

TOETSING IN BEWEGING: congresverslag van The Second Ottawa Conference on Assessing Clinical Competence, 27-30 juni 1987

H.W. van Lunsen, arts-seksuoloog, Hoofd Werkgroep Seksuologie,
Academisch Medisch Centrum, Amsterdam,
A.J.J.A. Scherpbier, arts-onderwijscoördinator, Vakgroep Chirurgie,
Rijksuniversiteit Groningen.

Inleiding

Onder de inspirerende leiding van Ian Hart en Ronald Harden werd voor de tweede maal een congres georganiseerd over het toetsen van medische competentie. Zowel in wetenschappelijk als in sociaal opzicht was het een congres van voortreffelijke kwaliteit.

Het aantal deelnemers aan deze conferentie was ongeveer 250. Canada en de Verenigde Staten waren uiteraard ruim vertegenwoordigd. Daarnaast waren er deelnemers uit o.a. Engeland, Ierland, Schotland, Zweden, Nederland (de rest van Europa zie je helaas zelden op onderwijscongressen), Australië, Nieuw-Zeeland en Maleisië. De Nederlandse deelnemers (15) waren afkomstig uit Leiden, Amsterdam, Rotterdam, Maastricht en Groningen.

Naast een tiental 'invited papers' van oude en nieuwe kopstukken (Anderson, Swanson, McGuire, Bordage, Neufeld, Stillman, Newble, Harden, Barrows, Hart, Maudsley en Norman) en vijf levendige workshops, waren er 69 'free papers' van gemiddeld goede kwaliteit. Nederland was met 7 voordrachten niet slecht vertegenwoordigd:

- Decomposition of OSCE's; some methodological considerations and empirical findings. C.P.M. van der Vleuten, S.J. van Luyk, Rijksuniversiteit Limburg, Maastricht;
- Long-term effects of a structured clinical examination on a training programme for basic clinical skills. H.J.M. van Rossum, M.A. von Bodegom, Th.W.M. Hoeks, J. Pols, Rijksuniversiteit, Leiden;

- The use of selfassessment as a means to improve the effectiveness of a senior-clerkship in paediatrics. Th.W.M. Hoeks, D. Bosschert, H.E. Zoethout, Rijksuniversiteit, Leiden;
- Assessment of some aspects of medical competence by means of evaluation by non-professional patients. H.W. van Lunsen, D. Soeters, Rijksuniversiteit, Groningen;
- Assessing history-taking skills orobservers' peculiarities? B.J. Roze, A.J.J.A. Scherpbier, R.J. Hiemstra, Rijksuniversiteit, Groningen;
- Assessing history-taking skills orsimulated patients' peculiarities? R.J. Hiemstra, A.J.J.A. Scherpbier, B.J. Roze, Rijksuniversiteit, Groningen;
- Assessing students' performances orexamination peculiarities? D. Soeters, A.J.J.A. Scherpbier, H.W. van Lunsen, Rijksuniversiteit, Groningen.

De 'invited papers' en de 'free papers' zullen worden gepubliceerd in de proceedings, die eind 1987 zouden verschijnen.

Toetsing in beweging

Het OSCE (Objective Structured Clinical Examination) is sinds de eerste beschrijving door Ronald Harden in 1975 op vele plaatsen ingevoerd. Dit bleek ook tijdens het congres uit het feit, dat er meerdere voordrachten uit plaatsen 'all over the world', over OSCE's werden gehouden. Het ene OSCE is het andere niet, of anders gezegd: de verschillende OSCE's zijn onderling vaak nauwelijks te vergelijken. Het OSCE is geen nieuwe examenvorm, zoals soms wordt gesuggereerd, maar een

organisatiemodel, waarbinnen diverse toetsingsvormen geïncorporeerd kunnen worden. De betrouwbaarheid van het totale examen (OSCE) is onder meer afhankelijk van het aantal stations en van het aantal verschillende toetsvormen. Meerdere sprekers wezen erop, dat voor het bereiken van een redelijke betrouwbaarheid de totale toetsingstijd voldoende lang moet zijn. Voor elke examen-vorm kan op basis van de 'generalizability' (David Swanson) worden uitgerekend wat de toetsingstijd moet zijn om de magische grens van 0.80 (betrouwbaarheidscoëfficiënt) te halen. Door in een examen verschillende toetsvormen te gebruiken, kan de totale examentijd voor zowel studenten als examinatoren dragelijk blijven. Bij een totale toetsingsduur van ongeveer drie uur (\pm twintig opdrachten van gemiddeld tien minuten) kan er dan sprake zijn van een betrouwbare toets, aldus berekende David Swanson voor de congresgangers. Een goed OSCE is dan ook een Multi Method Multi Media Multi stationsexamen (de vijf M's van Ian Hart).

