

O N D E R W I J S

DE ONTWIKKELING VAN EEN TRAININGSPROGRAMMA VOOR HET BE-
OORDELEN VAN PREKLINISCHE, TANDHEELKUNDIGE WERKSTUKKEN

G. J. J. M. STRAETMANS

*Uit het Instituut Conserverende Tandheelkunde voor Volwassenen
van de Katholieke Universiteit te Nijmegen.**Trefwoorden: Onderwijs – Beoordelen*1. *Probleemstelling*

Evenals in het buitenland¹⁻⁴ zijn er in Nederland verscheidene onderzoeksrapporten⁵⁻⁹ waarin geconcludeerd wordt dat tandartsen niet erg consistent zijn in het beoordelen van preklinische en klinische werkstukken.

In Nijmegen concludeerde de Subcommissie 'Toetsing en Beoordeling Motorische vaardigheden' in haar verslag dat '... zowel bij het toekennen van cijfers voor, als bij het nemen van voldoende/onvoldoende-beslissingen over werkstukken (c.q. handelingen) de tussenbeoordelaarsovereenstemming op totaalniveau matig tot laag is'.¹⁰

Inconsistente beoordelingen kunnen ernstige gevolgen hebben voor het onderwijs. De voornaamste worden hier genoemd.

1. Onbetrouwbare zak/slaag-beslissingen. Als beoordelaars verschillende cijfers toekennen voor hetzelfde werkstuk, dan kan het niet anders of hun oordelen komen tot stand op basis van verschillende of zelfs irrelevante factoren. Indien het mogelijk zou zijn om op ondubbelzinnige wijze alleen de kwaliteit van een werkstuk te beoordelen, dan zou het oordeel volstrekt betrouwbaar zijn. Helaas is dit ideaal onbereikbaar en moet genoeg worden genomen met het feit dat zo nu en dan studenten ten onrechte zakken of slagen.

2. Inconsistente evaluatie veroorzaakt verwarring en maakt het leren inefficiënt. Volgens de behaviouristische leertheorie is het leren voorspelbaar en efficiënt als de gevolgen ervan betrouwbaar en consistent zijn. Zelfs eenvoudige rekvaardigheden konden niet betrouwbaar aangeleerd worden als verschillende leerkrachten een kind dat een optelsom goed uitrekende, soms beloonden en soms straffen.¹

3. Onbetrouwbare beoordelingen kunnen emotioneel en bizar gedrag van de kant van de student tot gevolg hebben.

Proefdieren confronteren met een probleemsituatie waarin oplossingen geen consistente gevolgen opleverden, bleek een zeer betrouwbare methode om experimentele neurose op te wekken.¹ Hetzelfde geldt voor tandheelkunde-studenten. Inconsistente evaluatie is een bron van frustratie en kan het leergedrag negatief beïn-

vloeden door het oproepen van ongewenste emoties.¹¹

Dit artikel bespreekt op welke wijze geprobeerd wordt om door middel van beoordeelaarstraining, de betrouwbaarheid van werkstukbeoordelingen in de prekliniek op een aanvaardbaar niveau te brengen.

2. *Het probleem nader beschouwd*

Tot nu toe is alleen gesproken over de betrouwbaarheid van beoordelingen. Beoordelingen dienen echter niet alleen betrouwbaar te zijn, maar (vooral) ook valide. Alleen dan geven ze een juiste kwalificatie aan het werkstuk.

Een valide beoordeling kan slechts op basis van een valide beoordelingsinstrument tot stand komen. Een dergelijk instrument meet wat men wil meten; in de tandheelkunde is dat de mate waarin de student die vaardigheden bezit, die nodig zijn om problemen van tandheelkundige aard op te lossen.

Nu is de relatie tussen betrouwbaarheid en validiteit van dien aard, dat een instrument betrouwbaar kan zijn zonder dat het valide is. Omgekeerd kan een onbetrouwbaar instrument nooit valide zijn. Betrouwbaarheid is dus een noodzakelijke, maar niet-voldoende voorwaarde voor validiteit.

De maximaal haalbare validiteit wordt aangegeven door de betrouwbaarheidsindex (de wortel uit de betrouwbaarheidscoëfficiënt). Om een valide instrument te construeren voor het beoordelen van tandheelkundige werkstukken, moet getracht worden de betrouwbaarheid zo hoog mogelijk op te voeren. Bij de constructie van een betrouwbaar instrument spelen de volgende factoren een belangrijke rol:

- objectiviteit;
- specificiteit;
- differentiatie.

Objectiviteit wordt bereikt wanneer de kenmerken, waarop het werkstuk beoordeeld wordt, zo duidelijk gesteld zijn, dat de persoon van de beoordelaar, mits hij zich aan de scoringsregels houdt, geen invloed kan hebben op de score van de student.

