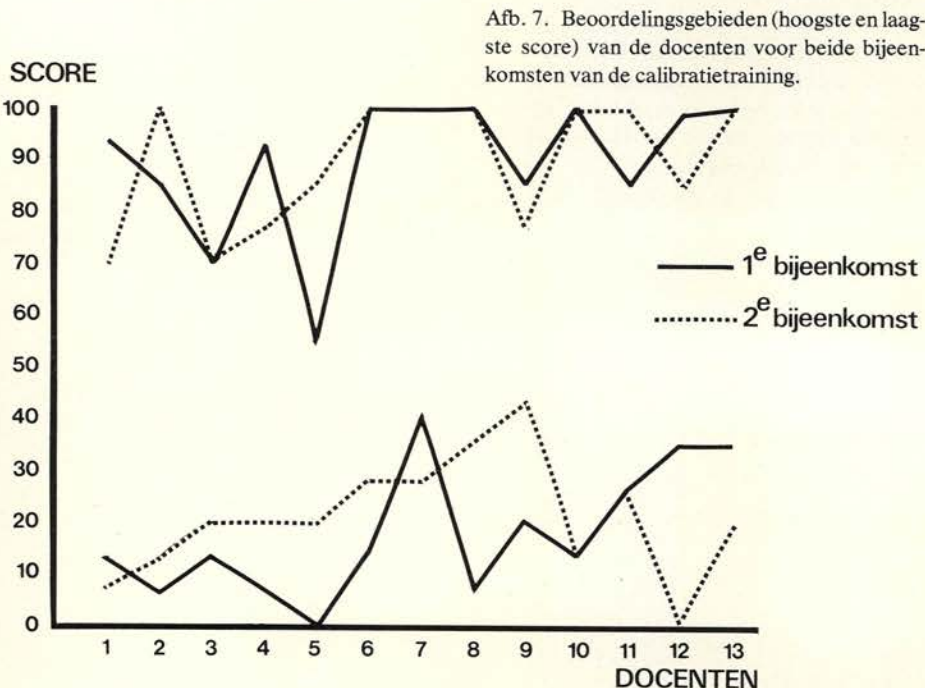
### 4.3 Resultaten van de training

#### 4.3.1. De eerste trainingsbijeenkomst

Bij wijze van voorbeeld wordt de calibratietraining van januari 1977 besproken. Het betreft nog een training van twee bijeenkomsten met een week tussentijd. De deelnemers zijn docenten van het tweedejaarspracticum, te weten 11 student-instructeurs en twee tandartsen (waaronder het jaarhoofd). De te beoordelen werkstukken betreffen de instructie voor totale omslijping voor een gegoten kroon, met 14 criteria.

De verschillen in beoordeling tijdens de eerste trainingsbijeenkomst zijn spectaculair te noemen.

Afbeelding 3 toont deze verschillen. Uit de afbeelding blijkt dat de twee tandartsen (docenten 13 en 14) het mildst in hun oordeel zijn en dat alle student-instructeurs strenger zijn. De gemiddelde strengheid voor alles is 58 (procent); met andere woorden, de gemiddelde instructeur geeft bijna zes op de tien keer een voldoende.



Er zijn twee docenten die vergeleken met de andere extreem beoordelen. De strengste docent geeft niet vaker dan drie op de tien keer een voldoende; de mildste docent geeft een kleine acht op de tien keer een voldoende. Met deze uitersten is het beoordelingsgebied van de docenten aangegeven. Van een schaal die loopt van nul tot honderd (procent) wordt ongeveer de helft gebruikt (van 29 tot 77); eveneens een illustratie van de eerder vermelde bevinding dat de docenten het onderling verre van eens zijn.

De calibratietrainingen leveren ook gegevens op met betrekking tot de moeilijkheid van de beoordelingscriteria (zoals door een bepaalde groep docenten beoordeeld). In afbeelding 4 zijn de verschillen tussen de criteria aangegeven. Moeilijke criteria, zoals bij de besproken instructie de nummers 3, 4, 5 en 13, worden tijdens de trainingsbijeenkomsten nader onderzocht. Als oorza-

ken voor de moeilijkheid van criteria kunnen worden genoemd:

1. de (deel-)handeling waarop een criterium betrekking heeft, is op zichzelf ingewikkeld;
2. de desbetreffende handeling is in het instructiemateriaal niet duidelijk beschreven;
3. een criterium is niet eenduidig geformuleerd, zodat de wijze waarop het moet worden beoordeeld niet duidelijk is.

Hoe de criteria worden gehanteerd blijkt in de eerste plaats van de docenten af te hangen. Deze conclusie komt het treffendst tot uiting als wordt nagegaan in hoeverre de docenten een criterium eensgezind waarderen. Het blijkt, dat aan één zelfde criterium door sommige docenten de hoogste en door andere docenten de laagste score wordt gegeven. Criterium nummer 8 scoorde zes keer het hoogst en één keer het laagst; criterium nummer 5 één respectievelijk drie keer; criterium nummer 4 één respectievelijk acht keer.

