

ONDERWIJS

HET BEOORDELEN VAN TANDHEELKUNDIGE VAARDIGHEDEN: EEN SYSTEMATISCHE BENADERING

J. E. WIEGMAN

Uit de vakgroep Parodontologie-Prothodontie-Sosiodontie van de rijksuniversiteit te Groningen.

Trefwoorden: Onderwijs – Beoordelen

1. Inleiding

Het leren van technische vaardigheden neemt een belangrijke plaats in in het tandheelkunde-onderwijs. In de regel gaat men met behulp van toetsen na of een student de doelstelling van een vak heeft bereikt.

Het beoordelen van werkstukken is een standaardonderdeel van het onderwijs. De vraag is dan ook gerechtvaardigd in hoeverre de beoordelingen objectief geschieden en een positieve functie vervullen in het leerproces van de student. Beoordelen moet meer zijn dan het selecteren van studenten, het moet allereerst gericht zijn op het (bij-)sturen van het leerproces. De student moet steeds weten hoever hij gevorderd is met het leerproces. Wanneer de beoordeling geschiedt op basis van duidelijk omschreven criteria dan is het mogelijk de student terugkoppeling te geven over zijn werk. Het spreekt welhaast vanzelf dat deze criteria van meet af aan bekend moeten zijn en ook in de oefenfase van de student al door de staf gehanteerd moeten worden.

Het belang van expliciete beoordelingscriteria kan wellicht nog het beste worden verduidelijkt aan de hand van een voorbeeld. Ter illustratie moge dienen de, wellicht voor velen herkenbare, wijze waarop tot voor enkele jaren het modelleren van een kroon in was in zijn werk ging binnen de vakgroep Parodontologie, Prothodontie

en Sosiodontie (P.P.S.) van de rijksuniversiteit te Groningen. De instructie bestond in feite uit niet veel meer dan de mededeling dat de student geacht werd zich het modelleren in was eigen te maken en dat na een aantal practica getoetst zou worden of hij daarin geslaagd was. Tijdens de toets werden de kronen beoordeeld op een tienpuntsschaal waarbij een zes of hoger voldoende was. Aan het begin van de practica werd de te verwerven vaardigheid door een assistent gedemonstreerd en in een vitrine stonden een aantal gemodelleerde kronen ten voorbeeld gesteld. Tijdens het oefenen waren er assistenten in de practicumzaal om de student te adviseren en te helpen. Bij ontstentenis van andere informatie moest de student aan het werk gaan op basis van het beeld dat hij zich tijdens de demonstratie had gevormd. Bij twijfel over de juiste vorm kon hij de assistent om advies vragen. Daar er niets op schrift stond was dit de enige informatiebron. Ook de assistent was echter ernstig gehandicapt door het ontbreken van expliciete criteria waardoor hij lang niet altijd in staat was consistente en relevante informatie te verstrekken aan de student, met als gevolg verwarring bij deze laatste.

Afbeelding 1, 2 en 3 geven karikaturaal de gevolgen van een dergelijke begeleiding weer.

Het zal geen nader betoog behoeven dat er sprake is van verloren tijd wanneer de

Samenvatting:

De beoordeling van tandheelkundige werkstukken binnen de vakgroep P.P.S. van de rijksuniversiteit te Groningen geschiedt aan de hand van vaste items en criteria. De intra- en interbeoordelingsbetrouwbaarheid van dit systeem is onderzocht. Er werd een intrabeoordelingsbetrouwbaarheid gevonden die lag tussen de 70% en 90%. De interbeoordelingsbetrouwbaarheid varieerde van 25% voor de totale groep deelnemende assistenten tot ongeveer 75% wanneer slechts vijf van de zeven in ogenschouw werden genomen.

Nagegaan is ook of het beoordelingsresultaat werd beïnvloed door bepaalde assistenten respectievelijk items.

De conclusie van dit onderzoek is dat de criteria eenduidig moeten worden omschreven en dat de training van de assistenten in het toepassen van deze criteria dient te worden verbeterd.

oefenfase verloopt zoals weergegeven in de afbeeldingen. De terugkoppeling die de assistenten verstrekken helpt de student totaal niet bij zijn leerproces maar zal dit zelfs ondermijnen en de motivatie van de student negatief beïnvloeden. Als het onderwijs op deze manier verloopt heeft het veel weg van het meedoen aan een kansspel.

Hinkelman en Long¹ concludeerden ook, evenals Glaser² in zijn 'Basic teaching model', dat het beoordelen een zeer belangrijke component is van het onderwijsleerproces. Als de methode om de door de student gemaakte werkstukken te beoordelen niet deugt of betrouwbaar is dan valt niet na te gaan of de student de doelstelling van het practicum heeft bereikt noch kan de student op adequate wijze terugkoppeling



Afb. 1. Als een assistent op deze manier terugkoppeling geeft, zal de student na het aanbrengen van wat was, vrij snel om het volgende commentaar vragen. Zie afb. 2.