Specificiteit houdt in, dat de kenmerken van dien aard moeten zijn, dat alleen degenen, die zich de vaardigheid in voldoende

Samenvatting:

Dit artikel beschrijft hoe door middel van staftraining geprobeerd wordt om de betrouwbaarheid van beoordelingen van preklinische werkstukken te vergroten. Onder training wordt verstaan: het herhaald bezig zijn met beoordelingsactiviteiten, waarbij die vaardigheid vergroot wordt door direct terugkoppeling te geven over de geleverde beoordelingsprestatie.

Ten behoeve van de training werd een werkstukkenverzameling aangelegd.

Voor ieder werkstuk daaruit is de kwaliteit op elk beoordelingsaspect vastgesteld (referentie-oordeel). Wijkt het oordeel van een assistent af van het referentie-oordeel, dan moet de beoordeling op een gedetailleerder niveau plaatsvinden. Daarvoor wordt gebruik gemaakt van een speciaal voor dat doel vervaardigd beoordelingsprotocol.

Het trainingsprogramma functioneert op een microcomputer.

mate hebben eigen gemaakt aan de gestelde eisen kunnen voldoen.

Differentiatie betekent in dit verband dat de kenmerken een docent in staat moeten stellen om studenten, die de vaardigheid goed beheersen te onderscheiden van studenten, die de vaardigheid minder goed beheersen.

Wordt aan deze voorwaarden voldaan, dan representeren verschillen in scores tussen studenten verschillen in de gemeten vaardigheid en is het instrument betrouwbaar.

Bij preklinische en klinische evaluatie van tandheelkundige werkstukken zijn twee soorten van betrouwbaarheid van belang:

1. de intra-beoordelaarsbetrouwbaarheid (gedefinieerd als de consistentie van de beoordelaar met zichzelf).

2. de inter-beoordelaarsbetrouwbaarheid (de consistentie tussen beoordelaars).

De intra-beoordelaarsbetrouwbaarheid zegt iets over de stabiliteit van het oordeel van een individuele beoordelaar.

De inter-beoordelaarsbetrouwbaarheid betreft de interne consistentie binnen een groep beoordelaars.

Beide betrouwbaarheidsmaten zijn gebaat bij gestandaardiseerde en geobjectiveerde beoordelingsprocedures.

3. *Een mogelijke oplossing voor het probleem*

Mackenzie onderzocht welke factoren van invloed zijn op de beoordelaarsovereenstemming bij de evaluatie van klinische tandheelkundige werkstukken.¹²

Analyse van de onderzoeksgegevens leerde dat overeenstemming gereduceerd werd

door zestien factoren. De belangrijkste daarvan waren:

A. Onduidelijke criteria

– Criteria moeten wederzijds uitsluitend zijn, zodat er geen dubbelzinnigheid kan bestaan over welk criterium van toepassing is als een fout wordt ontdekt.
– Beoordelaars kunnen een verschillende ‘aanpak’ hebben, zelfs wanneer criteria operationeel gedefinieerd zijn met betrekking tot de wijze van beoordelen. Criteria moeten complete informatie geven, ten einde te voorkomen dat beoordelaars op uiteenlopende manieren gaan beoordelen.
– Dubbelzinnigheden in de interpretatie van waarnemingen die volledig steunen op visueel onderzoek, kunnen niet voorkomen worden door verbale omschrijvingen alleen.

B. Ongestandaardiseerde hulpmiddelen voor de beoordelingen

Tandheelkundig instrumentarium dat gebruikt wordt om te bepalen of werkstukken aan de gestelde eisen voldoen, kan een bron van variatie zijn. Immers niet ieder instrument is in gelijke mate aan slijtage onderhevig en instrumenten van verschillende fabrikaat zijn vaak verschillend van afmeting.

C. Inconsistentie met betrekking tot de observatiemethoden

Als de methode om te observeren niet duidelijk gespecificeerd wordt, dan zijn beoordelaars geneigd om verschillende methoden te gebruiken.

D. Ontbreken van een objectieve standaard

Als beoordelaars in verschillende universiteiten zijn opgeleid, is het goed mogelijk dat hun oordelen niet met elkaar overeenstemmen doordat hun opleiders niet dezelfde autoriteit volgden.

E. Variatie in het optreden van de welwillendheidsfout

Bedoeld wordt de verschillende neiging van beoordelaars om consequent te gunstig of te ongunstig te beoordelen (‘milde’ of ‘strengere’ beoordelaars).

