

W.M.V. Dolmans, internist
Radboudziekenhuis
Katholieke Universiteit Nijmegen
Postbus 9101
6500 HB Nijmegen

M.D.-examination in Tanzania*

Tanzania is a vast country, which population (18 million) is mainly rural (90%). In its capital Dar es Salaam (1,3 million) the only University of the country has 7 Faculties, about 300 students and roughly 700 academic staff (67% of it is Tanzanian, a figure of 1981). The Medical Faculty was established in 1970. Before - from 1963 - Tanzania had a medical school, as part of the University of East Africa in Kampala, Uganda. The 1200 beds University Hospital has a staff of 240 doctors, of whom about 100 are medical specialists. The students (annual intake is 50) get a basic 5-year M.D.-course, consisting of 3 years tuition in the biomedical sciences, pathology, diagnosis and treatment, as well as demographic studies and health system management. The 4th and 5th year are taken up as junior and senior rotations through different clinical departments, and contain 2 elective periods. After the MD-exam the successful candidate enters 1 year of rotating internships before being allowed to practise independently c.q. a D.M.O.

The M.D.-exam is composed of 4 parts i.e.:

1. continuous assessment
2. written papers (M.C.Q. as well as essay-questions)
3. clinical exams in 5 subjects (medicine, surgery, obs/gynaecology, paediatrics and psychiatry)
4. orals are done in any case in community medicine, and as needed supplementary to 1-3 (either in weak candidates, or in very good ones for a possible credit).

The clinical exam (in each of the subjects mentioned) consists of one long case (45 min. for clarking and 15 min. for examination) and of 2-4 short cases (in total 20-30 min.). Two teams of examiners operate, each consisting of 2 internal and 1 external examiner, who judge candidates independently. A candidate who had team A on the long case will have team B for the short cases. The latter may consist of well circumscribed

tasks asked from the candidate in different patients (e.g.: "Would you examine the cardiovascular system in this patient?", or: "can you demonstrate to us the reflexes in the legs"), or he may be asked to examine a blood-film or stool-specimen under the microscope, etc.

The results of a candidate's performance are gathered first at departmental level, where for each of the 4 categories mentioned above a judgement is given, resulting finally in a pass or a fail, for that particular subject. The next step is to compile the data of all subjects of all candidates during the Examiners Board, which will conclude to a pass or fail for the M.D.-exam of all candidates, and which sets the period for repeat-exams in case of failure. After this the Faculty Board experts a marginal check on the results per Faculty. Finally the United Senate oversees the results of all exams of all Faculties during an official meeting. Both at Faculty Board and Senate the students may (through their representatives) dispute the exam-results of the Examiners Board. The system at the Faculty of Medicine, University of Dar es Salaam has some positive points compared to the Dutch system as it is at present:

1. The candidates are judged in a fair way: independent teams examine different aspects so that the candidate gets several chances.
2. the clinical exam is a nice combination of

a. a comprehensive judgement in the long case

b. a check on differing aspects in several short cases.

Negative points in the set-up in Dar es Salaam are its heavy organisational load, and that the system is time-consuming and expensive (e.g. because of the system of inviting 6 or more external examiners, mostly from other countries.

*Naar een voordracht op de NVMO Studiedag op 22 april 1986.

Presentatie van G. Norman*

Historie

Om een bespreking van de huidige stand van zaken in de "assessment of clinical competence" te introduceren is een korte historische schets van de ontwikkelingen rond studentenevaluatie aangewezen.

Gedurende de eerste jaren van min of meer systematische bemoeienissen (1955-1970) ligt de nadruk op de evaluatie van feitenkennis. Het onderzoek concentreert zich op de evaluatie van feitenkennis. Het onderzoek concentreert zich op de ontwikkeling van zgn. "objectieve" schriftelijke testmethoden, zoals Multiple Choice Questions (MCQ).

Vervolgens (1970-1980) tekent zich een toenemende belangstelling voor "Clinical Problem Solving" af. De vigerende veronderstelling was dat "Probleemoplossen" een afzonderlijke algemene vaardigheid was.

