

HET ARTSEXAMEN IN DISCUSSIE

A.J.J.A. Scherpbier, arts-onderwijscoördinator, Vakgroep Chirurgie, Rijksuniversiteit Groningen,
G.M. Verwijnen, huisarts, medewerker Project Evaluatie Studieresultaten Rijksuniversiteit Maastricht,
N.E.J.C. l'Espoir, arts, medewerker Project Evaluatie Studieresultaten Rijksuniversiteit Maastricht,
J.C.M. Metz, coördinator Onderwijsontwikkeling, Faculteit der Geneeskunde en Tandheelkunde, Katholieke Universiteit Nijmegen,
C.P.M. v.d. Vleuten, psychometricus, medewerker Project Evaluatie Studieresultaten, Rijksuniversiteit Maastricht,
H.J.M. van Rossum, arts-onderwijscoördinator, Vakgroep Interne Geneeskunde, Rijksuniversiteit Leiden.

Inleiding

In het voorjaar van 1987 organiseerde de Nederlandse Vereniging voor Medisch Onderwijs een dag over toetsingsmogelijkheden in het praktisch klinisch onderwijs. Deze praktische dag was een vervolg op een studiedag waarover eerder in dit bulletin is gerapporteerd (1). Samenvattend bleek, dat er nogal wat kritiek op de "traditionele" artsexamens mogelijk was. De medische competentie van de examenkandidaat wordt bijvoorbeeld vrijwel alleen beoordeeld op grond van een schriftelijke rapportage, de status, en een mondeling examen. Anamnese en lichamelijk onderzoek -als proces- worden niet geobserveerd en derhalve telt alleen het product -de status- mee voor het eindresultaat. Verbetering van het artsexamen betekent in elk geval observatie van de examenkandidaat tijdens anamnese en onderzoek, aldus Norman (McMaster University, Canada) in zijn voordracht. Hij vertelde ook, dat er niet één examenmethode is om "clinical competence" te toetsen. Kortom, meerdere examenvormen of examenmethoden zijn nodig om een kandidaat te beoordelen.

Als vervolg op deze theoretische studiedag met levendige discussies werd een praktische dag georganiseerd, waarbij de deelnemers zelf ervaring konden opdoen met verschillende examenvormen. In dit artikel een beschrijving van deze dag.

Het artsexamen in discussie

Naar aanleiding van deze titel en de toelichting in de folder schreven 56 deelnemers zich in voor een dag over toetsingsmogelijkheden in het praktisch-klinisch onderwijs, in Maastricht.

Het ochtendprogramma was georganiseerd in de vorm van een stationsexamen, zoals dat door Harden et al. (2) is beschreven.

Een stationsexamen is een organisatie-model voor examens, waarbij de examinandi worden geconfronteerd met een aantal examenopdrachten (stations). Elke examenkandidaat start in een ander station. Per tijdseenheid, bijvoorbeeld 20 minuten, wordt carrouselsgewijs doorgeschoven naar het volgende station, zodat elke examenkandidaat in verschillende volgorde dezelfde stations passeert.

Voor de deelnemers bestond het "examen" uit acht verschillende stations van 20 minuten. Vanwege het aantal deelnemers (56) werden er twee identieke stationsexamens (circuits) tegelijkertijd georganiseerd. De deelnemers konden in groepjes van twee à drie kennismaken met de verschillende examenvormen. De stations werden zodanig ingericht, dat de deelnemers ervaring konden opdoen als examiner en als examenkandidaat. Dankzij de ervaring van Scheltus van Luyk en Noël l'Espoir en de medewerkers van het Skillslab met het organiseren van stationsexamens verliep het ingewikkelde och-

tendprogramma (56 deelnemers verspreid over 16 kamers, die allemaal elke 20 minuten naar een andere kamer moesten verhuizen) vlekkeloos.

In een aantal stations werden gegevens verzameld betreffende de mate van overeenstemming tussen de deelnemers bij de beoordeling van een examenkandidaat. De gegevens werden tijdens de dag verwerkt door de mensen van het Projekt Evaluatie van Studieresultaten (PES) van de Rijksuniversiteit Limburg.

Tijdens het middagprogramma werd gediscussieerd over de verschillende examenvormen en werden de gegevens gepresenteerd.