Mackenzie concludeert, dat veel van de gesignaleerde problemen gereduceerd kunnen worden door gebruik te maken van ‘check-lists’, waarin omschreven wordt hoe geobserveerd moet worden en hoe de waarnemingen gecategoriseerd moeten worden met minimale interpretatie van de kant van de beoordelaars.¹² De in dit artikel voorgestelde oplossing

gaat in dezelfde richting als die van Mackenzie. Beknopt weergegeven komt dit op het volgende neer:

1. De beoordelingsprocedure wordt gestandaardiseerd. Er komen voorschriften met betrekking tot de aspecten waarop beoordeeld moet worden. In operationele beoordelingen wordt omschreven wat de eisen zijn waaraan voldaan moet worden. Verder hoe beoordeeld moet worden of aan de eisen voldaan is en hoe de beoordeling in een score uitgedrukt wordt.

2. Het subjectieve element in de beoordeling wordt gereduceerd door de voorschriften in zo objectief mogelijke bewoordingen vast te leggen.

Het standaardiseren en objectiveren van beoordelingsprocedures zijn noodzakelijke, maar helaas onvoldoende, voorwaarden voor betrouwbare beoordelingen. Training lijkt noodzakelijk om te bereiken dat alle beoordelaars het beoordelingssysteem op de bedoelde wijze hanteren. Want, praktisch bruikbare beoordelingssystemen kunnen nooit zo gestandaardiseerd en geobjectiveerd zijn, dat ondubbelzinnig gebruik ervan zonder meer verzekerd is.

In paragraaf 5 wordt beschreven op welke wijze geprobeerd werd een gestandaardiseerde en geobjectiveerde beoordelingsprocedure te vervaardigen en hoe trainings-sessies ingericht werden om beoordelaars te leren er op ondubbelzinnige wijze mee om te gaan.

Eerst, echter, volgt in paragraaf 4 een beknopte uiteenzetting over de effecten van training op de beoordelaarsovereenstemming.

4. Beoordelaarstraining

Het trainen van beoordelaars is niet nieuw; verscheidene onderzoekers hebben reeds geprobeerd om door middel van training de overeenstemming op een hoger peil te brengen.⁷ Patridge en Mast inventariseerden onderzoek op dit gebied en vonden de volgende conclusies met betrekking tot het gebruik van beoordelaarstraining:¹³

1. Natkin en Guild¹: Hoewel de trainings-sessies (beoordelen van tien werkstukken na door middel van lezen kennis te hebben genomen van het nieuwe beoordelingssysteem) geen significante bijdrage leverden aan de betrouwbaarheid van de evaluatie, lijkt het erop dat dergelijke sessies, waarin actuele evaluatieproblemen aan de orde komen, de meest geschikte methode vormen om fouten op te sporen en te classificeren.

2. Fuller¹¹: Het trainingsprogramma bestond hier uit de volgende onderdelen:

- bespreking van de gehanteerde begrippen;
- bespreking van menselijke factoren die het beoordelen beïnvloeden;
- uiteenzetting van de beoordelingsmethode;

d. beoordeling van werkstukken (vijftien per beoordelaar).

Geen bewijs werd gevonden dat beoordelingen betrouwbaarder werden door middel van training.

3. Abou-Rass³: De training omvatte zes zittingen binnen een tijdsbestek van acht weken. De nadruk lag op de definitie, betekenis, meting en evaluatie van elke ‘stap’ in de beoordelingsmethode. De overeenstemming tussen vier beoordelaars nam toe van 54 procent vóór de training tot 77 procent erna.

4. Hinkelman en Long¹⁴: Vier stafleden evalueerden dertig preparaties onafhankelijk van elkaar, gebruik makend van een nieuw beoordelingssysteem. Na een week evalueerden dezelfde vier personen dertig andere preparaties. In de periode tussen de twee evaluaties werd een trainings-sessie gehouden om de beoordelaars meer vertrouwd te maken met het nieuwe beoordelingssysteem. Slechts een geringe verbetering in overeenstemming tussen de beoordelaars werd gevonden. De onderzoekers zien daarin een reden om met enige regelmaat trainings-sessies te organiseren. Op die wijze hopen zij dat uniforme interpretatie niet alleen bereikt wordt maar ook gehandhaafd blijft.

5. Houpt en Kress⁴: Training werd omschreven als het geven van onmiddellijke terugkoppeling aan beoordelaars, nadat ze werkstukken beoordeeld hadden. De training had een klein positief effect op de intra-beoordelaarsovereenstemming bij tandartsassistenten, maar niet bij studenten en ook niet bij stafleden. Verder werden geen significante effecten gevonden met betrekking tot de inter-beoordelaars-overeenstemming.

Uit de aangehaalde literatuur blijkt hoe inconsistent de resultaten zijn van het onderzoek naar de effecten van training op de betrouwbaarheid van klinische en preklinische evaluatie.