Diverse schriftelijke simulaties van het medisch probleemoplossingsproces zoals met name de zgn. PMP's ("Patient Management Problems") doen hun intrede en raken op vele plaatsen zeer in zwang. Onderzoek spitst zich toe op de aard en het wezen van "problem solving" en "clinical reasoning". In het verlengde hiervan krijgt ook onderzoek naar de validiteit en betrouwbaarheid van de gehanteerde toetsmethoden meer en meer aandacht.

De laatste jaren tenslotte ontstaat het besef, dat "clinical competence" een complex multidimensioneel gebeuren is. Meerdere methoden zijn nodig om de verschillende aspecten van "clinical competence" adequaat in beeld te krijgen. Men moet niet op zoek gaan naar de beste examenmethode, maar naar de beste set van toetsinstrumenten om "clinical competence" te examineren. Het onderzoek richt zich meer en meer op de basale mentale processen, die aan probleemoplossen ten grondslag liggen.

In de huidige situatie is de algemeen geldende opvatting, dat medische competentie globaal een viertal onderscheidbare, doch onderling nauw samenhangende dimensies of vaardigheden kent:

1. Probleemoplossingsvaardigheden
 2. Kennis;
 3. Medisch praktische vaardigheden (anamnese, lichamelijk onderzoek e.d.);
 4. Sociale vaardigheden en attitude.
- Een combinatie van evaluatiemethoden is aangewezen om alle dimensies te bestrijken (Multi-trait/multi-method benadering).

Algemeen geaccepteerde opvattingen over de meest aangewezen methodologie zijn niet voorhanden. Volstaan wordt met een bespreking van de voor- en nadelen van de diverse benaderingen.

PROBLEEMOPLOSSINGSVAARDIGHEDEN

PMP's

In de nationale examens van de Verenigde Staten en van Canada wordt voornamelijk gebruik gemaakt van schriftelijke simulatie-vormen, zoals PMP's en variaties hierop. Ze zijn erg populair geworden, omdat ze op het eerste gezicht een nauwgezette weergave leken te realiseren van het probleemoplossingsproces, dat artsen in werkelijkheid doorlopen.

Inmiddels is overduidelijk aangetoond, dat het handelen van een kandidaat bij een PMP niet overeenstemt met z'n handelen als hij in werkelijkheid met hetzelfde probleem geconfronteerd wordt. Bovendien is er sprake van een grote mate van zgn. casusspecificiteit, zodat generalisaties over de kwaliteit van de probleemoplossingsvaardigheden vanuit één of enkele casus niet mogelijk zijn. Dit klemmt te meer daar PMP's en vergelijkbare schriftelijke simulaties in het algemeen zeer bewerkelijk zijn, zowel wat betreft constructie als wat betreft scoring. Al met al wordt de validiteit en betrouwbaarheid van deze toch min of meer objectieve toetsinstrumenten ernstig in twijfel getrokken, aldus Norman.

* Verslag door G.M. Verwijnen

MEQ

Een alternatief kan mogelijkserwijs gevonden worden in de zgn. Modified Essay Questions (MEQ's), ontwikkeld door Knox in Engeland. Series van duidelijk toegevoegde en goed afgegrensde beperkte doch open vraagstellingen in de context van een klinisch probleem zijn hierbij kenmerkend.

Er is duidelijk sprake van een grotere mate van objectiviteit en betrouwbaarheid dan b.v. bij de traditionele Essay vragen, welke nog steeds bij herhaling naar voren geschoven worden om denkprocessen te maten. Bovendien is er geen sprake van het zgn. "cueing" effect, dat bij PMP's en meer gesloten objectieve vraagvormen, zoals MCQ's zo'n storende rol speelt. Problemen ontstaan met name bij de scoring, die handmatig door middel van beoordeling door terzakekundige medewerkers uitgevoerd moet worden. Toepasbaarheid ("feasibility") is daardoor beperkt.

Voorts blijkt uit recent onderzoek van Norman zelf naar deze toetsvormen, dat een hoge correlatie tussen MEQ's en MCQ's gedemonstreerd kan worden. Mogelijk, dat MCQ's geplaatst in de context van een probleemsituatie toch een belangrijke rol kunnen spelen. Ook onderzoek heeft immers uitgewezen dat probleemoplossen nauw gerelateerd is aan kennis.