Afb. 2. Het zal er voor de student niet duidelijker op worden als de assistent hem op deze manier begeleidt. De student zal het dan snel eens bij een andere assistent proberen. Zie afb. 3.

Tabel I. Voorbeeld van een beschrijving van een preklinische practicumoefening: 'Het modelleren van een kroon in was'.

1. Doelstelling

Gegeven een Frasaco®-element voorzien van een MOD-preparatie, waarin ook de knobbels zijn betrokken, geplaatst in een fantoomkaak, moet de student in staat zijn maximaal twee uur dit element weer op te bouwen met was tot de oorspronkelijke vorm. Het wasmodel moet bovendien in een juiste relatie met antagonist en buurelementen zijn gemodelleerd. Het werkstuk wordt beoordeeld aan de hand van een lijst, die 25 items bevat. Een score van 20 of meer items goed is in de beoordeling van de toets voldoende.

2. Onderwijsmateriaal

a. Audiovisueel

Er is een videoband beschikbaar waarop het modelleren van de kroon in was wordt gedemonstreerd en waarop tevens het gebruik van de criteria is te zien. Deze band kan de student op elk moment bekijken. Bovendien kan hij tijdens de oefenfase de assistenten vragen om nadere uitleg van de criteria die voor bepaalde items gelden.

b. Syllabus

Een beschrijving van de fasen die kan worden gevolgd bij het modelleren van een kroon.

c. Modellen

Er zijn wasmodellen van vijf maal ware grootte aanwezig, die de verschillende fasen in het modelleren van de kroon laten zien.

3. Itemlijst

Algemeen	Linguaal aanzicht
1. De Frasaco®-kaak	14. hoogte mesiale randlijst
2. fissuurpatroon	15. mesiaal contactpunt
3. knobbelhelling	16. hoogte distale randlijst
4. fossa triangularis mesiaal	17. distaal contactpunt
5. randlijst mesiaal	18. hoogte linguale knobbels
6. contactpunt mesiaal	19. breedte linguale knobbels
7. contour mesiaal	20. hoogte buccale knobbels
8. fossa triangularis distaal	21. breedte buccale knobbels
9. randlijst distaal	
10. contactpunt distaal	
11. contour distaal	
12. buccale knobbel-punten	
13. linguale knobbel-punten	
	22. buccale contour
	23. linguale contour

Wasafwerking

24. glad wasoppervlak
25. randaansluiting was

4. Criteria bij het beoordelen van de items

De items worden beoordeeld met goed of fout. De criteria waarop de items beoordeeld worden zijn beschreven. De instrumenten die bij het beoordelen van deze criteria gebruikt worden zijn aangegeven. De items worden met behulp van de volgende criteria beoordeeld:

Item 1. Alle elementen moeten in de kaak zijn geschroefd, met uitzondering van de 46, die met was is opgebouwd. De kaak moet schoon zijn, en er mogen geen wasresten op voorkomen.

Item 2. De fissuur moet kruisvormig zijn. De mesiodistale fissuur eindigt in de mesiale en distale fossa triangularis en ligt op een lijn die getrokken kan worden door de mesiodistale fissuren van de buurelementen.

Item 3 enz. enz. tot en met item 25.

worden gegeven teneinde hem te helpen het werk(stuk) te verbeteren. Iron³ beweert zelfs dat terugkoppeling de enige variabele is, die het leren van een vaardigheid bepaalt.

Een oneerlijk toetsingssysteem beïnvloedt bovendien het moreel van de studenten negatief.⁴ Door Abou Rass⁵ werden acht, onderling samenhangende eisen geformuleerd waaraan een beoordelingssysteem moet voldoen. Eén van de eisen is dat er criteria en normen voor de te vervaardigen werkstukken moeten zijn opgesteld.

Kortom, om het niveau dat in het leerproces bereikt is te kunnen bepalen is een

beoordelingssysteem gebaseerd op duidelijke criteria noodzakelijk. Duidelijk geformuleerde en op schrift gestelde criteria zijn voor de beoordelingsbetrouwbaarheid van de assistent van groot belang. Een lage beoordelingsbetrouwbaarheid wordt onder andere veroorzaakt door onduidelijk geformuleerde criteria.⁶

Bovenbeschreven overwegingen en onderzoekgegevens zijn voor de vakgroep P.P.S. aanleiding geweest om het gebruik van eenduidig geformuleerde en op schrift gestelde criteria in te voeren. In het hierna volgende wordt beschreven op welke wijze

dit is gebeurd en hoe betrouwbaar de criteria door de assistenten worden gehanteerd.