Waarschijnlijk is dit te wijten aan de gebrekkige definiëring van de beoordelings-situaties. Onderzoek dat de elementen, waaruit de training bestaat (bijvoorbeeld soort training, aantal sessies, aantal werkstukken, hulpmiddelen), specificeert en systematisch manipuleert, is tot op heden nog niet uitgevoerd.¹³

Ondanks deze wat teleurstellende resultaten van training op de beoordelaarsovereenstemming werd toch besloten om trainings-sessies op te zetten. Reden was, dat de training behalve voor het vergroten van de overeenstemming ook voor andere doeleinden gebruikt zou kunnen worden. Bijvoorbeeld voor:

1. Zelfevaluatie van studentén.

Door zelf te beoordelen kunnen studenten een beter idee krijgen van de eisen waaraan hun werkstukken moeten voldoen.

2. De implementatie van nieuwe kwaliteits-

teitseisen.

Middels training kunnen nieuwe kwaliteitseisen sneller ingevoerd worden.

3. Het inwerken van nieuwe stafleden en student-assistenten.

Voor een deel kan het tijdrovende karwei van leren beoordelen efficiënter gebeuren door de nieuwe medewerker er in te trainen.

In het nu te bespreken project van de subfaculteit tandheelkunde te Nijmegen wordt het trainingsprogramma alleen gebruikt als middel om de beoordelaarsovereenstemming te vergroten. In die toepassing heeft training de volgende betekenis: het herhaald bezig zijn met beoordelingsactiviteiten, waarbij de vaardigheid van het beoordelen bevorderd wordt door onmiddellijke terugkoppeling te geven over de geleverde beoordelingsprestatie.

5. Een geïndividualiseerd trainingsprogramma

De titel van deze paragraaf geeft aan dat het trainingsprogramma zodanig is ingericht, dat individuele personen er gebruik van kunnen maken. Flexibel gebruik van de trainingsfaciliteit wordt hierdoor sterk bevorderd: groepsgewijze training is in het algemeen slechts beperkt mogelijk. Binnen de geplande practicumtijd moeten de docenten het merendeel van hun tijd aan begeleiding en beoordeling besteden: buiten de practicumtijd zijn de docenten nauwelijks beschikbaar.

In paragraaf 3 werd gezegd dat trainingen georganiseerd zouden moeten worden om beoordelaars op ondubbelzinnige wijze te leren omgaan met een gestandaardiseerde en geobjectiverde beoordelingsprocedure. Om terugkoppeling te kunnen verschaffen aan de beoordelaar is het noodzakelijk dat de beoordelingen verricht worden op werkstukken waarvan de kwaliteit bekend is. Met andere woorden: er moet een werkstukkenbestand zijn, waarvan de werkstukken op gedetailleerd niveau gekwalificeerd zijn door een aantal deskundigen.

Alvorens in te gaan op de inrichting van de trainingen, wordt eerst besproken hoe een gestandaardiseerde en geobjectiverde beoordelingsprocedure werd ontworpen en op welke wijze een werkstukkenbestand werd opgebouwd.

5.1. Het beoordelingsprotocol

'Beoordelingsprotocol' is een verzamelnaam voor de omschrijving, de beoordelingsmethode en het scoringsvoorschrift van te onderscheiden aspecten aan tandheelkundige werkstukken. Op dit moment is alleen een definitief protocol beschikbaar voor de klasse II-tweevlakspreparatie.

Uit de omschrijving blijkt het analytisch

karakter van dit beoordelingssysteem. Elk werkstuk wordt beoordeeld op zes kenmerken. Ieder kenmerk wordt geoperationaliseerd middels subkenmerken (zie tabel I).

Voor elk subkenmerk wordt in het beoordelingsprotocol in objectieve termen omschreven:

1. aan welke eisen werkstukken moeten voldoen;
2. hoe vastgesteld dient te worden of ze aan die eisen voldoen;
3. hoe de observatie in een score vertaald moet worden.

(Zie afb. 1 voor een voorbeeld.)

Ad 1. De eisen voor de subkenmerken wer-

den ontleend aan de in het preklinisch onderwijs gebruikte geïndividualiseerde cursus 'Preparatie en Restauratie I en II'.¹⁵ Teneinde die eisen in zo objectief mogelijke bewoordingen te kunnen formuleren, werden enige tips van Mackenzie ter harte genomen.¹⁶ Volgens hem dienen criteria:

A. direct gerelateerd te zijn aan de doelstellingen. Dat wil zeggen, dat de nadruk moet liggen op die aspecten die direct van invloed zijn op de kwaliteit van het werkstuk.

B. operationeel (meetbaar) gedefinieerd te worden. In plaats van: 'de preparatiebreedte is adequaat' formuleert men als volgt: 'bij premolaren mag de breedte van de step variëren van 1.0-1.3 mm.'

