

Observatietoetsen praktisch getoetst

Drs. S.J. van Luijk, Drs. C.P.M. van der Vleuten & Drs. R.M. van Schelven

De auteurs zijn werkzaam bij de Rijksuniversiteit Limburg bij de Projectgroep Evaluatie Studieresultaten

Binnen diverse takken van onderwijs wordt getracht om naast het meten van kennis ook andere aspecten, voor zover relevant voor de beroepsuitoefening, te meten. Ten aanzien van praktische vaardigheden zijn deze pogingen reeds eerder beschreven o.a. voor conservatoriumopleidingen (Colwell, 1970), middelbare beroepsopleidingen (Brown, 1941) en biologie (Tamir & Glassman, 1971). Dergelijke instrumenten zijn tot nu toe veel minder frequent en grootschalig toegepast dan instrumenten voor kennismeting. In de praktijk blijkt een instrument gericht op vaardigheidsmeting echter wel degelijk een belangrijke aanvulling te zijn op bestaande evaluatieinstrumenten. De nadelige effecten die bij de toepassing van een dergelijk instrument beschreven worden (Harden & Gleeson, 1979) zijn beheersbaar. In het nu volgende stuk wordt verslag gedaan van een aantal jaren ervaring met het toetsen van praktische vaardigheden aan de Medische Faculteit in Maastricht op grote schaal.

Hierbij worden positieve en negatieve aspecten van de toepassing van het instrument beschreven. Tevens volgen suggesties om de negatieve aspecten voor een deel te ondervangen.

Inleiding

Aan de Medische Faculteit in Maastricht bestaat al sedert geruime tijd ervaring met grootschalige afnames van vaardigheidstoetsen. Hierbij wordt door middel van observatie vastgesteld of de student de belangrijkste vaardigheden ten aanzien van competent medisch handelen beheerst. Ongeveer 900 studenten worden per jaar aan een dergelijke toets onderworpen. Aan het eind van elke toets worden structureel enquêtes afgenomen ten einde op het spoor te komen van facetten welke een goed verloop van de toets in de weg kunnen staan of bevorderen. De belangrijkste resultaten van de enquêtes zullen hierbij worden besproken. Om echter de uitkomsten goed te kunnen interpreteren is het noodzakelijk enig inzicht te hebben in het vaardigheidsonderwijs en het toetsstelsel van de medische faculteit.

Het vaardigheidsonderwijs

Het vaardigheidsonderwijs zoals dat vorm heeft gekregen binnen de Medische Faculteit in Maastricht zal hier kort worden toegelicht (zie verder Dochy & Van Luijk, 1987). Vanaf het begin van de opleiding worden aan de student vaardigheidstrainingen aangeboden welke gerelateerd zijn aan de inhoud van de leerstof waar de student mee bezig is binnen het curriculum. Het betreft het aanleren van al die vaardigheden die de basisarts nodig heeft om in de latere beroepspraktijk optimaal te kunnen functioneren. Vrijwel al het vaardigheidsonderwijs wordt georganiseerd vanuit een zogenaamd 'vaardigheidslaboratorium' (skillslab) voor de eerste vier studie jaren. De verschillende groepen vaardigheden die getraind worden zijn de communicatieve vaardigheden, de fysisch diagnostische vaardigheden, laboratoriumvaardigheden en therapeutische vaardigheden. Per week besteedt de student ongeveer drie uur aan het voorbereiden en volgen van de trainingen. Per onderwijsblok van zes weken worden gemiddeld acht onderwijsactiviteiten, (trainingen, simulatiepatiëntcontacten e.d.) georganiseerd. De meeste vaardigheden die getraind worden zijn uitgesplitst in de kleinste mogelijke zinvolle eenheden van handelen. Dit betekent dat een complexe handeling op die manier in stappen uiteengerfeld is waardoor het aanleren van die vaardigheid makkelijker wordt. Een dergelijk uitgesplitste handeling wordt 'standaard' genoemd. Iedere student leert de handeling volgens dezelfde standaard. Het oefenen vindt plaats met behulp van (plastic) modellen, instrumenten of er wordt op elkaar geoefend. Naast het aanbieden van vaardigheidstrainingen worden er ook regelmatig (simulatie) patiëntcontacten georganiseerd en bezoeken aan bejaarden- en verpleegtehuizen. Deze laatste activiteiten vinden met name plaats om studenten te leren hoe ze de pas verworven vaardigheden (geïntegreerd) kunnen toepassen.

