

Video als tentamenmiddel in de psychiatrie

door H. Pomerantz en J. Stijnen

Inleiding

De toepassingen van audiovisuele middelen in het onderwijs zijn veelvoudig. In de afgelopen drie jaren is er een algemene aanvaarding en ingebruikneming van video als technisch hulpmiddel te onderkennen in Nederland. Afgezien van de voor de hand liggende demonstratiemogelijkheden, die de video met zich meebrengt voor assistenten, verpleging en staf, zijn de toepassingsmogelijkheden van video als onderwijsmiddel veelal uitgebreid. Bleeker (1976) heeft uitvoerig de toepassingswijzen van video bij het onderwijs beschreven in zijn artikel. De functies van de videobanden zijn o.a.: het overdragen van praktische kennis van veel voorkomende klinische problemen, het beïnvloeden van attitudes t.o.v. psychiatrische behandeling, het duidelijker en meer doelgericht overbrengen van de stof m.b.t. concrete vaardigheden in het identificeren en diagnostiseren van toestandsbeelden en het vergroten van de betrokkenheid van de studenten bij de stof. De bovengenoemde doelen zijn grotendeels bereikt door het gebruik van de videobanden. Er is echter ook nog een ander probleem. Zoals ook gesignaleerd door Bleeker, is de ervaring dat het kennisniveau van de medische studenten vaak door ons overschat wordt en dat maar een deel van de behandelde stof wordt begrepen. Welk deel dit is, is tot nu toe onduidelijk. Het tentamenmiddel is van grote invloed op wat er wel of niet begrepen wordt door de studenten. Een multiple choice tentamen vergt bijvoorbeeld een andere aanpak van voorbereiding door de student dan een essay of mondeling tentamen. Mondelinge tentamens toetsen bijvoorbeeld apart van de stof ook sociale intelligentie (Ebel, 1972). Afhankelijk van het tentamenmiddel worden verschillende aspecten van de stof benadrukt en een ander niveau van betrokkenheid van de student gevergd. In de psychiatrie is er in de pe-

Schrijvers zijn resp. wetenschappelijk medewerker Psychiatrische Kliniek van het Academisch Ziekenhuis, Universiteit van Amsterdam, hoofd: Prof. dr. P. C. Kuiper, en research assistent afd. Klinische Psychologie van de Psychiatrische Kliniek, Academisch Ziekenhuis, Universiteit van Amsterdam, afd. hoofd: Dr. F. Beenen.

riode voorafgaand aan het co-assistentenschap weinig of geen gelegenheid voor de student om direct kennis te maken met de psychiatrie. Verondersteld wordt dat d.m.v. het video-onderwijs een student een behoorlijke kennis heeft van psychiatrische diagnoses en dat hij zeker de vaardigheden moet bezitten om n.a.v. een videoband een patiënt te kunnen bespreken. De in het algemeen als evaluatiemiddel toegepaste multiple choice tentamens zijn niet in staat deze veronderstelde vaardigheden te toetsen. Er kan hooguit een casus besproken worden maar ook dit blijft op een theoretisch en beslist niet praktisch niveau. Deze tentamenvorm maakt geen onderscheid tussen studenten die wel en studenten die niet de praktische vaardigheden geleerd hebben. Door het uitblijven van de feedback wordt de student niet gecorrigeerd tot aan zijn co-assistentenschap en dan is het vaak demotiverend en tijd verspillend om het op het laatste moment volledig te moeten bijsturen.

De video lijkt ook een uitstekend middel het onderwijs m.b.t. de praktische vaardigheden te valideren. De mogelijkheid van toepassing van video bij een tentamen is al eerder in de vakgroep psychiatrie ter sprake geweest (Bleeker, 1976). Er was een aantal praktische problemen met het opzetten van dergelijke tentamens. Het ontbreken van een geschikte tentamenzaal m.b.t. de aanwezigheid van audiovisuele middelen of het aantal studentenplaatsen, heeft ons gemotiveerd om in eerste instantie met een kleine groep studenten te beginnen, nl. bij het mondelinge deel van het doctoraal 2 tentamen psychiatrie.

Deze studie is vanwege de beperkte opzet bedoeld als pilot-studie, een exploratie van een nieuw tentamenmiddel. Er worden geen voorspellingen gedaan over de resultaten van het tentamen noch over de toekomstige carrière van degenen die hierop hoog scores. Deze studie exploreert de techniek en ervaringen van het werken met videoband als een klinisch-psychiatrisch evaluatiemiddel.