2. Het beoordelingssysteem

Het beoordelingssysteem dat door de vakgroep P.P.S. is opgezet is gebaseerd op de beoordeling van een aantal punten die te onderscheiden zijn bij een bepaald werkstuk. Deze punten, items genaamd, worden beoordeeld met goed of fout, andere kwalificaties zijn niet mogelijk.

Men heeft het gewenste niveau bereikt wanneer een vooraf vastgesteld aantal van het totale aantal items als zijnde goed is beoordeeld. De criteria waaraan een item tenminste moet voldoen voor het predikaat goed zijn op schrift gesteld en bekend bij staf en studenten. De individuele verschillen in de beoordeling van een werkstuk door de assistenten wordt op deze wijze zo veel mogelijk beperkt.

Tabel I geeft weer, wederom met het modelleren van een (partiële) kroon in was als voorbeeld, hoe dit systeem is uitgewerkt. Beschreven worden daarin de doelstellingen van de oefening, het aanwezige instructiemateriaal, de criteria voor de items en de norm voor het behalen van de toets.

3. De evaluatie van het beoordelingssysteem

Zoals al werd gememoreerd is het nieuwe systeem onder meer ingevoerd om te komen tot een objectievere beoordeling van



Afb. 3. De kritiek van de andere assistent kan de student geheel in verwarring brengen. De kroon was te groot en twee minuten later is dezelfde kroon te klein.

		assistenten						
		A	B	C	D	E	F	G
items	1	+	+		+	+		+
	2							
	3							
	4	+	+	+	+	+	+	+
	5		+	+			+	+
	6	+						+
	7	+	+				+	+
	8			+	+	+	+	+
	9	+				+	+	
	10			+		+		
	11	+	+	+	+	+	+	+
	12		+	+		+	+	+
	13			+	+		+	
	14	+						
	15					+		+
		7	6	6	4	4	2	5
		5	5	9	8	6	6	3
		Totale score						

Tabel II. De beoordeling van één van de werkstukken van oefening I. De resultaten van de eerste beoordeling zijn in de linkerkolom weergegeven en van de tweede beoordeling in de rechterkolom. + betekent dat het item als zijnde goed is beoordeeld.

de door studenten vervaardigde werkstukken. Getracht is in de reële onderwijs situatie na te gaan in hoeverre deze opzet geslaagd is. Onderzocht is in welke mate assistenten in staat zijn de diverse criteria consistent te hanteren. Er is sprake van een objectief beoordelingssysteem wanneer de intra- en interbeoordelingsbetrouwbaarheid hoog zijn. Aan de evaluatie van het beoordelingssysteem is door zeven stafleden meegewerkt die allen assisteerden bij een bepaald preklinisch practicum. De assistenten beoordeelden elk in totaal dertien, door studenten tijdens een toets vervaardigde werkstukken; zeven werkstukken van oefening I en zes van oefening II.

De eerste oefening werd op 15 items beoordeeld en de tweede op 25. Alle werkstukken werden door de assistenten twee maal beoordeeld en wel met een tussenpoos van minimaal twee uur.

4. De evaluatieresultaten

De verschillende assistenten blijken een werkstuk lang niet altijd hetzelfde te beoordelen. Tabel II, waarin is opgenomen de beoordeling van één van de zeven werkstukken van oefening I toont dit duidelijk aan. In dit voorbeeld loopt de totaalscore van het werkstuk (de linkerkolommen) van 3 tot 9. Neemt men ook de tweede beoordelingsronde in ogenschouw dan blijkt dat er items zijn die door de ene assistent twee maal met goed worden beoordeeld terwijl ze door de andere tot twee maal toe als fout worden bestempeld (b.v. item 1 door assistent A en B). De intra- en interbeoordelingsbetrouw-

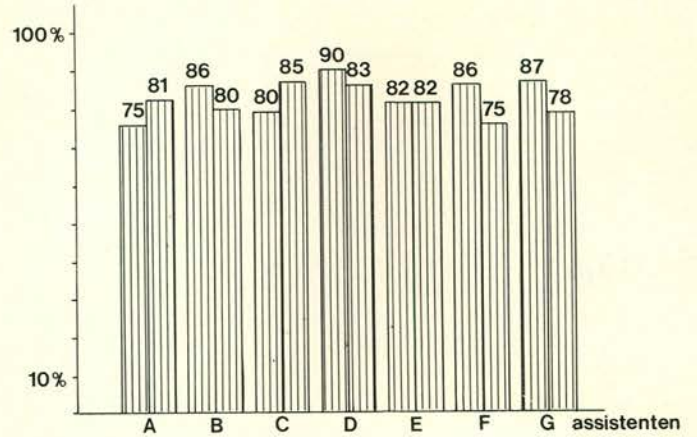
baarheid van de zeven assistenten is berekend op basis van de tabellen die analoog aan tabel II zijn opgesteld.