Het toetsysteem

De student wordt beoordeeld op een aantal aspecten van medische competentie. Deze aspecten betreffen het kennisdomein, het vaardigheidsdomein, het medisch denken en handelen

en attitude. Voor het laatste aspect bestaat nog geen formeel meetinstrument. Gezien de verdere inhoud van dit artikel zal hier verder alleen worden ingegaan op het kennis- en vaardigheidsaspect. Het kennisaspect binnen de medische studie wordt geëvalueerd door middel van blokttoetsen en voortgangstoetsen. De blokttoets is een schriftelijke toets bestaande uit gemiddeld 180 vragen van het juist/onjuist alternatief met een vraagtekenoptie. Deze is gericht op het evalueren van de kennis verkregen in de reguliere onderwijsblokken die de student doorloopt en wordt ook steeds afgenomen na afloop van een dergelijk blok, dat wil zeggen in de regel om de zes weken. Deze toets wordt alleen in de eerste vier studie jaren afgenomen. Dit instrument telt alleen maar in positieve zin mee (in het voordeel van de student) bij het nemen van beslissingen over de studie voortgang. Derhalve wordt deze toets beschouwd als een hulpbron voor kennis.

De voortgangstoets is de hoofdbron voor de beoordeling van kennis. Deze neemt derhalve een dominante positie in binnen het examenreglement. De voortgangstoets is een einddoelgerichte toets, met andere woorden, alle vragen die gesteld worden zijn afgestemd op het nivo van de basisarts. Iedere jaargroep wordt vier maal per jaar op hetzelfde moment onderworpen aan deze toets. De voortgangstoets heeft dezelfde vraagvorm als de blokttoets, dat wil zeggen juist/onjuist type met een vraagtekenoptie. Het gemiddelde aantal vragen bedraagt 250. Voor wat betreft het toetsen van vaardigheden wordt sedert acht jaar ervaring met een grootschalige afname van observatietoetsen voor het medisch onderwijs opgedaan. Voor elke jaargroep van ongeveer 150 studenten wordt een observatietoets georganiseerd. Dit betekent dat er tenminste 900 studenten per jaar getoetst worden, exclusief de herkansingen. De standaard toetsduur is twee uur.

In deze 'vaardigheidstoets' wordt aan studenten gevraagd om in verschillende kamers diverse vaardigheden te demonstreren. Deze vaardigheden kunnen betrekking hebben op anamnese, lichamelijk onderzoek, therapeutische vaardigheden en laboratoriumonderzoek. In elke kamer (een 'station' genoemd) die de student betreedt zit een beoordelaar (observator) welke de student beoordeeld aan de hand van een gedetail-

leerde beoordelingslijst (de criterialijst). Nadat de student de handeling verricht heeft gaat hij naar een volgend station. Af en toe is er ook een tweede observator (co-observator) aanwezig. Deze scoort (onafhankelijk) mee met de observator ter controle van de betrouwbaarheid. Het onderzoek wordt verricht op modellen, proefpersonen of (simulatie)patiënten. Na de toets krijgt de student na twee dagen de beschikking over zijn beoordelingsformulieren. De vaardigheidstoets is binnen het examenreglement de hoofdbron voor de beoordeling van het vaardigheidsdomein. Op de beslismomenten bepaald door de twee fasen structuur moet de vaardigheidstoets (eventueel na herkansing) met een voldoende resultaat worden afgesloten. Na iedere toets wordt systematisch aan studenten en observatoren de mening gevraagd over een aantal aspecten van deze vaardigheidstoetsen. Hierbij volgt een verslag van deze gegevens zoals die verzameld zijn in de afgelopen jaren.