5

1.2. PREPARATIEBREEDTE VAN DE STEP

KENMERK : OUTLINE

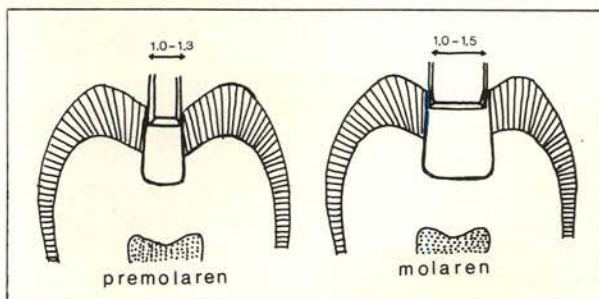
TYPE : Klasse II

Bij premolaren mag de breedte van de step variëren van 1.0-1.3 mm.

Bij molaren mag de breedte van de step variëren van 1.0-1.5 mm.

De preparatie is te smal als glazuurmes 10-8-12 niet door de isthmus heen kan en te breed als glazuurmes 13-8-12 en 15-8-12 door de isthmus van resp. premolaar en molaar heen kunnen.

SCORING : Preparatie-breedte is te smal = 1
Preparatie-breedte is juist = 2
Preparatie-breedte is te groot = 3



Afb. 1. Voorbeeld van een subkenmerk-beschrijving uit het beoordelingsprotocol.

Tabel I. Overzicht van de kenmerken en subkenmerken waarop de werkstukken worden beoordeeld.

SUBKENMERKOVERZICHT	
KLASSE II-TWEEVLAKSPREPARATIE VOOR AMALGAAM	
1. OUTLINE	1.1. HOOFDFISSUURPATROON 1.2. PREP. BREEDTE 1.3. APPROXIMALE WAND 1.4. BUCCALE WAND 1.5. PALATINALE/LINGUALE WAND 1.6. BUCCALE WAND BOX 1.7. PALATINALE/LINGUALE WAND BOX 1.8. CERVICALE WAND BOX
2. DIEPTE	2.1. STEP MESIAAL 2.2. STEP DISTAAL 2.3. RICHTING BODEM 2.4. DIEPTE BOX AXIAAL 2.5. RICHTING AXIALE WAND
3. CAV. OPP. HOEK	3.1. STEP BUCCAAL 3.2. STEP PALATINAAL/LINGUAAL 3.3. BOX BUCCAAL 3.4. BOX PALATINAAL/LINGUAAL 3.5. BOX CERVICAAL
4. CONVERGENTIE	4.1. APPROXIMALE WAND 4.2. BUCCALE WAND STEP 4.3. PALATINALE/LINGUALE WAND STEP 4.4. CONVERGENTIE ZIJFISSUUR 4.5. BUCCALE WAND BOX 4.6. PALATINALE/LINGUALE WAND BOX
5. PULP. AX. AFSCH.	5.1. BREEDTE VLAK
6. AFWERKING PREP.	6.1. OUTLINE STEP 6.2. OUTLINE BOX 6.3. BODEM STEP 6.4. BODEM BOX 6.5. BUCCALE WAND 6.6. PALATINALE/LINGUALE WAND 6.7. AXIALE WAND

C. te vermelden wat nog wel en wat niet meer acceptabel is. Het voorbeeld onder 'B' is hier ook van toepassing.

D. verduidelijkt te worden met behulp van een illustratie of een model.

Ad.2. Voor elk subkenmerk is aangegeven hoe vastgesteld moet worden of het te beoordelen werkstuk aan de geformuleerde eisen voldoet. Er zijn een drietal beoordelingsmethoden: meten, schatten en vergelijken.

Voor de subkenmerken van de kenmerken 'outline', 'diepte' en 'pulpo-axiale afschuining' wordt vastgesteld of aan de eisen voldaan is door metingen te verrichten met behulp van tandheelkundig instrumentarium. Voor de subkenmerken behorend bij de kenmerken 'caviteitsoppervlaktehoek' en 'convergentie' moet de beoordeelaar schatten of aan de gestelde eisen voldaan is.

Door te vergelijken met een referentiewerkstuk wordt vastgesteld of het te beoordelen werkstuk voldoet aan de eisen, geformuleerd in de subkenmerken onder 'afwerking'.

Referentie-werkstukken zijn door eerstejaars studenten vervaardigde klasse II-preparaties, die volgens overeenstemmend oordeel van een aantal stafleden, op de grens van voldoende-onvoldoende liggen. Een werkstuk is voldoende afgewerkt als de kwaliteit op dit aspect vergelijkbaar is met die van het referentie-werkstuk.

Ad 3. Bij de constructie van een beoordelingsschaal wordt men voor het dilemma geplaatst om te kiezen tussen zo hoog mogelijke overeenstemming en zo optimaal mogelijke terugkoppeling. De overeenstemming is gebaat bij zo weinig mogelijk schaalpunten. De terugkoppeling daarentegen bij zo veel mogelijk schaalpunten.^{4,14}

Gekozen werd voor een driepuntsschaal. Alle subkenmerken, behalve die onder 'afwerking', worden gescoord op een nominale schaal. Dat betekent dat de scores alleen een classificatiefunctie hebben. De '2'-score betekent dat aan de gestelde eisen voldaan is. Score '1' en '3' houden in dat niet aan de eisen voldaan is en verwijzen elk naar een andere geconstateerde fout. Voor de subkenmerken onder 'afwerking' wordt een ordinale schaal gehanteerd; hoe hoger de score hoe beter de afwerking van het werkstuk.