4.1. De intrabeoordelingsbetrouwbaarheid

De intrabeoordelingsbetrouwbaarheid is de mate waarin twee beoordelingen van een zelfde werkstuk door dezelfde assistent overeenstemmen.

Voor de zeven werkstukken van oefening I konden de assistenten maximaal 105 keer een item twee maal hetzelfde beoordelen. De zeven werkstukken moesten immers steeds op 15 items worden beoordeeld. De resultaten van de verschillende assistenten blijken duidelijk te verschillen (tabel III, linkerkolommen).

Assistent A bijvoorbeeld bereikte een intrabeoordelingsbetrouwbaarheid van 75% terwijl assistent D een score van 90% haalde. Er is ook geëvalueerd of de ene assistent bepaalde items met een hogere mate van betrouwbaarheid beoordeeld dan een andere. Uit tabel IV blijkt dat er duidelijk verschillen zijn bij item 2: assistent E geeft vier van de zeven keer niet dezelfde beoordeling terwijl assistent F bij herbeoordeling van dit item steeds hetzelfde judicium geeft. De intrabeoordelingsbetrouwbaarheid van de assistenten voor de werkstukken van oefening II varieert van 85% voor assistent A tot 75% voor assistent F (tabel III, rechterkolommen). De intrabeoordelingsbetrouwbaarheid van de assistenten voor de werkstukken van oefening II varieert van 85% voor assistent A tot 75% voor assistent F (tabel III, rechterkolommen).



Tabel III. De intrabeoordelingsbetrouwbaarheid van elke assistent in procenten. De resultaten van oefening I zijn in de linkerkolom weergegeven en van oefening II in de rechterkolom.

		assistenten						
		A	B	C	D	E	F	G
items	1	1	1					
	2	2	2	2	3	4		1
	3	1	1	1	1		2	1
	4	1		4		2	2	3
	5	2		1	1	1	2	
	6	3	1			1	1	1
	7	4	1	3		2	1	
	8			1		2	2	
	9	2				1	2	
	10		2	1	1			2
	11	2	2	1	1	2		1
	12	2	1			2	1	2
	13	1	2	2			1	
	14	2		2				
	15	3	2	3	3	1		4

Tabel IV. De intrabeoordelingsbetrouwbaarheid van de assistenten is voor de diverse items van oefening I verschillend. De cijfers in de verschillende kolommen geven weer het aantal keren, dat het betreffende item met een goed-fout score werd beoordeeld, na de twee beoordelingsronden.

De maximale goed-fout score is 7, d.w.z. minimale overeenkomst.

van de assistenten voor de verschillende items van oefening II geeft een beeld dat vergelijkbaar is met dat van oefening I (tabel V).

4.2. De interbeoordelingsbetrouwbaarheid

De interbeoordelingsbetrouwbaarheid is

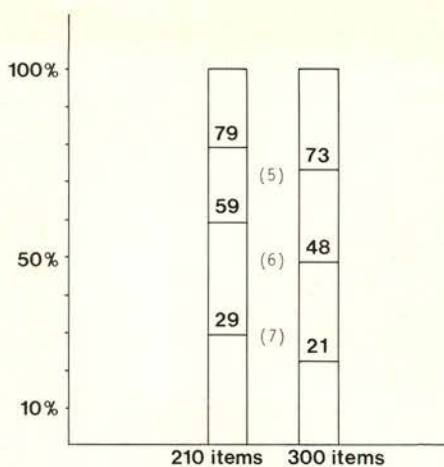
		assistenten						
		A	B	C	D	E	F	G
1	3			2	1	2	1	
2		2		1		2	2	
3	2		4	1		1	1	
4	2		2	1		3		
5	1	2	1	1	1	4	2	
6	1	1	1		1	1	2	
7	1	2	2		2			
8	1		1	1	2	3	3	
9	1	4		2	1	2	1	
10		2		1		3	2	
11	1	2	2	1	2	2	2	
12				1	2	2	2	
13	3	2	1	2	3	1	1	
14		3	2	1	3	1	5	
15		1			2	1		
16	1	2	1	1		1	1	
17				1	1	1	1	
18	2			1	2			
19	1		1	1	1	2	1	
20	2	1		2	1	1	2	
21	2	1	1			1	1	
22		2	1		1	1	1	
23	1	2		1	1			
24	1		1	1			1	
25	1	1		3		2	1	

Tabel V. De intrabeoordelingsbetrouwbaarheid van de assistenten is voor de diverse items van oefening II verschillend. De cijfers in de verschillende kolommen geven weer het aantal keren, dat het betreffende item met een goed-fout score werd beoordeeld, na de twee beoordelingsronden. De maximale goed-fout score is 6 d.w.z. minimale overeenkomst.

de mate waarin de beoordelingen van assistenten overeenstemmen wanneer ze allemaal hetzelfde werkstuk krijgen voorgelegd.