Vraagstelling

In de loop der jaren is er een aantal vragen opgekomen ten aanzien van de vaardigheidstoets. Deze zijn grotendeels te clusteren rond de volgende onderwerpen.

- Hoe bereiden studenten zich voor op vaardigheidstoetsen?
- Hoe ervaren studenten de vaardigheidstoetsen?
- Hoe ervaren de observatoren de vaardigheidstoetsen?
- Hoe ervaren studenten de vaardigheidstoetsen in relatie tot de kennistoetsen aan deze faculteit?

In dit stuk zullen aspecten van alle bovengenoemde vragen aan bod komen. De keuze van de aspecten werd bepaald door de wens negatieve en positieve effecten van de vaardigheidstoets zoveel mogelijk zichtbaar te maken opdat hiermee rekening gewonnen kon worden bij de verdere ontwikkeling en implementatie van het meetinstrument.

Methode

Instrumenten:

Om de bovengenoemde vraagstellingen te

beantwoorden zijn er in de loop der jaren verschillende enquêtes afgenomen.

Vanaf het studiejaar 1982/83 werden systematisch voor alle jaargroepen gegevens verzameld over de ervaringen van studenten en observatoren met de vaardigheidstoets. Dat wil zeggen dat er iedere keer vragen gesteld werden aan studenten over organisatie, toetsonderdelen (stations), observatoren en simulatiepatiënten. De observatoren beantwoordden vragen over organisatie, training voor observatoren en beoordelingslijsten. Daarnaast bestond er voor beide groepen, zowel observatoren als studenten, de mogelijkheid om zelf aanvullende opmerkingen te plaatsen in de daartoe vrijgehouden ruimte op de enquêtelijs.

Voor studenten bestond de enquête uit 15 gesloten vragen met een vijfpuntsschaal van het Likert type gaande van geheel eens naar geheel oneens. Tevens was de optie niet te beoordelen/geen mening toegevoegd. De vorm van de enquête voor observatoren was identiek aan die van de studenten alleen bevatte deze 11 gesloten vragen.

In het curriculumjaar 87/88 is er vervolgens naast de reguliere lijst voor studenten uit alle jaargroepen een aanvullende vragenlijst gemaakt. Deze vragenlijst bevatte vragen met betrekking tot de voorbereiding op de vaardigheidstoets (19 vragen uitgezet op een Likertschaal met vijf antwoord opties gaande van geheel eens naar geheel oneens). Naast deze vragen waren er vragen aan toegevoegd welke een indruk moesten geven over de beleving van de vaardigheidstoets enerzijds en de beide kennistoetsen, voortgangstoets en blokttoets anderzijds. Het aantal vragen hierover bedroeg 41, waarbij ook weer van een soortgelijke vijfpuntsschaal gebruik werd gemaakt. Tevens bestond de optie niet van toepassing.

Respondenten

Alle studentendie deelnamen aan de vaardigheidstoetsen werd gevraagd na afloop van de toets de vragenformulieren in te vullen. Door de jaren heen vulden vrijwel alle studenten deze formulieren in (respons van 90-95%). Voor de observatoren ligt dit percentage op ongeveer 80%.

Resultaten

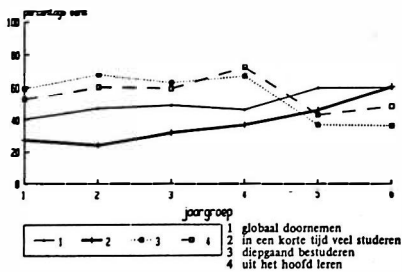
De nu volgende getallen hebben alleen betrekking op de som van de frequenties van de categorieën geheel eens en eens. Omdat men kan kiezen uit een zespuntschaal betekent het complement van geheel eens/eens niet noodzakelijk oneens/geheel oneens, omdat met name de neutraal/geen mening optie bij enkele vragen vaak wordt ingevuld.