Mondelinge examens

Van het van oudsher bekende mondelinge examen is bij herhaling de validiteit en betrouwbaarheid ernstig betwijfeld. Toch tekent zich de laatste tijd in de Verenigde Staten, waar het om bovenvermelde redenen in discrediet raakte, weer een stijgende populariteit af. Vooral het zgn. Gestructureerde Mondelinge Examen (GME) lijkt goede perspectieven te bieden. Door ven te voren nauwkeurig de relevante topics en vraagstellingen alsmede de beoordelingscriteria vast te leggen en bovendien afname en beoordelingsproces met de examinatoren te trainen is een belangrijke toename van validiteit en betrouwbaarheid te bewerkstelligen.

Uitgebreide ervaringen zijn opgedaan door de "Board of Emergency Medicine" in de Verenigde Staten. Men is zeer enthousiast. Het blijkt, dat een serie korte examens (casuïstiek besprekingen) beter uit de bus komen dan lange sessies. Op een bepaald moment wordt geen nader inzicht meer verkregen naarmate het examen voortduurt. Ondanks de strakke structurering is er toch een belangrijke mate van flexibiliteit, waardoor men als examinator kan bijsturen, nader uitdiepen of verhelderen, hetgeen bij alle mogelijke schriftelijke vormen uitgesloten is. Combineren met directe observaties van het patiëntenonderzoek is een aanbevelenswaardige uitbreiding. De ervaring leert dat men bepaald niet kan varen op wat de kandidaat rapporteert in dit verband.

Al met al blijkt dat met relatief geringe aanpassingen de gebreken van het traditionele klinische examen aanmerkelijke verbeterd kunnen worden.

Ondanks deze perspectieven is het over het algemeen echter van belang zich in dit domein de casus-specificiteit van het handelen te blijven realiseren. De onderlinge correlaties tussen verschillende problemen zijn laag. Men heeft op z'n minst 8-10 verschillende problemen nodig om zich reële generalisaties te kunnen veroorloven. De vraag of er zoiets als een afzonderlijke algemene probleemoplossingsvaardigheid bestaat, blijft actueel. Er zijn aanwijzingen dat het toch gaat om een bepaalde vrij algemene vorm van kennisbeheersing.

KENNIS

Als methoden noemt Norman de zgn. Essay-vragen (open vragen), diverse vormen van MCQ's en de Modified-Essay-Questions. Open vragen zijn uiterst inefficiënt (kunnen slechts een beperkt inhoudelijk gebied bestrijken), onbetrouwbaar (steeds opnieuw zeer lage interbeoordelaars' overeenstemming) en dubieus in hun doelstelling (gaat het om stijl of om inhoud).

MCQ's daarentegen vormen een uiterst efficiënte en zeer betrouwbare wijze van kennistoetsing, waarbij opgemerkt wordt dat, indien goed geconstrueerd, het mogelijk lijkt om behalve feitenkennis, ook toepassing van kennis en soms zelfs aspecten van probleemoplossen te meten.

Veel gemaakt is de kanttekening dat herkennen (waar vooral MCQ's een beroep op zouden doen) iets anders is dan herinneren (waar het juist om gaat en waar de open vraag vormen zich op richten). Deze weliswaar verschillende mentale processen blijken empirisch echter onderling zeer nauw met elkaar samen te hangen. Het vermogen tot herkennen is met andere woorden in hoge mate afhankelijk van het vermogen tot herinneren. Een belangrijk bezwaar van MCQ's is echter, dat ze neigen tot het meten van triviale, detaillistische feitenkennis, in een veelal nauwelijks relevante inhoudelijke context. Dit bezwaar is met name in de MEQ's ondervangen. Ze zijn meestal probleemgeoriënteerd, kunnen zich goed richten op de toepassing van beschikbare kennis, zijn relatief gemakkelijk te construeren en vanwege structurering redelijk betrouwbaar te beoordelen.

Problemen ontstaan vooral bij de ontwikkeling van scoringscriteria, terwijl bovendien zoals eerder gezegd de toepasbaarheid beperkt is vanwege het beroep dat op terzakekundige beoordelaars gedaan moet worden.