Opzet en methoden

De proefpersonen

De vijfdejaars studenten medicijnen moeten aan het einde van het cursusjaar het zg. D2-examen doen. Dit examen bevat een onderdeel psychiatrie, dat d.m.v. 70 vierkeuze-vragen schriftelijk wordt afgenomen. De studenten die hierop een score lager dan 7 behalen moeten een mondeling doen, waarna hun eindcijfer wordt bepaald als het gemiddelde cijfer van het mondeling en het schriftelijk. Deze groep omvatte 90 studenten, van wie er 68 de enquête inleverden. Bij 13 daarvan was door onvolledige invulling niet mogelijk hun cijfers op het schriftelijk en mondeling en hun eindcijfer te achterhalen, zodat we tenslotte van 55 studenten vrij volledige gegevens hebben. Omdat we wel beschikten over de cijfers van alle studenten die een mondeling moesten doen, was het mogelijk de totale groep herkansers te vergelijken met de groep van 55 die de enquête invulden. Vergelijking van de ge-

middelste cijfers van de totale mondelinge groep met de enquête-groep geeft geen aanleiding om ervan uit te gaan dat de 55 respondenten wat betreft cijferverdeling een speciale selectie vormden uit de totale groep herkansers en dat het niet een representatieve steekproef zou zijn.

Instrumenten

Er is gebruik gemaakt van acht verschillende videobanden, die alle acht opnames waren van interviews met vrouwelijke patiënten. Pogingen om de duur van de banden zoveel mogelijk gelijk te houden mogen niet echt geslaagd genoemd worden. De duur van de banden was resp. 16, 22, 14, 13, 18, 19, 20 en 25 minuten. De studenten kregen een korte samenvatting van de opnameklachten van de patiënt, een formulier met vragen betreffende het video-interview m.b.t. symptomen, toestandsbeeld, syndroom, etiologie, psychodynamiek en behandeling en de enquête. De examinatoren kregen de ontslagbrief van de patiënt of een demonstratieformulier uitgereikt. Er waren vier examinatoren, A, B, C en D, die resp. 30, 31, 21 en 8 studenten kregen toegewezen, van wie er resp. 12, 18, 20 en 5 de enquête invulden.

Opzet

Een aantal studenten werd opgeroepen voor een van de acht zittingen. Bij de oproep voor het mondeling kregen de studenten te horen aan welke examinerator ze waren toegewezen. Dit toewijzen is in principe at random gebeurd, afgezien van een heel enkel geval, waarin de student eerder nadrukkelijk om een bepaalde examinerator gevraagd had. Zij moesten daartoe naar de collegezaal komen, waar zij informatie over de patiënt die zij op de videoband zouden gaan zien, het formulier met korte antwoordvragen en de enquête kregen uitgereikt. Vervolgens werd de videoband onder aanwezigheid van de examinatoren gedraaid. Na afloop kregen de studenten gedurende korte tijd, 10 minuten, de gelegenheid hun antwoorden op papier te formuleren, waarna de formulieren werden opgehaald. Gedurende deze periode trokken de examinatoren zich in een bespreking terug, waarbij zij een eensluidende diagnose stelden. Hierna werd bij iedere examinerator de eerste student toegelaten. Op dit mondelinge gedeelte was het de bedoeling dat de student zijn op papier geformuleerde diagnose verdedigde. Zo'n mondelinge zitting duurde 15 minuten, aan het eind waarvan de examinerator de score van de student voor dit mondelinge tentamen bepaalde, die de student echter niet te horen kreeg. Hierbij is het van belang op te merken, dat examinerator A voor het begin van het examen op de hoogte was van de score van de betreffende student op het schriftelijk multiple choice tentamen. Dit in tegenstelling tot de andere examinatoren.

Analysetechnieken

De beschrijving van de resultaten is in twee categorieën ondergebracht:

- 1 Analyses van de scores, waaronder
 - a over de totale groep (n = 90);
 - b per docent;
 - c per videoband.
- 2 Analyses van de antwoorden op de enquête, waaronder
 - a over de totale groep;
 - b over 4 groepen, ingedeeld naar examinator;
 - c over 8 groepen, ingedeeld naar videoband;
 - d over 4 groepen, ingedeeld naar voorkeur voor tentamen-vorm;
 - e over groepen, ingedeeld naar cijfer op het schriftelijk.

Bij de analyses zijn de antwoorden op de enquête opgevat als gemeten op een intervalschaal met schaalwaarden 1, 2 en 3.