In een voorstudie werd het beoordelingsprotocol getest.¹⁷ Aan studenten en stafleden werd gevraagd werkstukken (afkomstig uit het werkstukkenbestand) te beoordelen met gebruikmaking van het beoordelingsprotocol. Ze kregen de opdracht om de omschrijvingen kritisch te bekijken en, indien nodig, kanttekeningen erbij te plaatsen.

5.2. Het werkstukkenbestand

Training in het beoordelen van werkstukken is niet goed mogelijk zonder te beschikken over een werkstukkenbestand waarvan de kwaliteit voor de afzonderlijke werkstukken bekend is.

De opbouw van een dergelijk bestand geschiedde in twee fasen:

fase 1: Verzamelen van het benodigde aantal werkstukken (klasse II-preparaties) aan de hand van bepaalde criteria.

fase 2: Beoordelen op subkenmerk-niveau van de werkstukken door expert-beoordeelaars. Uit deze gegevens werd een referentioordeel gedestilleerd.

Ad fase 1. Uit de verzameling werden 35 werkstukken geselecteerd aan de hand van de volgende regels:

- Er werd geselecteerd per kenmerk. Voor elk kenmerk werd in de verzameling gezocht naar werkstukken die aan alle eisen van dat kenmerk voldeden en naar werkstukken die een mankement vertoonden met betrekking tot dat kenmerk.

- De aantallen geselecteerde werkstukken per kenmerk waren zoveel mogelijk gelijk.



Afb. 2. Voorbeeld van een werkstuk uit het werkstukkenbestand.

Tabel II. Verdeling van de werkstukken over de selectiecriteria (m = molaar, p = premolaar)

	kenmerken												totaal
	outline		diepte		cav.opp.		conv.		pulp.ax.		afw.		
	m	p	m	p	m	p	m	p	m	p	m	p	
De werkstukken voldoen aan alle subkenmerken van het kenmerk waarop geselecteerd wordt.	2	1	2	1	2	1	2	1	1	1	2	1	17
De werkstukken voldoen <i>niet</i> aan minimaal een der subkenmerken van het kenmerk waarop geselecteerd wordt.	2	1	2	2	2	1	2	1	1	1	2	1	18
totaal	4	2	4	3	4	2	4	2	2	2	4	2	35

Tabel III. Voorbeeld van de terugkoppeling die een beoordelaar krijgt door middel van de microcomputer.

Beoordelaar: 15
Werkstuk: 374

Datum: 16-11-82

Oordeel	Kenmerken						
	Ou	Di	Ca	Co	Pa	Af	Cf
15	1	2	1	1	1	2	3
REF	1	3	1	1	1	2	4
17	1	3	1	1	2	2	4
44	2	3	1	1	2	2	4
23	2	3	2	1	3	2	5

Beoordeel ook de subkenmerken van: Di

Verklaring: Ou = outline; Di = diepte; Ca = caviteits-oppervlakte-hoek; Co = convergentie; Pa = pulpo-axiale afschuining; Af = afwerking; Cf = toegekend cijfer op grond van de som der afzonderlijke scores; REF = referentie-oordeel; 17, 44 en 23 zijn collega-beoordelaars die op een eerder tijdstip werkstuk 374 beoordeeld hebben.

In tabel II wordt het bestand schematisch weergegeven.

De geselecteerde werkstukken werden met bijbehorend buurelement gepositioneerd in kleine plastic bakjes. Op de bakjes werden identificatienummers (toevalsgedallen) aangebracht (zie afb. 2).

Ad fase 2. Fase 1 leverde een verzameling van 35 werkstukken op waarvan de kwaliteit op één kenmerk bekend was. In fase 2 werden deze werkstukken op alle kenmerken en subkenmerken beoordeeld door expert-beoordelaars (stafleden die in het preklinisch onderwijs assisteren).

Elk werkstuk werd beoordeeld door meerdere beoordelaars. Uit deze gegevens werd voor elk kenmerk en subkenmerk een referentie-oordeel gedestilleerd. Indien

mogelijk gebeurde dit door de modus (meest voorkomende score) te nemen.