Terwille van een verkleining van de beoordelingsverschillen tussen de docenten is het van groot belang, dat aan de hand van de BeoordelaarsxCriteria-tabel (die hier niet is opgenomen) de meest frappante tegenstrijdigheden en 'uitschieters' worden opgespoord.

Door het bespreken van de beoordelingsverschillen kan men in de toekomst tot een eenstemmiger oordeel komen.

#### 4.3.2. De tweede trainingsbijeenkomst

Het doel van de tweede trainingsbijeenkomst is vanzelfsprekend om na te gaan in hoeverre de geconstateerde verschillen in omvang zijn teruggebracht.

Afbeelding 5 toont de verschillen in strengheid tussen de beoordelaars tijdens de eerste en de tweede bijeenkomst. Alle docenten zijn de tweede keer gemiddeld gesproken iets minder streng geweest (van 58 naar 64), hetgeen is terug te vinden in het oordeel van de afzonderlijke docenten: acht zijn milder geworden, vijf strenger. De ver-

schillen tussen de eerste en tweede bijeenkomst zijn voor de afzonderlijke docenten aanzienlijk. Hier staat echter tegenover, dat de docenten die eerst het strengst dan wel het mildst oordeelden, elkaar thans nauwelijks meer ontlopen. Verschuivingen van de afzonderlijke beoordelingen in de (gewenste) richting van het groeps-gemiddelde komen voor, hoewel een vraagteken gezet zou kunnen worden bij de duurzaamheid van de meer spectaculaire verschuivingen.

Grote beoordelingsverschillen kunnen wellicht voor een deel blijvend worden verkleind door middel van een scherpere omschrijving van de criteria (minder ambiguïteiten e.d.).

In afbeelding 6 worden de verschillen weergegeven tussen de scores voor de verschillende criteria van de eerste en de tweede bijeenkomst. Per criterium is het verschil tussen beide bijeenkomsten niet erg groot. De tweede keer worden meer criteria milder beoordeeld dan de eerste keer. De verschillen tussen de criteria zijn daarentegen niet afgenomen.

Afbeelding 7 tenslotte, toont de beoordelingsgebieden van de respectieve docenten voor beide bijeenkomsten. Boven- en ondergrens van een beoordelingsgebied worden gevormd door de hoogste respectievelijk laagste score door de desbetreffende docent gegeven. Uit de afbeelding valt over het algemeen een (bescheiden) verkleining van de beoordelingsgebieden af te lezen.

Per saldo kan het resultaat van deze calibratietraining geslaagd worden genoemd en model staan voor het merendeel van de in de loop der tijd gehouden trainingen. Dit is met name het geval voor het gewenste effect, te weten een verkleining van de beoordelaarsverschillen, gemiddeld zowel als per criterium.

De schrijvers zeggen dank aan de collega's die het manuscript van kritisch commentaar hebben voorzien, in het bijzonder Prof. Dr. S. K. Thoden van Velzen, alsmede aan de audio-visuele dienst voor de tekeningen.

#### Summary:

Title: Changes in a preclinical practical course in dental skills. Part III. Functioning of the Personalized System of Instruction (2).

This paper, the third of three parts, continues the discussion of the functioning of the PSI-based preclinical practical course at the University of Amsterdam School of Dentistry.

Research results concerning the opinions of students and instructors, the value of assessments by students of their own work, and the calibration-training of instructors are discussed.

#### Literatuur:

1. Frick, T., Semmel, M. I. (1978): Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research* 48: 157-184.
2. Glass, G. V., Stanley, J. C. (1970): *Statistical methods in education and psychology*. Englewood Cliffs (N. J.), Prentice-Hall.
3. Hoekman, K. (1975): Onderzoek naar de (zelf-)beoordeling op een tandheelkundig practicum. In: *Empirische studies over onderwijs-18*. Onderwijsresearchdag 1974.) Groningen, Tjeenk Willink.
4. Kirk, R. E. (1968): *Experimental design: Procedures for the behavioral sciences*. Belmont (Cal.), Brooks/Cole.
5. Meer, A. van der, Plomp, Tj. (1977): Kontekstevaluatie van individuele studiesystemen. *Pedagogische Studiën* 54: 235-244.
6. Steures, R. W. R., Tromp, Th. J. M. (1978): Vernieuwing van een practicum voor tandheelkundige handvaardigheden. I. Opzet en invoering. *Ned Tijdschr Tandheelkd* 85: 421-426.
7. Steures, R. W. R., Tromp, Th. J. M. (1980): Vernieuwing van een practicum voor tandheelkundige handvaardigheden. II. Functioneren van het Individuele Studie Systeem (1). *Ned Tijdschr Tandheelkd* 87: 225-230.
8. Tromp, Th. J. M. (1979): Het oordeel van studenten in een Individueel Studiesysteem. *Onderwijsresearchdagen 1979*, Nijmegen

Februari 1980.

Louwesweg 1,  
1066 EA Amsterdam.