Resultaten

Tentamenscores

Tentamenscores van de gehele groep (n = 90). Teneinde na te gaan of er kwantitatieve verschillen waren tussen schriftelijk, mondeling en eindresultaat, zijn er t-tests berekend tussen de scores. Dit geeft de volgende resultaten:

Tabel 1: Gemiddelde en standaarddeviatie per tentamenonderdeel met bijbehorende t-tests

	gem.	sd	t-waarde met cijfer schrift.	t-waarde met cijfer mond.
cijfer schr.	5.16	.886		
cijfer mond.	6.00	1.31	-5.60**	
eindcijfer	5.72	.86	-7.17**	3.16**

**p < 0.01

Het gemiddelde cijfer op het mondeling en het gemiddelde eindcijfer was significant hoger dan het gemiddelde cijfer op het schriftelijk. Door afrondingen werd het gemiddelde eindcijfer niet gelijk aan het gemiddelde van de gemiddelden van het schriftelijk en het mondeling.

Vergelijking van de scores op het schriftelijk en de eindcijfers (n = 90). In 34 van de 90 gevallen (38%) bracht het mondeling geen verandering in het tentamenresultaat. Van de groep die aanvankelijk een score op het multiple choice < 6 had is 47% alsnog geslaagd (eindcijfer > 6). Van de groep die op het schriftelijk een 6 scoorde, is 16% alsnog gezakt, d.w.z. dat voor zes studenten het mondeling tentamen een nogal nadelige zaak was.

Tabel 2: Intercorrelaties tussen de scores op het schriftelijk, het mondeling en het eindcijfer (n = 90)

	schriftelijk	mondeling
mondeling	.19	
eindcijfer	.63	.78

Video als tentamenmiddel in de psychiatrie

De cijfers op het schriftelijk correleren nogal matig met de cijfers op het mondeling. De redelijke hoge correlaties van de cijfers op het schriftelijk en het mondeling met het eindcijfer bieden geen nieuwe informatie, aangezien het eindcijfer is samengesteld uit de score op het schriftelijk en die op het mondeling. Ook moet hier opgemerkt worden dat de lage correlatie tussen de scores op het schriftelijk en op het mondeling mede veroorzaakt kunnen zijn door de 'restricted range' van de scores van de herkansingstudenten (3 t.m. 6) en de traditionele lage betrouwbaarheid van mondelinge tentamens (Carter, 1962).

In de onderstaande tabel is niet duidelijk of de scoreverschillen op het mondeling het gevolg zijn van een examinatoreffect — d.w.z. dat de ene examinerator hogere cijfers geeft dan de andere of dat er sprake is van toevalsfluctuaties.

Tabel 3: Gemiddelde score op schriftelijk, mondeling en eindcijfers per examinatorengroep

examinator	S	M	E	N
A	5,20	5,90	5,70	30
B	5,29	6,19	5,87	31
C	5,00	5,70	5,50	21
D	5,13	6,50	6,00	8
totaalgroep	5,16	6,00	5,72	90

Voor het onderscheiden van een mogelijk examinatoreffect is vijfmaal een covariante analyse uitgevoerd, waarbij het cijfer op het schriftelijk als covariaat werd ingevoerd en het cijfer op het mondeling de afhankelijke variabele was. Op deze manier moesten de verschillen in scores tussen de examinatoren gecorrigeerd worden voor de verschillen in kwaliteit tussen de studenten. De analyse is eenmaal uitgevoerd op de totale groep ($n = 90$) en eenmaal op de totale groep, met uitzondering van de studenten die resp. examinerator A, B, C en D hadden. Bij A is namelijk een groter effect van de covariaat te verwachten, daar deze immers op de hoogte was van de score op het schriftelijk.

Tabel 4: Resultaten van de covariantieanalyse op de cijfers van het mondeling, met als hoofdeffect de examinatoren en als covariaat de score op het schriftelijk

	SS	df	MS	F	p-waarden
covariaat	5,73	1	5,73	3,44	.064
examinator	4,68	3	1,56	.94	.999
residu	141,59	85	1,67		
totaal	152,00	89	1,71		

Tabel 5: Resultaten van de covariantieanalyse op de cijfers van het mondeling, met als hoofdeffect de examinatoren en als covariaat de score op het schriftelijk, met uitsluiting van de examinator A groep

	SS	df	MS	F	p-waarden
covariaat	1,88	1	1,88	1,09	.301
examinator	4,53	2	2,27	1,32	.276
residu	96,44	56	1,72		
totaal	102,85	59	1,74		

Uit bovenstaande analyse blijkt, dat het covariaateffect niet meer significant is, als de A-groep niet mee doet in de analyse. In dat geval zakt tevens de p-waarde van het examinatoreffect aanzienlijk.