Vorbereiding vaardigheidstoetsen

Gevraagd is of de specifieke voorbereiding op de toets, dat wil zeggen de extra tijd die aan de voorbereiding wordt besteed omdat er een toets is, veel tijd in beslag neemt.

Over alle zes studiejaar is ongeveer 80% van de studenten het met die uitspraak eens voor wat betreft de vaardigheidstoets. De percentages voor de bloktoets en voortgangstoets liggen op 38% respectievelijk 17%. De jaren 5 en 6 zijn iets minder tijd kwijt aan de specifieke voorbereiding op de vaardigheidstoets, gemiddeld 64%. Deze specifieke voorbereiding gebeurt met name door het studeren in pieken en veel minder door regelmatig studeren. De manier waarop de vaardigheidstoets wordt beoordeeld stimuleert het uit het hoofd leren.

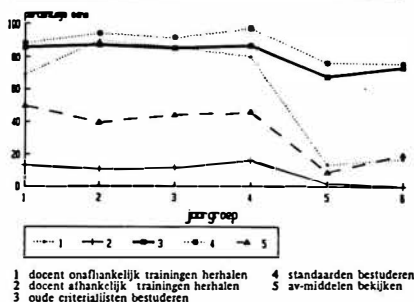
De meest gehanteerde voorbereidingsstrategieën ten aanzien van de vaardigheidstoetsen betreffen het bestuderen van eerder gebruikte criterialijsten, het bestuderen van standaarden en het bestuderen van aantekeningen (zie figuur 1). Voor de vijfde en zesdejaars studenten geldt



Figuur 1: Voorbereidingsstrategieën voor de vaardigheidstoets

dat al deze voorbereidingsstrategieën in relatief mindere mate worden toegepast aangezien deze jaargroepen zich zowieso minder voorbereiden op de toets. De druk vanuit het reguliere onderwijs (co-assistentenschap) lijkt groter dan de examendruk.

Het studeren voor de vaardigheidstoets in de eerste vier studiejaar betekent voor wat betreft de studiestijl dat voor het grootste deel de leer- en oefenstof zowel diepgaand bestudeerd wordt als ook uit het hoofd geleerd. Opvallend is dat de keuze voor de leerstijlen in jaar 5 en jaar 6 precies omgekeerd zijn aan die van de eerste vier jaargroepen (zie figuur 2).



Figuur 2: Leermiddelen ter voorbereiding op een vaardigheidstoets

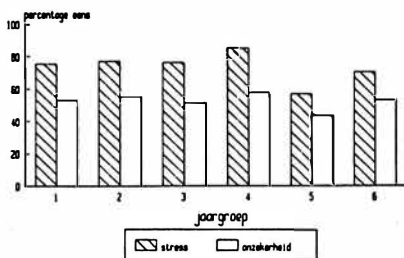
Uit de evaluaties blijkt verder dat de studenten de vaardigheidstoets met name zien als een goed middel om te komen tot het herhalen van vaardigheden en het verkrijgen van een overzicht over vaardigheden. Studenten zijn het meest negatief in hun oordeel over de bijdrage van de toets aan het tot inzicht komen over het hoe en waarom van de vaardigheden. Hierbij is sprake van een afnemende waardering voor deze (op onderwijs gerichte) toets in de laatste twee (op praktijk gerichte) studiejaar.

