

Verslag van het congres 'EVALUATION IN MEDICAL EDUCATION' te Beersheva, mei 1987

Y. van Leeuwen, huisarts, Vakgroep Huisartsgeneeskunde, Rijks-universiteit Limburg, Maastricht.

Wandelgangen zijn geliefd, ook op medische congressen. Men verwacht er het heil te halen, dat men in de congreszalen vergeefs zoekt. In hoeverre de 'vruchtbare contacten' ook werkelijk vrucht dragen, valt overigens te betwijfelen. Ook de Nederlandse Beersheva-gangers, ongeveer 20 in getal, afkomstig uit met name Maastricht, Nijmegen, Utrecht en Groningen, stelden hun verwachtingen van het aanbod niet te hoog. Daarentegen togen zij met tassen vol sheets, dia's en computervellen naar het beloofde land, om daar allerlei blijde boodschappen betreffende gedegen 'Hollands' onderzoek te brengen.

Gastvrouw was de medische faculteit van de Beersheva universiteit, die evenals Maastricht, behoort tot het netwerk van faculteiten waar het onderwijs op probleem-gestuurde leest is geschoeid. Problemen te over trouwens, als men zich realiseert dat de regio van de faculteit wordt gevormd door de uitgestrekte zonbeschroeide Negev-woestijn, waar trekkende bedoeïnen hun eigen cultuur langzaamaan zien opgaan in de heersende cultuur die op Joodse, maar meer nog op Westerse leest is geschoeid. In het onderwijs aan medische studenten in Beersheva wordt veel aandacht besteed aan 'community health' met een groot accent op gezondheidsvoorlichting en -opvoeding. Helaas was er tijdens het congres weinig gelegenheid om daarmee kennis te maken.

Van siësta's had men in deze hete regionen, in de middag ongeveer 40 graden!, blijkbaar nog nooit gehoord. De dagen waren vol-gestouwd met praatjes en forum-discussies, soms twee uur achter-een zonder pauze. Joden hebben een traditie waarin leren centraal

staat. Dat viel ook nu niet te miskennen.

In tegenstelling tot de terughoudende verwachtingen viel de kwaliteit van het gebodene mee. Door de overvloed aan thema's en de aanwezigheid van internationaal befaamde onderzoekers was er zowel voor beleidsmakers als voor onderwijskundig dan wel meet-technisch geïnteresseerde artsen en sociale wetenschappers iets van hun gading. Alleen de psychometrici pur sang kwamen mogelijk niet geheel aan hun trekken.

Evaluatie van klinische competentie had de volle aandacht.

Maatsch, van de Michigan State University, USA, opende de discussie met zijn LUMP-theorie. 'LUMP' is te vertalen als 'klont'. Hij illustreerde met een scala van onderzoeksresultaten, dat de theorie die er vanuit gaat, dat klinische competentie te onderscheiden is in los van elkaar staande deelaspecten als kennis, vaardigheden en attituden, niet langer stand houdt. Toetsen, gericht op de verschillende deelaspecten blijken in hoge mate te correleren: goede studenten scoren op alle gebieden doorgaans goed, slechte studenten op alles slecht. 'Houdt maar op', was zijn boodschap, 'met het ontrafelen van het begrip; het is één grote klont waarvan de exacte samenstelling ons vooralsnog ontgaat. Wees echter getroost: met elke toets, of het nu kennis, vaardigheid of iets anders betreft, meet je wel iets dat in de klont thuishoort. Veel meten, hoe dan ook, geeft veel informatie over de klont. Eén enkele toets is een kruimeldief, die een onrechtvaardig oordeel velt over studievoortgang van studenten'.

Norman, van de McMaster University, Canada, deed er nog een schepje bovenop met zijn stelling, dat klinische competentie niet veel anders is dan het op de juiste wijze gebruik maken van vergaarde kennis. Vooral de aanname, dat de competentie is gelegen in een juiste strategie voor het verzamelen en interpreteren van verkregen gegevens van de patiënt, moet overboord. Artsen blijken bij de aanvang van het consult, na het aanhoren van de klacht, al hypothesen in hun hoofd te hebben ten aanzien van de diagnose. De kwaliteit van deze hypothesen, zo blijkt uit onderzoek, bepaalt in hoge mate de kans, dat uiteindelijk de juiste diagnose wordt gesteld. Norman baseert zijn uitspraken echter op onderzoek bij dermatologen en radiologen, patroonherkenners bij uitstek. Het is goed mogelijk, dat de resultaten anders uitvallen indien casuïstiek uit bijvoorbeeld de huisartspraktijk en de praktijk van de internist wordt gepresenteerd. Ik kan me althans moeilijk voorstellen, dat de klacht: 'Ik voel me de laatste tijd niet zo goed', of 'Mijn baby heeft de hele nacht gehuild', onmiddellijk associaties met welomschreven ziektebeelden oproept. Vlekjes en schaduwen lenen zich daar meer voor.