Analyses, uitgevoerd met uitsluiting van resp. examinatoren B, C en D toonden geen noemenswaardige verandering in het covariantie-effect.

Er is getracht te onderzoeken of de tentamenresultaten afhankelijk zijn van de vertoonde videobanden. Daartoe zijn van de 8 videobandgroepen gemiddelde tentamenscores berekend en zijn de verschillen hiertussen d.m.v. t-tests getoetst. Hier is geen gebruik gemaakt van covariantie-analyse omdat de aantallen studenten per groep te klein zijn.

Uit deze analyse is er gebleken dat de banden 2 en 5 op het mondeling de meest extreme gemiddelden leveren, nl. 7,50 en 5,00. Gezien het kleine aantal studenten waarop deze gemiddelden betrekking hebben resp. $n = 2$ en $n = 5$, is het gevaarlijk hieraan conclusies te verbinden. Als men ook groep 8 buiten beschouwing laat ($n = 3$) zijn de meest extreme gemiddelden 5,56 en 6,71, die behoren bij resp. groep 1 en groep 4. De t-toets tussen deze groepen leverde een t-waarde van $-1,77$ met $df = 14$, d.w.z. een overschrijdingskans van $p = 0,10$. Deze niet bijzonder significante waarde was de laagste overschrijdingskans van alle in dit verband uitgevoerde t-toetsen.

Resultaten van de enquête

Gegevens betreffende de gehele groep ($n = 68$). Op de gegevens van de enquête zijn over de totale groep ($n = 55 - 68$) rechte tellingen uitgevoerd en zijn de verschillende items, voor zover mogelijk, geïntercorreleerd.

De nevenstaande resultaten geven aan dat in het algemeen het mondeling videotentamen moeilijker gevonden werd dan het multiple choice tentamen, dat het meer terdege toetste dan het multiple choice en dat het prettiger was. De vraagstelling vond men eerlijk. Over de vraag of men genoeg tijd had om zijn antwoord te formuleren en op te schrijven waren de meningen verdeeld. Wel vond men in het algemeen dat men genoeg kans kreeg op het mondeling om zich te verdedigen. De voorkeur ging het meest

Video als tentamenmiddel in de psychiatrie

Tabel 6: Absolute aantallen en percentages van gekozen alternatieven per enquêtevraag (m.c. = multiple choice)

vraag 1	moeilijker	hetzelfde	makkelijker	geen mening
gem. mondeling	37	8	19	4
moeilijker dan m.c.	54,4	11,8	27,9	5,9
vraag 2	meer terdege	hetzelfde	minder terd.	geen mening
mondeling toets	41	6	14	7
meer terdege dan m.c.	60,3	8,8	20,6	10,3
vraag 3	ja	gaat wel	neen	geen mening
vond u de vraagstelling eerlijk	52	13	1	2
	76,5	19,1	1,5	2,9
vraag 4	prettiger	hetzelfde	minder prett.	geen mening
mondeling prettiger dan m.c.	41	1	24	2
	60,3	1,5	35,3	2,9
vraag 6	ja	gaat wel	neen	geen mening
genoeg tijd om antwoord op te schrijven	23	25	20	
	33,8	36,8	29,4	
vraag 7	ja	gaat wel	neen	geen mening
genoeg kans zich te verdedigen	47	12	9	
	69,1	17,6	13,2	
vraag 8	vid. + mond.	vid. + mond.	vid. + mc.	tent. zonder vid.
welke voorkeur voor tentamenvorm	+ ess.	+ mc.		
	42	9	8	6
	64,6	13,8	12,3	9,2
vraag 9	hoger op mc	hoger op mond.	blanco	
waarop heeft u hoger gescoord	15	37	16	
	22,1	54,4	23,5	

*Tabel 7: Correlatiematrix van de enquêtevragen en de tentamenscores**

	1	2	3	4	5	6	7	8	9	10
1 = cijfer schriftelijk										
2 = cijfer mondeling	.23									
3 = eindcijfer	.64	.81								
4 = vraag 1	.01	.26	.19							
5 = vraag 2	.02	-.22	-.16	.02						
6 = vraag 3	-.07	.07	.02	-.27	.09					
7 = vraag 4	.04	-.11	-.04	-.49	.24	.39				
8 = vraag 6	.02	-.18	-.11	.05	-.07	.11	-.05			
9 = vraag 7	.15	-.12	-.10	-.17	.14	.01	.13	.03		
10 = vraag 9	-.22	.59	.37	.34	-.14	-.02	-.49	-.07	-.29	

* De r-waarden hebben betrekking op aantallen die variëren van 41 tot 68. Hierdoor verschillen de r-waarden tussen de tentamenscores in deze tabel van die in tabel 3 die gebaseerd zijn op N = 90.

uit naar deze tentamenvorm — d.w.z. video plus mondeling plus essay — waarbij men ook verwachtte dat men er hoger op gescoord zou hebben dan op het multiple choice.