De inhoud van de stations

Het stationsexamen bestond uit de volgende stations:

- anamnese;

Eén van de deelnemers uit elk subgroepje nam de anamnese af bij een ervaren simulatiepatiënt. De patiënt was een nieuwe patiënt met als klacht "pijn op de borst". In tien minuten moest de examinandus de anamnese afnemen en een waarschijnlijkheidsdiagnose aan de examiner mededelen. De andere deelnemers moesten als examiner een observatieformulier bestaande uit 52 items (!) invullen. De anamnese werd alleen inhoudelijk beoordeeld. Gesprekstechniek werd in een ander station beoordeeld. De prestatie van de kandidaat werd ook beoordeeld door een ervaren examiner. Het invullen van een gedetailleerd observatieformulier bleek voor onervarenen moeilijk te zijn. "Heeft hij wel of niet de uitstraling van de pijn goed uitgevraagd?", "Is de samenhang tussen pijn op de borst en emoties goed uitgevraagd?". Het lijkt gemakkelijker dan het is - aldus een deelnemer - om een examenkandidaat zo te observeren dat je weet of hij/zij iets wel of niet heeft gevraagd.

- gesprekstechniek;

De patiënt in dit station was een

simulatiepatiënt, die vervroegd terugkwam op de polikliniek Cardiologie. De examenkandidaat kreeg als opdracht om anamnestiche klacht te exploreren en vervolgens uitleg te geven omtrent de vermoedelijke diagnose, behandelingsmogelijkheden en adviezen.

De observatielijst bestond uit 15 items. Het aantal items was voor de deelnemers geen probleem, in tegenstelling tot het observatieformulier bij het anamnesestation (52 items). Voorbeelden van te beoordelen items zijn: "Stelt op de juiste wijze en het juiste moment gesloten vragen", "Reflecteert op adequate wijze gevoelens die non-verbaal of verbaal door de patiënt geuit worden". Het beoordelen van sommige items vond men moeilijk en het was volgens de deelnemers noodzakelijk om de examinatoren te trainen, waarbij afspraken over een aantal begrippen, zoals bijvoorbeeld "reflecteren van gevoel", moeten worden gemaakt.

- computersimulatie;

Computerprogramma's waarmee het oplossen van een patiëntenprobleem kan worden gesimuleerd, zijn de laatste jaren sterk verbeterd. De mogelijkheid tot koppeling met andere informatiedragers, zoals beeldplaat en compact-disc, leidt tot steeds fraaiere simulaties.

Bij de programma's RUPS (Van Rossum, Leiden) en CASES (Verbeek, Leiden) kunnen de prestaties van de kandidaat aan het eind van het programma in een score worden weergegeven. Dit betekent voor de examiner een tijdsbesparing bij het afnemen van een toets. Interessante programma's voor de deelnemers aan deze dag dus.

De problemen bij computerprogramma's als toetsinstrument zijn te vergelijken met de problemen bij Gestructureerde Open Vragen (GOV). De gouden standaard van anamnese en onderzoek is bijvoorbeeld een groot probleem bij de produktie. Wat is een goede vraag bij de anamnese? Welke vraag is een slechte vraag? Welke gedeelten

van het lichamelijk onderzoek zijn noodzakelijk? Welke overbodig? De subjectiviteit van de docent-auteur van het programma kan worden bestreden door een panel deskundigen bij de produktie te betrekken. Een ander probleem is de (nog?) noodzakelijke "simplificatie" van het probleemoplossend proces, althans in vergelijking met de manier waarop men nu denkt dat artsen in werkelijkheid problemen oplossen. Hoewel er nog wel "problemen op te lossen zijn" is, gezien de ontwikkelingen van de laatste jaren, te verwachten dat het toetsen van diagnostische vaardigheden met behulp van de microcomputer een plaats zal gaan verdienen als onderdeel van examens in het praktisch klinisch onderwijs.

- neurologisch onderzoek;

In dit station kon kennis worden gemaakt met de mogelijkheid om fysisch-diagnostische vaardigheden te toetsen.