Afname vaardigheidstoetsen

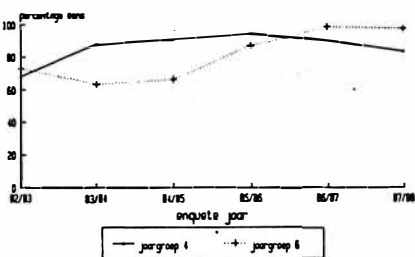
Stress en onzekerheid van studenten

Ten aanzien van de factoren stress en onzeker-

heid bij de uitvoering van de toets blijkt de vaardigheidstoets zoals te verwachten was systematisch een zeer prominente positie in te nemen in de beleevingswereld van studenten hoewel er een afname is in de laatste twee studiejaar (zie figuur 3). Uit andere gegevens wordt duidelijk dat stress en onzekerheid bij kennistoetsing op een aanzienlijk lager nivo liggen. De opgeroepen stress lijkt duidelijk samen te hangen met het observeren van gedrag, wat niet anoniem kan gebeuren en derhalve een meer bedreigende uitwerking op de meeste studenten heeft dan bijvoorbeeld schriftelijke toetsing.



Figuur 3: Stress en onzekerheid tijdens de vaardigheidstoets



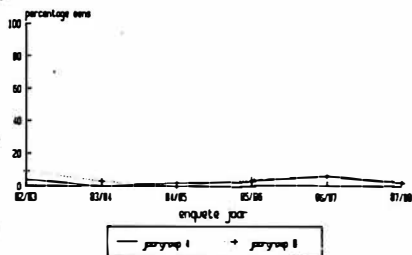
Figuur 4: Relevantie van vaardigheden meningen van studenten

Inhoud van de vaardigheidstoets

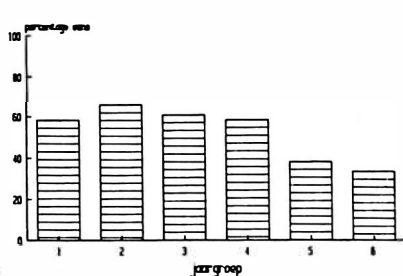
De relevantie van de getoetste vaardigheden wordt door zowel de studenten als de observatoren hoog ingeschat. Gemiddeld 85 procent van de studenten vindt dat men deze vaardigheden nodig heeft als basisarts, al lijken deze percentages voor het zesde jaar iets lager te liggen (figuur 4).

De vraagstelling naar de observatoren is anders, nl. of men de vaardigheden overbodig vindt. Ja, vindt gemiddeld 4 procent (figuur 5), nee vindt gemiddeld 80 procent.

Tevens is gevraagd of de praktijk van de ge-



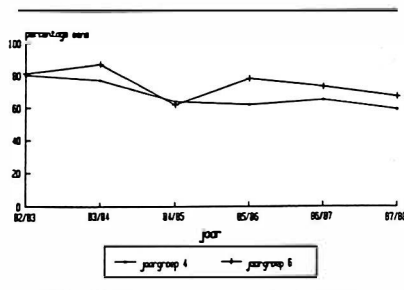
Figuur 5: Overbodige vaardigheden meningen van observanten



Figuur 6: Weerspiegeling van de praktijk in de vaardigheidstoets (studenten)

zondheidszorg zoals studenten die hebben mee-gemaakt zich goed weerspiegeld in de diverse toetsinstrumenten. Hierbij scoort de vaardigheidstoets systematisch hoger dan de schriftelijke kennistoetsen, hoewel hier sprake is van een groot verval in de laatste twee studie-jaren. In de eerste vier studie-jaren is 60% het met deze stelling eens, terwijl in jaar vijf en zes 35% het hiermee eens is (figuur 6).

Op de vraag of het vaardigheidsniveau goed geëvalueerd wordt door de vaardigheidstoets, beantwoord gemiddeld 75 procent van de observatoren bevestigend. De vierdejaarstoets wordt in dit opzicht minder gewaardeerd, aldus de enquête (figuur 7).



Figuur 7: Het vaardigheidsniveau wordt goed geëvalueerd door de toets.