De correlatie van het cijfer op het schriftelijk met vraag 9 (de verwachting van een hoger cijfer op het schriftelijk dan wel op het mondeling) was matig $- .22$, terwijl die van het cijfer op het mondeling en vraag 9 zeer redelijk was $+ .59$.

Naarmate de studenten deze tentamenvorm moeilijker vinden (vraag 1) vinden ze de vraagstelling minder eerlijk (vraag 3, $r = -.27$), de vorm minder prettig (vraag 4, $r = -.49$) en verwachten zij hoger gescoord te hebben op het m.c. ($r = .34$).

Er is ook een redelijke correlatie tussen vraag 4 en vraag 3 ($-.39$) en vraag 4 en vraag 9 ($r = -.49$), d.w.z. naarmate de studenten dit tentamen prettiger vinden, vinden ze de vraagstelling eerlijker en verwachten ze ook een hogere score op het mondeling gehaald te hebben dan op het multiple choice. Naarmate men een hoger cijfer heeft op het mondeling vindt men deze tentamenvorm enigszins makkelijker dan het multiple choice tentamen ($r = .26$).

Analyse van enquêtegegevens van de vier examinatorgroepen.

Ook per examinatorgroep zijn er rechte tellingen geproduceerd, waarbij tevens per enquêtevraag t-tests zijn berekend op de verschillen in gemiddelde tussen de groepen. Om na te gaan of er significante ervaringsverschillen waren tussen de videobanden is er per docent een enquête-analyse uitgevoerd. De resultaten geven aan dat wat de moeilijkheidsgraad betreft, de studenten dit tentamen moeilijker vonden bij B dan bij C of A, die onderling geen verschil vertoonden. Ook vonden de examinandi van B deze tentamenvorm minder prettig in vergelijking met het multiple choice tentamen dan de examinandi van A en C dat vonden, en dat zij minder kans hadden zich te verdedigen. De studenten die moesten verschijnen voor A waren het meest van mening dat zij meer terdege getoetst waren.

Analyse enquêtegegevens van de acht videobandgroepen.

Om te trachten na te gaan of er alsnog ervaringsverschillen waren onder de deelnemers zijn er gemiddelden en standaarddeviaties uitgerekend op de enquêtevragen per videobandgroep. De op de gemiddelden tussen de groepen uitgevoerde t-toetsen leverden alleen voor de vraag betreffende de tijd, vraag 6, een redelijk aantal significante resultaten op. Bovendien bleek dat bij band 1 dit tentamen meer als moeilijker dan het schriftelijk werd ervaren dan bij de banden 3 en 6 (resp. $p = 0.07$ en $p = 0.01$).

Hieruit blijkt dat er enige verschillen waren bij het ervaren van de tijdsfactor tussen de banden. Voornamelijk op banden 1 en 2 is er een duidelijke indicatie dat om de een of andere reden de deelnemers vonden dat ze minder tijd hadden in vergelijking met de deelnemers bij de andere banden.

Resultaten van de analyse van gegevens van de subgroepen ingedeeld naar hun voorkeur voor tentamenvorm. In een poging om kenmerken te onderscheiden van de groepen die een verschil-

Video als tentamenmiddel in de psychiatrie

Tabel 8: T-toetsen op gemiddelden van enquêtevraag 6 tussen de 8 videogroepen

band	N	x	SD	t1	t2	t3	t4	t5	t6	t7
1	9	2,56	.53							
2	2	3,00	0	-2,53**						
3	10	1,5	.153	4,36***	9,00***					
4	7	1,29	.49	4,99***	9,3***	,86				
5	6	1,5	.84	2,75**	5,0***	0	-,55			
6	10	2,2	.632	1,34	4,0***	-2,69**	-3,36***	-1,77**		
7	8	2,0	.93	1,5	3,06**	-1,36	-1,9*	-1,06	.52	
8	3	1,33	.58	3,24**	5,00**	,45	-,12	,35	2,23*	1,43

* = $p < .10$

** = $p < .05$

*** = $p < .01$

lende voorkeur voor tentamenvorm uitspraken, is een aantal kruistabellen geconstrueerd. Uit één hiervan is gebleken dat van degenen die band 1, 3 of 6 gezien hebben, nog meer de voorkeur uitgaat naar deze tentamenvorm.