Het neurologisch onderzoek werd verricht bij een simulatiepatiënt met lage rugpijn, uitstralend naar het been. De simulatiepatiënt simuleerde afwijkingen bij het lichamelijk onderzoek. Eén van de deelnemers onderzocht de patiënt, de andere deelnemers beoordeelden dit onderzoek aan de hand van een observatieformulier.

De formulering van deze opdracht was zodanig, dat de kandidaat "gestuurd" werd. De kandidaat wordt geacht slechts bepaald onderzoek te doen; alleen dat onderzoek wat in de opdracht staat vermeld.

In dit station werd de kandidaat niet alleen beoordeeld op de technische handeling -de onderzoeker slaat met de reflexhamer "los uit de pols" op de quadricepspees, vlak onder de patella- maar ook op de bevinding -de kandidaat komt tot de juiste bevinding met betrekking tot de kniepeesreflex- en de instructies aan de patiënt -bij de sensibiliteit wordt de patiënt geïnstrueerd de ogen te sluiten of niet te kijken.

De observatielijst bestond uit 58 items. Daarentegen werd in het volgende station met korte observatielijsten van 3 items gewerkt.

- korte opdrachten;

In dit station drie korte opdrachten: beoordeling van het gezichtsveld, de stemvorkproef van Weber en de stemvorkproef van Rinne. Een heel kort observatieformulier voor elke opdracht. Elke opdracht werd beoordeeld met drie items.

Eén deelnemer was kandidaat, één deelnemer was patiënt. De overige deelnemers moesten als observator aan de hand van een kort observatieformulier hun oordeel geven. Bij de volgende korte opdracht verwisselden de deelnemers hun rol; de examenkandidaat werd observator etcetera. De deelnemers konden op deze wijze examenervaring opdoen in meerdere opzichten.

Waarschijnlijk mede doordat de deelnemers zelf ook examenkandidaat waren, werd er levendig gediscussieerd over de juiste methode om de opdracht uit te voeren. In de echte examensituatie moet zowel voor de examenkandidaat als voor de examinerator duidelijk zijn wat er gedaan moet worden en op welke manier.

- traditioneel mondeling examen;

Bij een "traditioneel artsexamen" wordt de examenkandidaat niet geobserveerd tijdens het afnemen van de anamnese en het verrichten van het lichamelijk onderzoek. Het cijfer wordt bepaald naar aanleiding van het mondeling examen na afloop van anamnese en onderzoek. Een dergelijk mondeling examen was tevoren op video opgenomen. Alle deelnemers werd gevraagd de examenkandidaat te beoordelen en een cijfer te geven. In figuur 1 is de verdeling van de cijfers te zien. Opvallend -maar niet onverwacht- is de grote spreiding tussen de beoordelaars. Deze grote spreiding zou kunnen worden verklaard door het verschil in achtergrond van de deelnemers. Het oordeel van examinatoren met dezelfde achtergrond blijkt echter ook wei-

nig overeenkomst te vertonen (3 4).

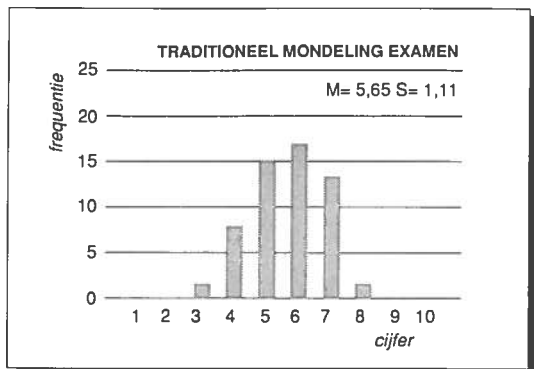


Fig. 1. Verdeling van de cijfers die bij beoordeling van dezelfde kandidaat, door de deelnemers zijn gegeven naar aanleiding van het op video opgenomen traditioneel mondeling examen.

Voor de National Board of Medical Examiners waren dergelijke gegevens in 1965 aanleiding om het klassieke artsexamen af te schaffen (5).