5.3. De trainingen

In totaal worden er zes trainings-sessies georganiseerd. Om organisatorische redenen wordt de beoordelingstaak niet in een plenaire zitting verricht. Iedere sessie bestaat op haar beurt uit een aantal op verschillende tijdstippen geplande beoordelingsronden, waarbij steeds één beoordelaar aanwezig is. In elke ronde van een bepaalde sessie worden dezelfde werkstukken aangeboden, om berekening van de inter-beoordelaarsovereenstemming mogelijk te maken. Na afloop van de laatste beoordelingsronde in een sessie wordt een plenaire zitting gehouden waarin ge-

discussieerd kan worden over het beoordelingsprotocol, het referentie-oordeel en de trainingsopzet.

Het beoordelen in de trainings-sessies geschiedt in eerste instantie op de wijze zoals gebruikelijk in de Nijmeegse prekliniek; dat wil zeggen op kenmerk-niveau. De scores worden ingevuld op zogenaamde schrapkaarten en vervolgens ingevoerd in een microcomputer via een kaartlezer.

De microcomputer zoekt voor elke werkstuk-kenmerkcombinatie het referentie-oordeel op. Bovendien zoekt het systeem naar eventueel door andere beoordelaars gegeven oordelen over dezelfde werkstuk-kenmerkcombinaties. Alle relevante informatie wordt daarna afgedrukt (zie tabel III).

In geval van een afwijkend oordeel ten opzichte van het referentie-oordeel krijgt een beoordelaar opdracht om voor het betreffende kenmerk de beoordeling te specificeren naar subkenmerken. In dat geval wordt gebruik gemaakt van het beoordelingsprotocol. Ook over de beoordelingen op subkenmerk-niveau krijgen de beoordelaars onmiddellijke terugkoppeling via de printer.

6. Discussie

Het besproken trainingsprogramma zal de komende tijd gebruikt worden ten behoeve van assistenten en student-assistenten van het eerstejaars blok: 'Preparatie en Restauratie I'.

Na afloop van de laatste sessie zal nagegaan worden of de training een positieve invloed heeft gehad op de beoordelaars-overeenstemming. Te zijner tijd zullen de resultaten van het onderzoek eveneens ter publikatie aan dit tijdschrift worden aangeboden. Als er sprake is van een toegenomen beoordelaarsovereenstemming, dan zal het resultaat van een kosten-batenanalyse beslissend zijn voor het antwoord op de vraag of de trainingen gecontinueerd en/of uitgebreid moeten worden.

Blijkt het trainingsprogramma geen positief effect te hebben op de overeenstemming, dan dient serieus nagedacht te worden over de vraag of zak/slaag-beslissingen op basis van werkstukbeoordelingen gerechtvaardigd zijn.

Het trainingsprogramma werd ontwikkeld in het kader van een onderwijsstimulerings-onderzoek binnen het Instituut Conserverende Tandheelkunde voor Volwassenen. Medewerkers van het Instituut en eerstejaarsstudenten hebben door hun inzet meegeholpen aan de realisatie van het beoordelingsprotocol en het werkstukbestand.

Voor een belangrijk deel steunt het onderzoek op voorbereidende werkzaamheden en op de begeleiding van Ir. A. J. Sanders. Van belang waren ook de adviezen van Drs. T. G. Borgesius (Instituut voor Onderzoek van het Wetenschap-

pelijk Onderwijs), Drs. E. W. Linn (Instituut voor Volledige Prothese en Maxillo Faciale Prothetiek) en Drs. F. L. Otto (Onderwijsgroep van het Faculteitsbureau).

Voor het kritisch doorlezen van het manuscript is dank verschuldigd aan Prof. Dr. A. J. M. Plasschaert en aan Drs. E. H. A. M. Verdon-schot.

De foto's in dit artikel werden vervaardigd door de heer J. L. M. van de Kamp.

Summary:

Title: Improving the evaluation of preclinical dental students by training instructors.

Keywords: Education – Assessment

This article describes the development of a training-program for instructors to increase reliability of judgments.

Training consists of rating class II cavity preparations on six criteria. After each rating trainees are provided with the correct rating figures as defined by the consensus score of expert raters (standard rating). In case of disagreement with the standard rating, the instructor has to specify his rating. For that purpose a so-called 'raters protocol' has been developed.

A microcomputer provides feedback and is also used for the administration of scores.

Literatuur:

1. *Natkin E, Guild RE.* Evaluation of preclinical laboratory performance: A systematic study. *J Dent Educ* 1967; 31: 152-161.

cal laboratory performance: A systematic study. *J Dent Educ* 1967; 31: 152-161.

2. *Lilley JD, Brugge Cate HJ ten, Holloway PJ, Holt JK, Start KB.* Reliability of practical tests in operative dentistry. *Br Dent J* 1968; 125: 194-197.

3. *Abou-Rass M.* A clinical evaluation instrument in endodontics. *J Dent Educ* 1973; 37: 22-36.

4. *Houpt MI, Kress G.* Accuracy of measurement of clinical performance in dentistry. *J Dent Educ* 1973; 37: 34-46.