Het advies van 'veel toetsen', geuit door Maatsch, werd aangevuld door het advies 'lange toetsen' door Swanson, vertegenwoordiger van de American Board of Internal Medicine, USA. Hij liet zien hoe slecht het is gesteld met de betrouwbaarheid van de vigerende observatie-toetsen, waarbij studenten worden beoordeeld op hun denken en handelen, terwijl ze een vaardigheid uitvoeren, een foto beoordelen of een gesprek voeren met een (simulatie-)patiënt. In vele faculteiten wordt de Objective Structured Clinical Examination (OSCE) gebruikt, waarbij studenten, al opdrachten uitvoerend, een circuit doorlopen, terwijl een observator hun handelingen scoort.

Een dergelijke OSCE, aldus Swanson, zal uit zeer veel onderdelen, hij noemde een getal van minstens 30, moeten bestaan, wil een betrouwbaar oordeel over het competentie-niveau van de student kunnen worden geveld. De vraag is dan echter of je studenten een dergelijke toetsbatterij kunt aandoen.

'Veel toetsen' en 'lang toetsen' werd op vele manieren aangevuld met 'anders toetsen'.

Beersheva propageerde bij monde van Mevr. Mahler, een schriftelijke toets, gericht op het achterhalen van het inzicht van studenten in de samenhang tussen medische begrippen. Studenten werd bijvoorbeeld geleerd de werking van insuline bij een diabetes-patiënt in schema weer te geven. Aspecten, zoals opname van glucose in de spieren, de omzetting van eiwitten en vetten in glucose in de lever, moesten in het schema worden verdisconteerd. Een toets is ontwikkeld om het geconstrueerde schema op kwaliteit te beoordelen. Een zeer interessante optiek, zeker indien men bedenkt dat dit aansluit bij de theorie, gepropageerd door cognitief psychologen, dat de structuur van kennis bepalend is voor de wijze, waarop deze adequaat kan worden gemobiliseerd, hetgeen volgens Norman weer bepalend is voor datgene wat klinische competentie wordt genoemd.

'Anders toetsen' demonstreerde Beersheva bij monde van Insler, gynaecoloog, ook door demonstratie van een computersimulatie. Op zich niet nieuw, maar de uitgebreide feedback-faciliteiten wekten bewondering.

Over feedback gesproken.

Jason, van de University of Miami, USA, benadrukte op pathetische wijze, een televisie-dominee waardig, dat het gebruik van toetsen als instrument voor feedback en zelfevaluatie sterk wordt verwaarloosd, hetgeen in onderwijskundig opzicht als een doodzonde moet worden beschouwd.

Ook Norman deed in zijn slotrede

een duit in dit collecte-zakje en geenszins ten onrechte. In een tijd, waarin levenslange intercollegiale toetsing wordt gepredikt, is het de plicht van onderwijsgevers toetsneuroses te voorkomen, o.a. door de aantrekkelijke kanten van beoordelen meer te benadrukken.

Kagan, van de University of Houston, USA, had het over 'gevoel'. Hij demonstreerde een beoordelings-instrument: de Affective Sensitivity Scale, waarbij het open staan van de arts voor de gevoelens van de patiënt werd beoordeeld. Dit gebeurde met behulp van een videoband, waarop een gesprek tussen patiënt en arts was opgenomen. De band werd verscheidene malen stopgezet, waarop de student werd gevraagd wat hij dacht, dat er in (één van) de personen omging. De beoordelings-schaal zou nog moeten worden geperfectioneerd, maar was reeds redelijk bruikbaar.

En daarmee komen we op attitudes. Rezler, van de University of New Mexico, USA, stelde nogmaals vast, dat het met het beoordelen van attitudes slecht was gesteld in heel de medische wereld. Zij merkte scherpzinnig op, dat het weliswaar moeilijk is vast te stellen wat een gewenste attitude is, maar dat het daarentegen een veel minder moeilijke opgave is aan te geven wat zeker een ongewenste attitude is.

Goede meet-tips bleven eigenlijk achterwege, maar de gemaakte opmerking is mogelijk al inspirerend genoeg.

De Nederlandse bijdragen, waarover ook elders in dit nummer wordt gesproken, vielen op door hun gedegenheid en goede presentatie-vormen. De breed georiënteerde toets-interesse van huisarts-docenten, vooral uit Groningen,

Utrecht en Nijmegen, viel daarbij extra op. Terwijl in de huisarts-opleiding niet wordt getoetst met selectie-doeleinden, wordt juist daar een groot scala van toetsmethoden uitgetoetst, waarbij zowel kennis als technische en sociale vaardigheden en attitude aan de orde komen. Wellicht juist omdat de toetsen vooral als zelfbeoordelingsinstrument worden gebruikt, wordt aan het element van feedback veel zorg besteed.