- gestructureerd mondeling examen; Bij een gestructureerd mondeling examen zijn de vragen die de examiner stelt, tevoren door een panel van deskundigen vastgesteld. Er is ook van tevoren een criterialijst opgesteld, met behulp waarvan de examiner de kandidaat beoordeelt. Dit alles met de bedoeling om meer gestandaardiseerd te examineren. Een gestructureerd mondeling examen werd voor deze dag op video opgenomen. De deelnemers werd gevraagd de prestatie van de student te beoordelen en een cijfer te geven. In figuur 2 is de cijferverdeling weergegeven.

Helaas is evenals bij het traditioneel mondeling examen, het oordeel van de deelnemers over de kandidaat niet éénduidig. In figuur 2 zijn de cijfers van alle deelnemers weergegeven. Dit is niet

helemaal correct, omdat de cijfers op twee verschillende manieren tot stand zijn gekomen. De helft van de deelnemers heeft namelijk alleen een cijfer gegeven, de andere helft heeft eerst een criterialijst ingevuld voordat het cijfer werd gegeven.

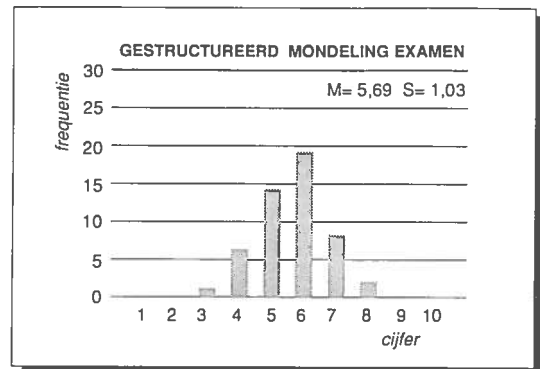


Fig. 2. Cijferverdeling naar aanleiding van een op video opgenomen gestructureerd mondeling examen, waarbij alle deelnemers dezelfde kandidaat beoordeelden.

Uit verdere analyse blijkt, dat de deelnemers die na het invullen van de criterialijst een cijfer hebben gegeven, gemiddeld een lager cijfer geven dan de deelnemers, die alleen een cijfer hebben gegeven. Op de spreiding van de cijfers heeft het invullen van de lijst echter weinig invloed.

De overeenstemming tussen de beoordelaars wat betreft het invullen van de criterialijst is uitgezocht door de somscore van de items te berekenen. Uit figuur 3 blijkt, dat de beoordelaars vrijwel allemaal dezelfde somscore hadden. Dit wil nog niet zeggen dat de beoordelaars dezelfde items positief of negatief hebben beoordeeld. Hiervoor moet de overeenkomst op item-niveau worden uitgerekend. Het oordeel van de deelnemers blijkt niet alleen wat betreft de somscore maar ook op itemniveau een grote mate van overeenstemming te

vertonen.

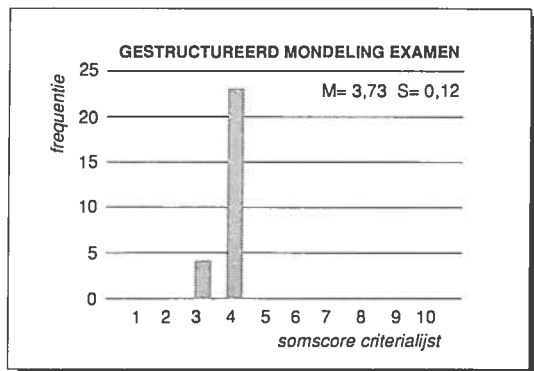


Fig. 3. Van de deelnemers die het op video opgenomen gestructureerd mondeling examen met een criteria-lijst oftewel observatieformulier beoordeelden, zijn de somscores van de items in deze figuur weer-gegeven.

De deelnemers waren kennelijk dus wel in staat om een observatielijst goed in te vullen, zelfs zonder training. Daarentegen geeft het wegen van de kandidaat -het uitdrukken van de prestatie in een cijfer- problemen.

- gestructureerde open vraag (g.o.v.);

Dit is een schriftelijke toetsvorm, waarbij de kandidaat stap voor stap verschillende fasen van een medisch probleem doorneemt. Elke fase bestaat uit informatie over het probleem gevolgd door één of meer open vragen. De deelnemers werd gevraagd een cijfer te geven over een door een student ingevulde g.o.v. De helft van de deelnemers gaf alleen een cijfer, de andere helft vulde eerst een beoordelings-lijst in. Ondanks een zekere over-eenkomst met het gestructureerd mondeling examen verloopt hier het beoordelingsproces op een aantal punten duidelijk anders.