Voor de items van de werkstukken van oefening I zou een interbeoordelingsbetrouwbaarheid van 100% betekenen dat de zeven assistenten de 105 items zowel in de eerste als in de tweede ronde allemaal gelijk beoordelen. Voor de tweede oefening betreft het 150 items die twee maal door allen hetzelfde beoordeeld moeten worden, wil er sprake zijn van 100% overeenstemming.

In tabel VI zijn de feitelijke resultaten weergegeven; wanneer alle zeven assistenten in ogenschouw worden genomen dan blijkt de interbeoordelingsbetrouwbaarheid bij oefening I 29% en bij oefening II 21% te bedragen. Laat men een assistent buiten beschouwing dan stijgt de betrouwbaarheid tot 59% respectievelijk 48%. Berekend over vijf assistenten wordt de score 79% voor de eerste oefening en 73% voor de tweede.



Tabel VI. Het percentage van interbeoordelingsbetrouwbaarheid voor oefening I is weergegeven in de linkerkolom en van oefening II in de rechterkolom. Bovendien zijn de scores aangegeven wanneer er (5), (6) of (7) assistenten meededen.

oordeling gaf die afweek van zijn zes collega's. Rechts binnen de kolommen staat vermeld hoe vaak de beoordeling afweek van vijf anderen en m.a.w. dus overeenkwam met (slechts) één andere assistent. In deze tabel valt op dat assistent A een erg lage interbeoordelingsbetrouwbaarheid had voor item 6. In acht van de veertien keer gaf hij beoordelingen die afweken van vijf andere assistenten. Assistent G had problemen met item 10; acht van de veertien keer beoordeelde hij dit item op een wijze die verschilde van de overige zes. De uitkomsten op dit punt voor oefening II zijn te vinden in tabel IX. In deze tabel valt op dat assistent C een erg lage interbeoordelingsbetrouwbaarheid scoorde bij item 24 terwijl assistent B bij item 23 nogal eens alleen stond in zijn oordeel.

5. Nabeschuiving

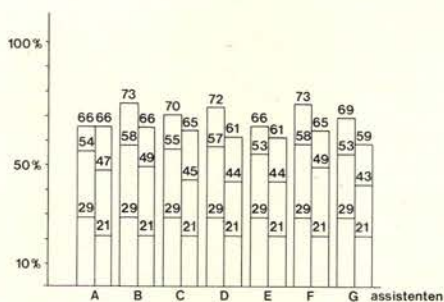
De situatie zou ideaal zijn wanneer de intra- en de interbeoordelingsbetrouwbaarheid 100% zou bedragen. In dit onderzoek varieerde de intrabeoordelingsbetrouwbaarheid van 75% tot 90% en de interbeoordelingsbetrouwbaarheid van 21% tot 79%, afhankelijk van de scores die werden samengenomen. De gevonden resultaten komen overeen met die welke men in de literatuur aantreft.^{1 5 7 9}

De betrokken assistenten vonden de resultaten echter enigszins teleurstellend. Dit is ook wel begrijpelijk wanneer men bedenkt dat alle assistenten enthousiast waren over de wijze waarop het onderwijs werd gegeven. Verder hadden zij zich voor de aanvang van het onderwijs middels een training van een week voorbereid op de cursus. Gedurende deze training werd het onderwijsprogramma naar inhoud en vorm doorgenomen evenals de criteria die bij het beoordelen van de verschillende items zouden worden gebruikt.

Voor de betrokken assistenten was het resultaat des te meer teleurstellend omdat zij op het moment dat het hier beschreven onderzoek werd uitgevoerd de studenten al zes maanden hadden begeleid.

Een positief effect van het onderzoek is dat de vakgroep P.P.S. nu over cijfermateriaal beschikt, hetgeen de discussie over het gebruik van criteria bij het beoordelen in een stroomversnelling heeft gebracht. De staf bleek bereid geheel open de discussie over deze materie te voeren.

Duidelijk werd dat de vakgroep de criteria beter moet gaan formuleren opdat een betrouwbaarder beoordelingssysteem dan het nu gehanteerde kan ontstaan. Ook kwam de vraag aan de orde of het mogelijk was een beter trainingsprogramma te ontwikkelen dan het huidige. Dit zou dan gericht moeten zijn op een identiek leren gebruiken van de criteria voor de verschillende items door alle assistenten, teneinde zo de intra- en interbeoordelingsbetrouw-



Tabel VIII. Aangegeven is het aantal keren, dat een assistent een item verschillend beoordeelde ten opzichte van zes andere assistenten (links in de kolom) of vijf anderen (rechts in de kolom).