Met betrekking tot het 'multi-method' karakter van een examen werd er ook nog op gewezen, dat het observeren van studenten alleen dient te geschieden bij die competentiedomeinen, die alleen door directe observatie kunnen worden beoordeeld. Andere domeinen kunnen vaker efficiënter op een andere manier worden getoetst (multiple choice, gestructureerde open vraag e.d.).

Voor het beoordelen van onderzoeks-, gespreks- en anamnesevaardigheden blijken de 'standardized patients' in opkomst te zijn. Paula Stillman heeft zelfs de neiging de hele proceskant van medische competentie uitsluitend door 'standardized patients' te laten beoordelen. Het is de vraag of deze 'patiënten' daar waar het gaat om de beoordeling van de communicatieve facetten van de arts-patiënt relatie, niet teveel zijn gebiased door hun dubbele rol.

Hoewel op veel plaatsen het besef, dat medische competentie meer is dan alleen kennis, nog nauwelijks is doorgedrongen, is er onder de avant-garde sprake van een herwaardering van cognitieve elementen. De avant-garde spreekt ook niet meer van kennis pur sang, maar van kennis ingebed in alle domeinen van competentie en toepassing van kennis by 'skills', 'attitudes' en 'problem-solving abilities'.

Welke boodschappen werden door de 'big shots' nog meer uitgedragen?

- Er is teveel aandacht voor inter-en intrabeoordelaarsbetrouwbaarheid. Bij goede training en selectie van beoordelaars kan betrekkelijk eenvoudig en snel een goede betrouwbaarheid van de beoordelaars worden bereikt. Training vindt in het algemeen gestandaardiseerd plaats met behulp van op video opgenomen examens. Als de mogelijkheid bestaat meer examinatoren in te schakelen, kunnen er dus beter méér stations worden georganiseerd dan dat één examen door twee beoordelaars wordt geobserveerd.

- Er is te weinig aandacht voor inter-case reliability en case specificity: bij het ontwerpen van opdrachten moet er meer dan voorheen worden geijkt; er wordt nogal eens van uitgegaan dat twee verschillende opdrachten hetzelfde meten (inter-case reliability). Daarnaast moet voor het goed samenstellen van toetsen meer worden nagegaan of een bepaalde opdracht wel echt specifiek is voor een bepaalde kennis- of vaardigheidsdomein.

- Er is toenemende aandacht voor de invloed van toetsing op het leer- en studeergedrag van studenten. Wellicht het grootste belang van een multi-method OSCE is, dat studenten zich beter, meer praktijkgericht en minder monomaan op een toets voorbereiden.

Ook toekomstig leergedrag wordt door OSCE's beïnvloed en is afhankelijk van de kwaliteit van de

feedback. Er bestaat dan ook consensus over het feit, dat het beter is om niet te werken met normscores (x% van de maximale score moet worden gehaald), maar met profielscores, die aangeven waar de student zich op elk onderdeel van de toets bevindt ten opzichte van het optimum en/of ten opzichte van collega-studenten. Norman signaleerde in zijn overzichts-voordracht in dit verband, dat de aandacht voor het effect van toetsing op leergedrag van studenten vóór en na die toetsing vooral in Nederland tot belangrijke nieuwe ontwikkelingen leidt.

Conclusie

De 'Ottawa' conferenties hebben hun bestaansrecht duidelijk bewezen. Er is een internationaal platform ontstaan, waarop een kleurrijk gezelschap van psychometrici, onderwijsontwikkelaars en

medici ideeën en ervaringen uitwisselt met betrekking tot het toetsen van medische competentie. In tegenstelling tot de 'closed shop' van de verpolitiekte AMEE, hebben bij de 'Ottawa' conferenties zowel beginners als gevorderden de gelegenheid hun ervaringen en onderzoeksresultaten te presenteren en te bediscussiëren. Zeker gezien de vele positieve reacties, mag dan ook worden verwacht, dat ook de derde 'Ottawa' conferentie in Groningen goed zal worden bezocht. Dit congres (22-24 mei 1988) krijgt als titel: TICTAC: Third International Conference on Teaching and Assessing Clinical Competence. Tijdens dit congres zal worden geprobeerd om niet alleen aandacht aan toetsing, maar ook aan 'teaching' en competentie te besteden.