De eerder gerapporteerde hoge correlaties tussen MEQ's en MCQ's worden vooral veroorzaakt door het context aspect. MCQ's geplaatst in een willekeurige probleem~~on~~afhankelijke context correleren niet of nauwelijks met vergelijkbare MEQ's, terwijl MCQ's geplaatst in de context van dezelfde probleemsituatie als de MEQ's een nagenoeg perfecte samenhang vertonen.

Het devies luidt dan ook om altijd te proberen de toetsing in de relevante context van een (klinische) probleemsituatie te plaatsen.

MEDISCH PRACTISCHE VAARDIGHEDEN

In dit domein zijn vooral observatiebeoordelingsschalen (ratingscales en checklists) van toepassing. Wijdverbreid is het gebruik van al dan niet globale beoordelingen door klinische begeleiders tijdens de coschappen (stagebeoordelingen). Er bestaat een enorme variatie aan mate van detaillering en structurering. Het is een uiterst eenvoudige, gemakkelijk toepasbare en goedkope benadering, waarmee in principe een rijk geïllustreerd beeld van het functioneren van de kandidaat gevormd kan worden. De praktijk leert echter dat dit meestal niet gebeurt en dat er sprake is van bijzonder lage interbeoordelaars betrouwbaarheden en twijfelachtige validiteit, door onvoldoende standaardisering en training.

Direkte observatie en registratie met behulp van uitgewerkte checklisten en beoordeling aan de hand van tevoren vastgestelde beoordelingscriteria onder gestandaardiseerde omstandigheden is een andere uitwerking die een redelijke betrouwbaarheid en validiteit waarborgt. Het vereist echter een flinke input aan tijd en mankracht zowel wat betreft voorbereiding als wat betreft afname, zodat de toepasbaarheid beperkt is.

Tenslotte wordt de zogenaamde OSCE (Objective Structured Clinical Examination) genoemd als een goed gedocumenteerde benadering voor het meten van basale klinische vaardigheden. Het is in feite een organisatie-model, waarbinnen vooral de directe observatietechniek als methode gehanteerd wordt. Kandidaten moeten in een serie zogenaamde stations van beperkte duur en omvang verschillende basale vaardigheden demonstreren. Deze door Harden in Engeland ontwikkelde benadering wordt momenteel gepropageerd als de meest aangewezen voor het meten van medische competentie. Het concept biedt ontegenzeggelijk vele mogelijkheden, doch is zeer arbeidsintensief en in zijn huidige uitwerking veelal beperkt tot geïsoleerde basale vaardigheden.

SOCIALE VAARDIGHEDEN EN ATTITUDE

Een problematisch gebied, niet in de laatste plaats vanwege onduidelijkheden omtrent definiëring en afgrenzing alsmede plaats en betekenis bij examenbeslissingen. Afgezien hiervan lijkt directe observatie onder gecontroleerde omstandigheden toch de meest geschikte methode te zijn. Behoudens beoordeling door observatoren is ook de beoordeling door de betrokken patiënten van belang. Vooral getrainde simulatiepatiënten voldoen daarbij het best. Gegevens omtrent betrouwbaarheid en validiteit zijn redelijk te noemen.

Ook stagebeoordelingen door klinische begeleiders en zgn. peer-assessments (beoordeling door collega-studenten) worden toegepast, maar hebben meestal een lage betrouwbaarheid.

SAMENVATTING EN CONCLUSIES

Samenvattend kan men stellen dat de zgn. multi-trait/multi-method benadering de meest aangewezen aanpak is voor de evaluatie van medische competentie. De uitwerking van de noodzakelijke matrix van trekken vs methoden/instrumenten is sterk afhankelijk van de lokale mogelijkheden, onderwijsdoelstellingen en de sociale en politieke sfeer binnen de betreffende faculteit.

Een belangrijk vraagstuk in dit verband is bv. of men voor het artsexamen kan volstaan met een selectie van problemen (8-10) uit het gehele gebied van de geneeskunde, dan wel dat een toetsing per discipline noodzakelijk is. Duidelijk is echter dat men niet kan volstaan met één enkel probleem, noch met één enkele methoden, terwijl bovendien van belang is dat voorzover mogelijk elke methode uitgewerkt worden in de context van relevante klinische probleemstellingen.