De interbeoordelingsbetrouwbaarheid is ook geëvalueerd voor de individuele assistenten. Bij de werkstukken van oefening I komt 29% van de items van de ene assistent overeen met die van de zes anderen. Laat men de assistent die de meest afwijkende beoordelingen geeft buiten beschouwing dan stijgt de betrouwbaarheid afhankelijk van de assistent die als uitgangspunt wordt genomen van 53% tot 58%. Laat men de twee meest afwijkende assistenten buiten beschouwing dan varieert dit percentage van 66 tot 73 (tabel VII, linkerkolommen). De scores voor oefening II treft men aan in de rechterkolommen van tabel VII.

Nagegaan is ook nog in hoeverre deze resultaten werden veroorzaakt door een lage interbeoordelingsbetrouwbaarheid van één of twee assistenten bij bepaalde items. Hiervoor werd bepaald welke assistent voor welke items een beoordeling gaf die afweek van de zes overigen. De resultaten zijn voor oefening I weergegeven in tabel VIII. Links in de kolom staat het aantal keren aangegeven dat de assistent een be-

		assistenten						
		A	B	C	D	E	F	G
items	1	2	1 1	1 1				
	2			1 1	1	1	1	1
	3	1		2	2			1 2
	4	1 2		3 2	1		1 2	3 3
	5	3 1				1	1 1	
	6	5 3	2 2	2	1 1			1
	7	1 2	1		1	1		1
	8	1	1			2		1
	9	1				2		
	10		1 1	1 1				8
	11	3	1	1	1 2	1		
	12	1	1			4	2 3	
	13		2	2 1	1 1	2		
	14	2 1		2 2				1
	15				2 1			1
		12 18	4 10	10 12	6 10	14 16	4 10	14 10

Tabel VII. De percentages van interbeoordelingsbetrouwbaarheid voor elke assistent met resp. (6), (5) of (4) andere assistenten. De linkerkolom geldt voor oefening I en de rechterkolom voor oefening II.

baarheid te verbeteren. Helaas zijn de resultaten uit de literatuur over het effect van trainen niet eensluidend.⁵⁻⁷

Momenteel wordt overwogen een trainingsprogramma te ontwikkelen dat individueel en op elk gewenst tijdstip door de assistenten kan worden gevolgd. Het is mogelijk trainingsmateriaal te verkrijgen door het verzamelen van werkstukken. Deze moeten dan worden beoordeeld door een vakexpert. Deze 'expertbeoordelingen' zouden kunnen dienen als referentiekader. Wanneer men dit materiaal met de bijbehorende beoordelingen in een computerprogramma invoert kan de individuele assistent het gebruik van de criteria op elk gewenst moment oefenen. In eerste instantie beoordeelt hij dan bijvoorbeeld tien werkstukken uit de verzameling waarna de computer de intra- en interbeoordelingsbetrouwbaarheid t.o.v. de ingevoerde sleutel berekent. De assistent krijgt vervolgens terugkoppeling over zijn gebruik van de criteria en kan zijn wijze van beoordelen verbeteren door de criteria voor bepaalde items nog eens te bestuderen of door deze anders toe te leren passen. Door vervolgens nog eens een aantal werkstukken te beoordelen kan hij met behulp van het computerprogramma opnieuw nagaan of zijn beoordelingsbetrouwbaarheid is toegenomen.

Een geheel ander discussiepunt blijft of het mogelijk is de items van een werkstuk afzonderlijk te beoordelen zonder het totaalbeeld in de beoordeling te betrekken. Het is

heel wel denkbaar dat een slechte of goede eerste totaalindruk van invloed is op de beoordeling van de afzonderlijke items. Dit lijkt echter niet de belangrijkste oorzaak voor de lage scores; immers de problemen met de afzonderlijke items zijn verschillend voor de verschillende assistenten (tabel IV, V, VIII en IX).

Het onderzoek heeft ook de discussie weer op gang gebracht over de vraag hoeveel items er als goed beoordeeld moeten zijn om voor het werkstuk geheel een voldoende te krijgen. Met andere woorden: waar dient de norm of cesuur te liggen. Deze kwestie blijkt aan de hand van een beoordelingslijst met items beter bespreekbaar te zijn.

Niet onderzocht is in dit kader of en in hoeverre de beoordeling van een assistent wordt beïnvloed doordat hij weet van wie een bepaald werkstuk is of doordat hij weet dat hetgeen hij voorgelegd krijgt een oefening dan wel een examenwerkstuk betreft. Wel is bekeken of een beter resultaat zou kunnen worden verkregen indien het beoordelen van de toetsen door twee stafleden geschiedt. In een afzonderlijk artikel zal daarover in dit tijdschrift nog verslag worden gedaan.