De Maastrichtse bijdragen getuigden van een gezond meetverleden, waaruit steeds meer 'metens-waardigheden' te voorschijn komen.

Norman vatte het congres in een hem toevertrouwde spitsvondige notendop samen. Hij formuleerde de volgende adviezen:

- richt onderzoek naar klinische competentie op onderzoek naar de kwaliteit van de initiële hypothesen. Hij prees daarbij het in Nederland lopende onderzoek met schriftelijke open casus (De Graaff, Maastricht);
- ga te rade bij de cognitief psychologen, die hun aandacht richten op de structuur van kennis, waarbij voor evaluatiedoeleinden het toetsen van inzicht in verbanden (Beersheva) past;
- toets lang en veel, om een rechtvaardig oordeel over studenten te kunnen vellen en geef fatsoenlijke, inzichtelijke feedback;
- vergeet de attitudes niet!

Na deze bondige afsluiting volgde een dag van dwalen door Jeruzalem en een rondgang langs de Dodenzee, waarmee het congres meer dan waardig werd afgesloten.

'Next year in Jerusalem', bij wijze van spreken.

Nederlandse bijdragen

- A simple paper-and-pencil test for the assessment of medical problem-solving.
E. de Graaff.

- A written test as an alternative measure of performance: is it possible?
C.P.M. van der Vleuten en S.J. van Luyk, Rijksuniversiteit Limburg, Maastricht.
- An integrated evaluation of the vocational training of general practitioners.
J. Bulte, R. Grol, A. Helsper-Lucas, H. Mookink en V. Tielens, Katholieke Universiteit, Nijmegen.
- Assessment integrated in a long-term continuing medical education program in family medicine.
C.H.G.M. Foolen, G.M. Verwijnen, S.J. van Luyk, G.H.M.I. Beusmans en C.P.M. van der Vleuten, Rijksuniversiteit Limburg, Maastricht.
- Assessment of the performance of trainees in the vocational training of general practitioners.
H. Mookink, A. Helsper-Lucas, J. Bulte, R. Grol en V. Tielens, Katholieke Universiteit, Nijmegen.
- Cognitive similarity between trainer and trainee as a determinant of success in general practice vocational training.
K.H. Groenier, J. van der Ende en S. Visser, Rijksuniversiteit, Groningen.
- Development of a medical content classification system for item-banking.
P.A.W. van Hessen, N.F.C. Cobben en G.M. Verwijnen, Rijksuniversiteit Limburg, Maastricht.
- Evaluation of communication skills.
J. van der Ende, S. Visser en K.H. Groenier, Rijksuniversiteit, Groningen.
- Necessity of a test-review committee in test-construction.
P.A.W. van Hessen en G.M. Verwijnen, Rijksuniversiteit Limburg, Maastricht.
- Peer review in the vocational training of general practice.
R. Grol, J. Bulte, A. Helsper-Lucas, H. Mookink en V. Tielens, Katholieke Universiteit, Nijmegen.
- Some examples of the possibilities of program independent achievement testing with progress tests.
Tj. Imbos, P.A.W. van Hessen, A. Muijtens, H. Snellen, G.M. Verwijnen en W.H.F.W. Wijnen, Rijksuniversiteit Limburg, Maastricht.
- Taking decisions about students: lottery or science?
C.P.A. van Boven, B. Stalenhoef, H. Snellen, Tj. Imbos en C.P.M. van der Vleuten, Rijksuniversiteit Limburg, Maastricht.
- The assessment of the professional performance of trainees in vocational training of general practitioners.
R. Pieters, Rijksuniversiteit, Utrecht.
- The construction of a simulation of initial medical problem-solving (SIMP). (Workshop)
Y. van Leeuwen, E. de Graaff en M.J. Drop, Rijksuniversiteit Limburg, Maastricht.
- The Dutch diabetes mellitus test for residents in family medicine.
Y. van Leeuwen, J. Zuidweg, C.P.M. van der Vleuten en L. Tan, Rijksuniversiteit Limburg, Maastricht en Gemeente Universiteit, Amsterdam.
- The effect of vocational training on knowledge of morbidity in general practice.
K.H. Groenier, S. Visser en J. van der Ende, Rijksuniversiteit, Groningen.

- The skills-test: an adequate predictor of clinical performance.
N.E.J.C. L'Espoir, S.J. van Luyk en C.P.M. van der Vleuten,
Rijksuniversiteit Limburg, Maastricht.
- The use of a self-assessment questionnaire for student counseling.
S.J. van Luyk, N.E.J.C. L'Espoir, H.L.J. Bartstra, G.M. Verwijnen
en Tj. Imbos, Rijksuniversiteit Limburg, Maastricht.
- What's in the score?
An exploratory analysis of true/false test scores in the assessment of
knowledge.
G.M. Verwijnen, Tj. Imbos, P.A.W. van Hessen en W.F.H.W.
Wijnen, Rijksuniversiteit Limburg, Maastricht.