5. *Otto FL.* Evaluatie beoordelingsprocedures blok 152 (1977-1978) en blok 160C (1976-1977). Katholieke Universiteit te Nijmegen, 1979.

6. *Borgesius TG.* Beoordelen van practicum-werkstukken: evaluatie van tandheelkundig onderwijsblok 164. Katholieke Universiteit te Nijmegen, 1980.

7. *Penning Ch, Steures RWR, Thoden van Velzen SK, Tromp ThJM.* Een klinisch instructie- en beoordelingssysteem voor caviteits-preparatie en -restauratie. *Ned Tijdschr Tandheelkd* 1980; 87: 34-43 en 88-94.

8. *Sanders AJ.* Evaluatierapport blok 155 studiejaar 1977-1978. Intern Rapport, Katholieke Universiteit te Nijmegen, 1980.

9. *Reenen GJ van.* Resultaten van onderwijs in de studierichting der Tandheelkunde. Dissertatie Vrije Universiteit te Amsterdam, 1981; 130-138.

10. *Otto FL.* Beoordelingsprocedures beoordeeld: Een systematische evaluatie van beoordelingsprocedures van de motorische

onderwijsblokken eerste cursusjaar Tandheelkunde. Katholieke Universiteit te Nijmegen, 1981.

11. *Fuller JL.* The effects of training and criterion models on interjudge reliability. *J Dent Educ* 1972; 36: 19-22.

12. *Mackenzie RS, Antonson DE, Weldy PL, Welsh BB, Simpson WJ.* Analysis of disagreement in the evaluation of clinical products. *J Dent Educ* 1982; 46: 284-289.

13. *Patridge MI, Mast TA.* Dental clinical evaluation: A review of the research. *J Dent Educ* 1978; 42: 300-305.

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

15. *Instituut Conserverende Tandheelkunde voor Volwassenen.* Syllabus bij de geïndividualiseerde cursus Preparatie/Restauratie I en II. Blok 155, Blok 255. Katholieke Universiteit te Nijmegen, 1981.

16. *Mackenzie RS.* Factors essential to evaluation of clinical performance. *J Dent Educ* 1974; 38: 214-223.

17. *Straetmans GJJM.* Het onderwijsstimuleringsproject 'Een geïndividualiseerd trainingsprogramma voor het beoordelen van praktikumwerkstukken'. Intern Rapport CE 82-04, Katholieke Universiteit te Nijmegen, 1982.

December 1982. Philips van Leydenlaan 25, 6525 EX Nijmegen.

FEUILLETON

VAN MOND TOT MONDIG: DE LANGE WEG NAAR DE MODERNE MONDVERZORGINGSMENTALITEIT*)

H. BAUDET

Trefwoorden: Geschiedenis

Inleiding

Beroepsprestige-stratificaties zijn in Nederland betrekkelijk weinig gemaakt. De bekendste is nog altijd die van Van Heek uit 1958 (Sociale stijging en daling in Nederland). Op een schaal van 1 tot 57, die werd aangevoerd door achtereenvolgens het be-

roep van hoogleraar, arts, burgemeester van een grote gemeente, rechter, ingenieur, notaris en advocaat, volgde op de achtste plaats dat van tandarts – boven dat van directeur van een grote onderneming en 48 andere.¹⁾

Natuurlijk betreft dat momentopnamen en dit soort prioriteiten is nogal aan fluctuaties onderhevig. Heden ten dage zou het beeld wel eens dramatisch veranderd kunnen zijn. Maar men mag veronderstellen, dat de tandarts zijn fraaie plaats op z'n minst heeft behouden. Zijn naoorlogse wetenschappelijke opleiding, de riant ont-

wikkeling van zijn inkomen en allerlei daarmee samenhangende factoren hebben zijn sociale prestige sinds, zeg, 1950 waarschijnlijk nog versterkt. De grote toeloop naar het beroep in de laatste dertig of vijfendertig jaren valt zeker ten dele daaruit te verklaren. In elk geval heeft onze twintigste eeuw een sensationele stijging te zien gegeven van het aantal tandartsen. De meesten zullen de cijfers wel kennen. Tede Nederland er in 1900 nog slechts 169, en aan de vooravond van de Tweede Wereldoorlog circa 1300, thans zijn er dat zo'n 6000 (tabel I). Dit betekende in de loop van deze eeuw een stijging per honderdduizend inwoners van 3 naar 42.

Opleiding

Ik noemde opleiding en inkomen. Daar wil ik nog een ogenblik bij stilstaan. Wat de opleiding betreft heeft de tandheelkunde in Nederland pas na de Tweede Wereldoorlog volledige universitaire status gekregen

*) Naar een rede, Nationaal Tandheelkundig Congres 'Mondjesmaat', 30 september 1983, Congresgebouw, Den Haag. Onderzoek: H. Baudet en G. Gommers.