Een hamvraag die opkwam was natuurlijk of een betrouwbaarheid van 70% acceptabel is in een onderwijssituatie. Anders geformuleerd: leert een student de criteria goed als de begeleidende assistent op een dergelijk 70%-niveau werkt? Ofschoon een betrouwbaarheid van 100% zonder

		assistenten						
		A	B	C	D	E	F	G
items	1			2				
	2	3			1 1			
	3	2	1 2	1 1		2 1	1	
	4	1	1	1		1	1	1
	5		1		2 1		2 2	
	6				1	2 3	1 2	2
	7	1	3	1	1			3 2
	8	1		1	4 1	2		2 1
	9		1 3		2 4	1	1 1	
	10	2	2			1 2	1 2 3	
	11		1			3	1 2	1 2 2
	12				1	2	1 3	2 1
	13	1 1	1	1	1	1	2	1
	14			2	1	2		1 3
	15			1		1 2		1 2
	16	1	2				1	
	17		2			1 1	3 3	
	18	2	2		1	2 2	1	
	19			1		2	3	
	20	2 1	1 1		2	1	1	
	21	1 1		2	1		2	3
	22		1	1	1 4	1		4
	23	1 1	4 2		1 1		1	
	24	1		5 1				
	25	1		1	1			1 1

Tabel IX. De resultaten van oefening II van dezelfde evaluatie als weergegeven in tabel VIII. Aangegeven is het aantal malen, dat een assistent een item verschillend beoordeeld ten opzichte van zes andere assistenten (links in de kolom) of vijf anderen (rechts in de kolom).

meer ideaal zou zijn blijft de vraag of dit ook noodzakelijk is. Bij de beantwoording van deze vraag dient men zich te realiseren dat de studenten in de gekozen onderwijsopzet het gebruik van de criteria leren door een combinatie van verschillende onderwijsleermiddelen, waarvan de assistent er slechts één is.¹¹ Dit neemt niet weg dat het de taak van de vakgroep blijft het onderwijsproces te onderzoeken en de instructiemethoden te verbeteren om erop deze manier voor te zorgen dat de beschikbare studietijd door de student op een zo effectief mogelijke wijze kan worden gebruikt.

De auteur is dank verschuldigd aan Prof. Dr. A. C. M. van de Poel en Ir. E. H. Vermeer voor de met hen gevoerde discussies die veel hebben bijgedragen aan het tot stand komen van dit artikel.

Summary:

Title: Assessment of dental skills: a systematic approach.

Keywords: Education - Assessment

The object of this study was to evaluate the assessment system of dental skill performances using items based on described criteria.

The performance items were judged as either right or wrong. The intra- and inter examiner reliability was determined when the instructors used this assessment system. The figures of this investigation have an intra examiner reliability

of 70% to 90%. The scores for the inter examiner reliability of the item assessment varied from about 25% for the total group participating instructors to about 75% for a group of five participating instructors. It was also investigated as to whether these were affected by instructors and specific items.

The conclusion of this evaluation is that, the department must not only improve the criteria for the items but must also improve the training of the instructors to use them.

Literatuur:

1. Hinkelman KW, Long NK. Method for decreasing subjective evaluation in preclinical restorative dentistry. *J Dent Educ* 1973; 37: 13-18.

2. Glaser R. Psychology and instructional technology. In: Glaser R (Ed.): Training research and education. Pittsburg, 1962.
3. Irion AL. A brief history of research on the acquisition of skill. In: Bilodeau EA (Ed.): Acquisition of skill. New York: Academic Press, Inc., 1965.
4. Nedelsky L. Science teaching and testing. Chicago: Harcourt, Brace and World, Inc., 1965.
5. Abou Rass M. A clinical evaluation instrument in endodontics. *J Dent Educ* 1973; 37: 22-36.
6. Natkin E, Guild RE. Evaluation of preclinical laboratory performance: A systematic study. *J Dent Educ* 1967; 31: 152-161.
7. Houpt MI, Kress G. Accuracy of measurement of clinical performances in dentistry. *J Dent Educ* 1973; 37: 34-46.
8. O'Conner P, Lorey R. Improving interrater agreement in evaluation in dentistry by use of comparison stimuli. *J Dent Educ* 1978; 42 (4): 174-179.
9. Patridge MA, Mast TA. Dental clinical evaluation: A review of the research. *J Dent Educ* 1978; 42 (6): 300-305.
10. Fuller JL. The effects of training and criteria models in interjudge reliability. *J Dent Educ* 1972; 36: 19-22.
11. Vermeer EH. Evaluatie van het waspracticum. Bulletin 15. Afdeling Onderwijs Subfaculteit Tandheelkunde te Groningen. Interne publikatie, 1979.

Augustus 1982.

Antonius Deusinglaan 1,
9713 AV Groningen.

BERICHTEN

Verenigingsverslagen en mededelingen

NEDERLANDSE VERENIGING VOOR ORTHODONTISCHE STUDIE

Aankondiging najaarsvergadering d.d. 4 november 1983

De najaarsvergadering van de Nederlandse Vereniging voor Orthodontische Studie zal dit jaar worden gehouden op vrijdag 4 november a.s. (aanvang 16.00 uur) in het Postiljon Motel Nulde-Putten.