Uit een kruistabel over voorkeur voor tentamenvorm en al dan niet moeilijk vinden van het videotentamen bleek er een interessante tegenstelling in mening te bestaan. Het grootste percentage personen die deze tentamenvorm gemakkelijker vonden dan het multiple choice (39%) werd aangetroffen binnen de groep die de voorkeur gaf aan deze tentamenvorm (video + mondeling + essay). Toch gaf binnen deze groep een groter percentage (48,8%) aan deze tentamenvorm moeilijker te vinden dan het multiple choice. In alle groepen waren de meesten van mening dat deze tentamenvorm moeilijker was, maar desalniettemin ging bij de meesten toch de voorkeur uit naar deze vorm. Dit geeft enigszins aan dat, hoewel men deze tentamenvorm als moeilijker had ervaren, men toch deze vorm meer gewent achtte.

In de overige berekende kruistabellen was geen significant onderscheid te zien in antwoordkeuzen tussen de groepen met een verschillende voorkeur voor tentamenvorm.

Inventarisatie van toegevoegd commentaar op de enquête. De meest voorkomende commentaren betroffen klachten over de geluidskwaliteit. Daarbij viel het op dat deze klachten *niet* steeds betrekking hadden op dezelfde band, maar verdeeld waren over verschillende banden. Door een student werd geklaagd over de lange wachttijd tussen de videozitting en het mondeling (de desbetreffende student moest één uur wachten).

Voorts waren er wat klachten over de eigen zenuwachtigheid. Bij een zitting (band 2) gebeurde het dat de staf tijdens het afdraaien van de videoband binnenkwam. Dit leverde prompt commentaar van enkele studenten, die dit als nogal storend hadden ervaren.

Conclusies

In het algemeen wordt er op het mondeling hoger gescoord dan op het schriftelijk. Dit kan mede veroorzaakt zijn doordat de studenten nu de stof voor de tweede keer bestudeerd hadden; de docenten neigen ertoe de cijfers rondom de 5-6 te geven en zeer lage cijfers niet te geven. De nogal matige correlatie tussen het schriftelijk en het mondeling geeft o.a. aan dat het hier om nogal verschillende instrumenten gaat, die waarschijnlijk iets andere vaardigheden meten. Hierdoor werd het mogelijk dat een zestal studenten, die op hun schriftelijk een zes hadden, op het mondeling dermate slecht voldeden, dat ze een vijf als eindcijfer kregen en alsnog zakten. Als men ervan uitgaat dat zowel het schriftelijk als het mondeling iets meet dat relevant is voor de evaluatie van de prestaties, is het de vraag of niet ook de studenten die een zeven of hoger hadden op hun schriftelijk een videotentamen zouden moeten doen.

Uit de covariantieanalyses op de cijfers mondeling bleek dat in de totale groep het cijfer op het schriftelijk een significante variantiebron was, terwijl deze niet langer significant te noemen was als men de examinandi van examiner A uit de analyse weglief. Hieruit blijkt dat de voorkennis van A van het cijfer op het schriftelijk van invloed is op de te geven score voor het mondeling of dat A over bijzondere kwaliteiten beschikt. Het examinatoreffect was in beide analyses niet significant, d.w.z. dat van de scoreverschillen op het mondeling tussen de groepen die een andere examiner hadden, niet aangetoond kon worden dat deze systematisch zijn en dat het dus voordeliger zou zijn om bij de ene examiner tentamen te doen dan bij de andere.

Wat betreft het verband tussen de score op het mondeling en de vertoonde videoband, moet men er rekening mee houden dat het mogelijk is dat er op band 1 lager en op band 4 hoger gescoord wordt. Doordat de gegevens hier toch op wat weinig studenten betrekking hebben (resp. $n = 9$ en $n = 7$) is de conclusie hierover weinig definitief. Voor de andere videobanden is geen effect op het cijfer gevonden.

Wat betreft het voordeel van de examinandi over het tentamen in de huidige vorm mag het volgende geconcludeerd worden. In het algemeen vond men deze tentamenvorm moeilijker dan het multiple choice, maar tevens vond men dat men meer terdege getoetst werd, de vraagstelling eerlijker was, de tentamenvorm prettiger, dat men genoeg kans kreeg om zich op het mondeling te verdedigen; de voorkeur ging het meest uit naar deze tentamenvorm en men verwachtte op het mondeling hoger gescoord te hebben dan op het multiple choice tentamen, wat veelal terecht was. Alleen wat betreft het punt of men genoeg tijd had gekregen om de antwoorden te formuleren en op te schrijven, waren de meningen verdeeld te noemen. Als geheel mag geconcludeerd worden dat deze tentamenvorm positief beoordeeld werd, zeker in vergelijking met het multiple choice tentamen.