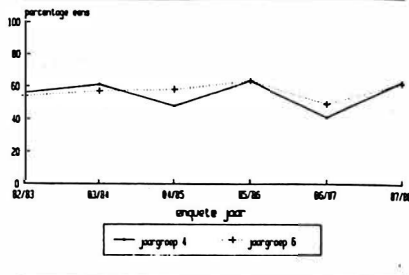
Dit hoeft niet zozeer te liggen aan de vorm en inhoud van het examen, maar misschien ook aan de manier waarop studenten met de vaardigheidstoets omgaan. Veel studenten leren nl. bij de voorbereiding van dit belangrijke examenmoment (eindpunt eerste fase) de criterialijsten uit het hoofd. Hierdoor wordt het onduidelijk of de toets op een adequate manier vaardigheden meet, of dat meer een beroep wordt gedaan op kennis (onthouden en reproduceren).

Het gebruik van criterialijsten door observatoren

Per toets en per jaar wisselen de antwoorden

nogal op de vraag of de observatoren veel moeite hebben om zich te oriënteren op de criterialijsten.

Het valt op dat de bruikbaarheid van de criterialijst relatief laag wordt ingeschat, slechts tussen de 50 en 70 procent van de observatoren vindt deze zeer bruikbaar (figuur 8).



Figuur 8: Criterialijsten zijn zeer bruikbaar; meningen van observatoren

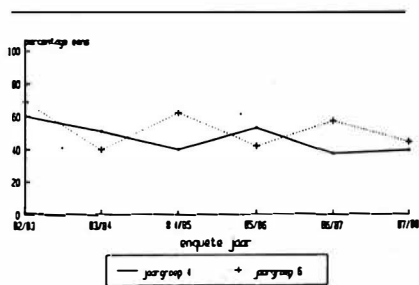
Gemiddeld 15 procent vindt de criterialijsten niet bruikbaar. Het is niet zo dat – door de jaren heen – het commentaar betrekking heeft op één vaardigheidsgebied. Per toets is dit verschillend. Een mogelijke verklaring hiervoor is het veel gehoorde commentaar dat de lijst – in relatie tot de uit te voeren vaardigheid – te gedetailleerd zou zijn en dat het heel moeilijk wordt om op een juiste manier om te gaan met de criterialijsten wanneer de observatietrainingen niet gevolgd worden.

Bovendien worden de lijsten door veel studenten uit het hoofd geleerd, hetgeen de bruikbaarheid van de criterialijst als een meetinstrument voor vaardigheden, niet ten goede komt.

Voorbereiding/training van observatoren

De observatortrainingen worden met name in de eerste jaren goed bezocht. Later wordt dit minder, diverse observatoren hebben de voorbereidingen dan al eerder gevolgd.

Steeds minder observatoren zeggen deze trainingen een goede voorbereiding te vinden, maar dit is deels te verklaren doordat de observatoren die de training niet gevolgd hebben een neutraal antwoord geven (figuur 9).



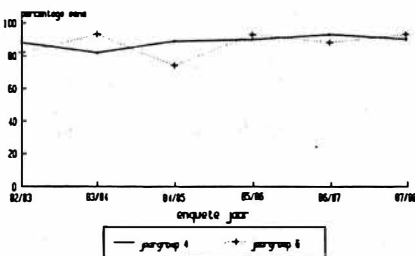
Figuur 9: Observatortrainingen zijn een goede voorbereiding

Voor een ander deel geldt dat men het niet nodig vindt op de eigen vakspecifieke deskundigheid getraind te worden.

Meningen van studenten over (co-)observatoren

Ongeveer 60 procent van de studenten zegt dat de meeste observatoren hem/haar gerust stellen. Het eerste jaar is minder te spreken over de aanwezigheid van de co-observatoren, terwijl gemiddeld 40 procent van de ouderejaars hun aanwezigheid juist positief vindt. Dit komt waarschijnlijk doordat de ouderejaars graag een extra controle zien op het invullen van hun – soms complexe – criterialijsten.

Tachtig tot negentig procent van de studenten zegt feedback prettig te vinden, dit is nagenoeg constant (figuur 10) ..