's Middags zal Prof. Dr. R. C. W. Burgersdijk (afdeling Kindertandheelkunde Katholieke Universiteit te Nijmegen) het onderwerp 'Kindertandheelkunde en Orthodontie' behandelen. Tijdens de avondbijeenkomst zullen Dr. R. Brons (kaakchirurg) en R. M. Mulié (orthodontist) spreken. Hun gemeenschappelijk onderwerp is 'Geïntegreerde multidisciplinaire behandeling van dentofaciale afwijkingen', een onderwerp dat door hen als volgt wordt samengevat.

Onze groep die zich bezighoudt met de behandeling van dentofaciale afwijkingen bestaat uit een orthodontist, een kaakchirurg en een algemeen-practicus met speciale belangstelling voor gnathologie en parodontologie. Ondersteunende therapie wordt gegeven door de mondhygiëniste en de logopediste.

De patiënten welke geacht worden voor gezamenlijke behandeling in aanmerking te komen zijn die, welke niet door één van de disciplines zelfstandig naar tevredenheid kunnen worden geholpen.

Gezamenlijk wordt een beschrijving van de afwijking gegeven, aan de hand van criteria die betrekking hebben op de esthetiek, cefalometrie, gnathologie, parodontologie, de gebitstoe-

stand en een al of niet afwijkend groeipatroon. Dit leidt tot een zgn. *causaal omschreven diagnose*, waarbij naast de statische evaluatie van de afwijking ook de oorzakelijke storing in het groeipatroon wordt aangegeven. De volgende stap is het opstellen van een gezamenlijk, geïntegreerd behandelplan, dat in extenso met de patiënt wordt besproken. Eén van de doelstellingen hierbij is de zgn. *blijvende herkenbaarheid van de patiënt*. Voor iedere patiënt afzonderlijk wordt bepaald welke orthodontische, chirurgische en algemeen-tandheelkundige therapie noodzakelijk is en de volgorde, al dan niet met overlappingen, waarin deze dienen plaats te vinden.

In het algemeen geldt het volgende schema: opheffen van gnathologische en parodontologische afwijkingen, zo nodig goede tijdelijke prothetische voorzieningen, orthodontische voorbehandeling, chirurgische interventie(s), orthodontische nabehandeling en definitieve tandheelkundige maatregelen, zoals inslijpprocedures, zo nodig gevolgd door logopedische therapie.

J. M. Beek

NEDERLANDSE VERENIGING VOOR MONDZIEKTEN EN KAAKCHIRURGIE

Jaarverslag 1982

In 1982 werd de wetenschappelijke voorjaarsvergadering op 24 april 1982 te Utrecht gehouden, maar de gecombineerde vergadering met het Collegium Chirurgicum Neerlandicum vond dit jaar niet plaats.

Er werd speciaal aandacht besteed aan de gecombineerde behandeling van chronische osteomyelitis met hyperbare zuurstof, chirurgische en antibiotische therapie. Als hoogtepunt moet de voordracht van collega Bakker van het Academisch Medisch Centrum van de Universiteit

van Amsterdam worden vermeld.

Deze goed bezochte vergadering werd gecombineerd met een huishoudelijke vergadering.

De najaarsvergadering werd door de afdeling Mondziekten en Kaakchirurgie van de Vrije Universiteit te Amsterdam georganiseerd. Deze vergadering, op 12 en 13 november, werd druk bezocht. Er waren een groot aantal voordrachten. De inmiddels traditionele festiviteiten werden op vrijdagavond 12 november in Heemstede gehouden.

Het Bestuur heeft zich, samen met een daartoe ingestelde commissie, in de verslagperiode o.a. bezig gehouden met de taak van de kaakchirurgen op het gebied van de parodontologie. Uit een hiertoe gehouden enquête werd onder meer geconcludeerd dat, hoewel er al voldoende kennis en kunde voor dit deel van ons vakgebied bestaat, met name de opleiding voor dit onderdeel nog de nodige aandacht behoeft.

De voorzitter werd in het verslagjaar benoemd tot hoogleraar in Adelaïde. Voorlopig zal hij daar gedurende één jaar als docent werkzaam zijn. De beste wensen van de Vereniging verzellen hem. Voor zover bij het secretariaat bekend werden door de leden van de Vereniging een vijftigtal voordrachten gehouden voor circa 4900 collegae.

Door de leden werden in binnen- en buitenlandse vaktijdschriften 58 wetenschappelijke bijdragen gepubliceerd.

J. M. Onland

NEDERLANDSE VERENIGING VAN TANDARTSEN

Dr. Th. Dentz-medaille uitgereikt aan Prof. Dr. G. Boering

Tijdens de opening van het Nationaal Tandheel-