Wat de analyses van de enquêtes per examiner betreft kan worden opgemerkt dat deze tentamenvorm bij de examinandi die het mondeling bij B moesten doen, er het minst gunstig afkwamen. Onder hen was een groter percentage van mening dat deze tentamenvorm moeilijker was, minder prettig en een kleiner percentage was van mening dat ze genoeg kans hadden gehad om zich te verdedigen.

De analyses per band gaven, behalve het feit dat de tentamenvorm bij band 1 moeilijker gevonden werd dan bij band 3 en band 6, vooral verschillen in de beantwoording van vraag 6 te zien. Bij de banden 3, 4, 5 en 8 gaf men in het algemeen aan genoeg tijd te hebben gehad voor het formuleren en opschrijven van de antwoorden, bij de banden 6 en 7 vond men 'het gaat wel', terwijl men bij de banden 1 en 2 niet genoeg tijd gehad zou hebben. In vergelijking met de overige groepen werd bij de groepen die de banden 1, 3 of 6 hebben gezien, het meest de voorkeur uitgesproken voor deze tentamenvorm, d.w.z. video + mondeling + essay. De totale groep studenten die de voorkeur gaf aan deze tentamenvorm, bevatte, zoals ook te verwachten is, een groter percentage studenten die deze tentamenvorm gemakkelijker vonden, dan de percentages in de groepen die een andere voorkeur uitspraken. Toch was ook binnen de eerstgenoemde groep het percentage dat deze tentamenvorm moeilijker vond groter (48,8%) dan het percentage dat deze tentamenvorm gemakkelijker vond (39%). Het bleek niet mogelijk om op grond van het cijfer van een student op het schriftelijk, d.w.z. zijn kenniskwaliteitsindex, een voorspelling te doen omtrent zijn houding t.o.v. deze tentamenvorm, gemeten aan de beantwoording van de enquête.

Een punt dat hierbij niet aan de orde is geweest, maar dat voor tentamens ook van groot belang is, is het oordeel van de examinatoren. De informatie hierover bestaat uit mondelinge reacties, die van alle vier examinatoren zeker positief te noemen zijn.

Tenslotte mag gesteld worden:

- 1 — dat de videoband een nuttig en goed hanteerbaar middel is voor de evaluatie van studentenprestaties, dat door de studenten goed wordt ontvangen, en
- 2 — dat als het tentamen zijn huidige structuur behoudt, de 'borderline cases', d.w.z. de studenten met een zes op het multiple choice gehandhaafd dienen te blijven als deel van de herkansingsgroep,
- 3 — dat het tentamineren aan de hand van een video dermate verschilt van een 'traditioneel' multiple choice tentamen, waarbij de vaardigheden die bij de videovorm noodzakelijk zijn, dermate relevant geacht worden, dat het de vraag is of niet alle examinandi een videovorm-tentamen moeten afleggen.

Discussie

In dit onderzoek is in eerste instantie getracht na te gaan hoe

een tentamen in deze vorm bevalt. Daarbij was het beter geweest als men zich van tevoren er meer op had bezonnen wat 'deze vorm' nu precies inhield. De bedoeling was dat de inhoud van de videoband, het stukje schriftelijk en het mondeling dezelfde zou zijn. Op de inhoud van het mondeling was echter te weinig controle, zoals dat vaak het geval is met mondelinge tentamens. Dit is voor een deel op te vangen door het min of meer standaardiseren van het mondeling, d.w.z. op papier vastgelegde vragen. Het probleem hierbij is, wat men moet doen als een examinandus blijkt geeft van een duidelijk gebrek aan kennis op het betreffende gebied. Moet men een student laten zakken op grond van een meting op één beperkt gebied of moet men hem nog een kans geven door over te stappen op andere onderwerpen.

Verder is in dit verslag aangenomen dat wat door de studenten als video + mondeling + essay werd aangekruist ook werd opgevat als zijnde de vorm van dit tentamen. De antwoorden die zij in plm. 10 minuten moesten opschrijven, kunnen echter nauwelijks essays genoemd worden.