Figuur 10: Feedback van observatoren ervaart ik als prettig

Dit is een opvallend gegeven aangezien veelal beweerd wordt dat men studenten tijdens een toetsituatie juist geen feedback moet geven omdat ze dan te nerveus zouden worden als gevolg van een mogelijke negatieve feedback.

Discussie

Men kan zijn twijfels hebben over de betekenis van het verzamelen van opinies voor onderzoek en beleid.

Het is bekend dat er interpretatieproblemen kunnen optreden bij 'consumentenoordeelen' gebaseerd op rating scales (Saal, 1980). Afgezien van de meettechnische problemen, onder andere dat mensen geneigd zijn op een rating schaal zoveel mogelijk rondom het gemiddelde te scoren, is bij deze gegevens bijvoorbeeld niet duidelijk wat het effect is van het afnemen van deze enquêtes direct na een vaardigheids-toets.

Desalniettemin stelt March (1984) in een overzichtsartikel dat oordelen op een dergelijke manier verkregen onder een aantal condities zeker betrouwbare en valide informatie opleveren. Dit werd recentelijk bevestigd door onderzoek naar beleving van onderwijs op basis van student-oordelen (Gijssels, 1989).

De stabiliteit van de hier beschreven gegevens door de jaren heen en de ondersteuning van de interpretaties van de gestructureerde vragenlijst door de eveneens verzamelde antwoorden op 'open vragen' geven eveneens steun aan de validiteit van de hier gepresenteerde gegevens.

Het beschrijven van effecten van vaardigheids-toetsen op het onderwijs en studenten is niet nieuw. Reeds bij de introductie van dergelijke toetsen bij de studie geneeskunde zijn voor- en nadelen al beschreven (Harden & Gleeson, 1979). Als grootste nadelen worden genoemd:

- het gevaar van het geïsoleerd testen van de vaardigheid min of meer los van de patiënt
- de belasting voor patiënten en observatoren
- de veel grotere voorbereidingstijd in vergelijking met andere examenvormen.

Dit spoort volledig met bovenstaande ervaringen van afgelopen jaren.

Als grote voordelen worden echter beschreven:

- de meer valide manier van toetsen van praktisch (klinisch) onderwijs waar vaardigheden

- een belangrijk deel van uitmaken
- het betrouwbaar kunnen toetsen aangezien de patiënt en observatorvariantie deels in de hand gehouden kunnen worden
- het van tevoren kunnen bepalen van de blauwdruk van het examen
- het niet afhankelijk zijn van het patiënten aanbod
- het in de hand houden van de complexiteit van het examen
- het grote aantal casus (problemen) wat kan worden aangeboden
- het grote aantal studenten wat aan een dergelijke toets kan worden onderworpen.

Ook dit komt grotendeels overeen met ervaringen binnen de medische faculteit in Maastricht, zij het dan dat dit meer argumenten zijn gereedeneerd vanuit een 'meetperspectief' dan vanuit de student. Effecten op studiegedrag worden hierbij niet genoemd.

Invloed van observatietoetsen op studiegedrag wordt binnen het medisch onderwijs voor het eerst vermeld door Newble (1983). Hierbij draagt de aanwezigheid van de toets bij tot verandering van het studiegedrag (meer praktisch gericht leren en oefenen) in de gewenste richting. Ook recentelijk heeft Newble (1988) een aantal effecten van observatietoetsen beschreven: Studenten ervaren de 'Structured Clinical Examination (SCE)', afgenomen in het laatste studiejaar, prettiger als de kennistoets (MCQ), vinden de SCE zeer relevant maar ervaren ook een steeds grotere invloed van dit examen op hun manier van studeren in de door docenten gewenste richting. Ook Newble geeft aan dat de voorbereidingstijd voor dergelijke examens groot is. Als grote voordelen van de test wordt aangegeven: de relevantie en de faire beoordeling.

Op grond van bovenstaande ervaringen lijkt het zinnig om bij het vormgeven en ontwerpen van dergelijke toetsen naast de gebruikelijke toets-technische vragen (wie maakt en controleert het examenmateriaal, hoe moet de norm worden vastgesteld e.d.) en organisatorische vragen (wie regelt welke zaken) zich ook af te vragen wat de effecten van de toetsing zijn op het leergedrag van studenten. Dit wordt voor een groot deel bepaald door de status die een examen heeft binnen het examenreglement, maar

ligt deels ook aan de vormgeving, het lijkt zinnig richting te geven aan een meer globale manier van toetsing zeker als deze ook nog gericht wordt op het meten van uitkomsten van vaardigheden. Ook observatoren krijgen in dat geval wat meer beoordelingsruimte dan het keurslijf van de criterialijst. Ook hierbij zal echter moeten worden nagegaan wat de consequenties zijn voor betrouwbaarheid en validiteit van het oordeel en tevens welke de gevolgen zijn op het studiegedrag.

Een goede voorlichting over wat studenten én observatoren kunnen verwachten kunnen bijdragen tot het verlagen van de stress en onzekerheid van studenten.

Elke toets behoort een leermoment voor de student en een moment van kennisoverdracht voor de docent te zijn. Derhalve moet er ook binnen dergelijke toetsen ruimte gecreëert worden voor een goede feedback.

Uit de gegevens blijkt dat een dergelijke toets als relevant wordt beoordeeld, door zowel studenten als door docenten. Voor een deel kan de 'praktijksituatie' naar tevredenheid in een dergelijke toets verwerkt worden. Het lijkt er derhalve op dat een dergelijke toets – naast kennistoetsing – binnen het 'praktijk onderwijs' zinvolle informatie zal opleveren over studenten en de kwaliteit van het genoten praktijk onderwijs.

Observatietoetsen zijn vaak in het nadeel ten opzichte van kennistoetsen vanwege de organisatorische complexiteit en acceptabiliteit. Jarenlange toetsing van vaardigheden en daaraan gekoppeld onderzoek heeft aangetoond dat er een meetinstrument is ontwikkeld welk een aantal positieve aspecten kent en de toets der kritiek kan doorstaan. Negatieve bij-effecten kunnen, mits onderkent, beperkt blijven. Dit zal ten goede komen aan de acceptabiliteit en maakt daarmee de kans op implementatie van dergelijke toetsen binnen de diverse curricula groter.

Literatuur

Brown, C.M. *Evaluation and Investigation in Home Economics*, (178-179). New York: Crofts & Co, 1941.

Colwell, R. *The evaluation of music teaching and*

learning, (108-109-116). New York: Prentice- Hall Inc. 1970.

Tamir, P. & F. Glassman. A practical Examination for BSCS Students: a progress report. *Journal of Research in Science Teaching*, 1971, 8, 307-315.

Dochy, F.J. & S.J. van Luijk (red). *Handboek Vaardigheidsonderwijs*, Lisse: Swets & Zeitlinger, 1987.

Saal, F.E., R.G. Downey & M.A.L. Lahey. Rating the ratings: Assessing the Psychometric Quality of Rating Data. *Psychological Bulletin*, 1988, 88, 413-428.

March, H.W. Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility. *Journal of Educational Psychology*, 1984, 76, 707-754.

Gijselaers, W.H. *Kwaliteit van het onderwijs gemeten, studies naar betrouwbaarheid van studentoordeelen*. Dissertatie Rijksuniversiteit Limburg, 1989.

Harden, R.M. & F.A. Gleeson, Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education*, 1979, 13, 41-54.

Newble, D.I. & K. Jeager, The effect of assessments and examinations on the learning of medical students. *Medical Education*, 1981, 17, 165-171.

Newble, D.I. (1988), Eight yearsl experience with a structured clinical examination. *Medical Education*, 1988, 22, 200-204.