Uit dit onderzoek mag niet de conclusie getrokken worden dat er *geen* sprake was van een 'examinator'-effect op de scores van het mondeling. De toetsing hiervoor is alleen geschikt om aan te tonen dat er *wel* een examinatoreffect is. De toetsing van de afwezigheid van effecten is met de huidige statistische technieken niet bevredigend uitvoerbaar. Bovendien bleek bij de covariantie-analyse het examinatoreffect zonder de A-groep al een grote sprong te maken in de richting van een significant niveau. Het is niet uitgesloten dat als de tentamens door A zonder voorkennis waren afgenomen, de analyse van de totale groep wel een significant examinatoreffect had aangetoond.

Het begrip moeilijk zoals gebruikt in enquêtevraag 1, vraagt om nader bestudeerd te worden. Men zou verwachten dat 'moeilijk' een negatieve betekenis heeft, dat het samen gaat met afwijzing van datgene wat moeilijk is, dat het gepaard gaat met lage scores. Dit tentamen werd echter enerzijds moeilijker gevonden, maar anderzijds verwachtte men er hoger op te scoren en oordeelde men er nogal positief over.

Men heeft in deze tentamenopzet enigszins nagegaan wat het effect van de wachttijd tussen de videovoorstelling en het mondeling was op de scores. Daartoe heeft men op de eerste zitting de scores, verkregen op verschillende tijdstippen, met elkaar vergeleken en geconcludeerd dat het zo best wel kon. Toch is het aan te bevelen dit nauwlettend in het oog te houden en voor een aantal tentamens bij te houden hoe de tentamenscores na de verschillende wachttijden zijn. Dergelijke invloeden zijn immers uit den boze op tentamenscores. Zo is het ook aanbevelenswaardig om per videoband bij te houden wat de daarop behaalde scores zijn, aannemende dat de banden geregeld voor tentamens gebruikt zullen worden. Op deze manier kunnen per videoband gegevens bewaard worden, zoals dat ook gebruikelijk bij de itempools voor multiple choice tentamens is. Het is dan beter mogelijk om uit

te maken welke banden gelijk zijn wat moeilijkheidsgraad betreft. Als men tevens in de enquêtes bij het tentamen wat specifiekere vragen omtrent de vertoonde videoband stelt, behoort het tot de mogelijkheden om meer eigenschappen van de band te registreren, zoals bijvoorbeeld de geluidskwaliteit. Door een handige registratie zal het dan ook mogelijk zijn om een eventueel videoband-examinator interactie-effect na te gaan.

Betreffende de verschillen per band omtrent enquêtevraag 6, is het waarschijnlijk dat deze verschillen niet te wijten zijn aan de lengtes, de duur van de banden, maar aan een subjectief ervaren verschil tussen de zittingen. Bijvoorbeeld bij de eerste zitting waren er enkele technische en organisatorische onzekerheden, die aanleiding hebben kunnen geven voor een negatiever oordeel van de toegestane tijd. De moeilijkheidsgraad, de geluidskwaliteit, etc. kunnen allemaal bijgedragen hebben aan deze ervaringsfactor. Het is te hopen dat d.m.v. deze manier van analyseren een aantal 'goede' banden (homogeen wat betreft een aantal factoren — tijd, geluid, moeilijkheidsgraad) uitgeselecteerd kunnen worden voor gebruik op volgende tentamens.

Hoewel de huidige organisatorische tentamenstructuur wel acceptabel lijkt, is het toch aanvechtbaar dat de studenten die hoog scoren op het multiple choice niet getoetst worden op vaardigheden waarop de studenten die laag scoren wel getoetst worden. Een hoge score op het multiple choice garandeert geen bekwaamheid in praktische vaardigheden, zoals die gehanteerd moeten worden tijdens het co-schap psychiatrie, maar door de videobanden ook te gebruiken op het multiple choice, komt men al meer in de richting van de toetsing van praktische vaardigheden. Aan de validatie van een dergelijk tentamenmiddel kan iets gedaan worden door de scores op een dergelijk tentamen te vergelijken met het optreden tijdens het coassistentenschap psychiatrie. Bovendien zou op deze manier het D2 allicht een betere voorspeller kunnen zijn voor het basisartsexamen.

Literatuur

- Bleeker, J. A. C. (1976), Video bij het onderwijs in de psychiatrie. *Tijdschrift voor Psychiatrie* 5, 347-354.
- Carter, H. D. (1962), How reliable are good oral examinations? *California Journal of Educational Research* 13, 147-153.
- Ebel, R. L. (1962), *Essentials of Educational Measurement*, Prentice Hall, New Jersey.