Uit figuur 4 blijkt, dat er een grote spreiding is wat betreft de cijfers. Het invullen van een cri-

terialijst had weinig invloed op de spreiding van het gegeven cijfer.

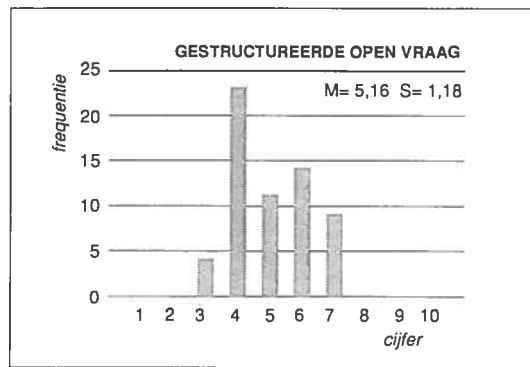


Fig. 4. Frequentieverdeling van de cijfers die door de deelnemers werden gegeven naar aanleiding van een -door een student ingevulde- gestructureerde open vraag (g.o.v.). Alle deelnemers beoor- deelden dezelfde g.o.v.

Ook wat betreft de somscore van de criteria-lijst is er weinig overeenstemming tussen de beoor- delars. Bij deze g.o.v. zijn de examinatoren het dus op geen en- kele manier met elkaar eens.

Beschouwing

De discussie over de examenvor- men, die volgens het programma 's middags zou plaatsvinden, start- te bij vrijwel alle groepen in het eerste station. Als kennismaking met examenvormen en aanzet tot discussie over examenvormen kan deze dag ons inziens als geslaagd worden beschouwd. Uit de evalua- tie achteraf bleek, dat de deel- nemers deze mening deelden.

Zou er nu iets gaan veranderen in de praktische examens in de kli- nische fase? De deelnemers aan deze dag zijn waarschijnlijk niet representatief voor de examina- toren in Nederland. Er is sprake van een positieve selectie. Dit bleek ook uit een door hen van tevoren ingevulde vragenlijst.

Vrijwel alle deelnemers waren het eens over de noodzaak tot observatie bij klinische examens, het standaardiseren van de beoordeling en van de examensituatie, het trainen van examinatoren en het nemen van slaag/zak-beslissingen op grond van meerdere casus. Meer dan de helft van de deelnemers was zelfs van mening, dat een beoordeling van een examenkandidaat door meerdere examinatoren dient te geschieden. Als suggestie om veranderingen te bereiken, werd door meerdere deelnemers gezegd, dat het goed zou zijn als een dergelijke dag op elke faculteit georganiseerd zou worden. Hoewel de organisatoren hier volledig mee instemmen, zijn er nog twijfels over de haalbaarheid van dit plan.

Uit de gegevens die tijdens deze dag van een aantal stations zijn verzameld, is duidelijk dat het

oordeel van de deelnemers over de prestatie van een kandidaat weinig overeenstemming vertoonde; zoals bij het gestructureerd mondeling examen bleek, geeft met name het "wegen" van de kandidaat problemen. Tijdens deze dag werd ook duidelijk, dat het traditioneel mondeling examen niet als "kema-keur" voor de studie geneeskunde kan gelden. Voordat we kunnen spreken over een artsexamen als een goede toets voor de klinische competentie, moet er nog wel wat gebeuren. Dit soort problemen bij het toetsen van klinische competentie bestaat niet alleen in Nederland. In 1985 en in 1987 is er in Ottawa een conferentie aan dit thema gewijd. In de proceedings van deze conferenties is voor geïnteresseerden veel informatie over toetsen en problemen bij toetsen te vinden (zie ook de rubriek "Mededelingen").

Literatuur

1. Bulletin Medisch Onderwijs 1987; 6: 3-13.
2. Harden RMCG, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975; 1: 447-51.
3. Scherpbier AJJA, Bender W, Cohen-Schotanus J. Over de betrouwbaarheid van examinatoren. In: Zwierstra RP, Scherpbier AJJA, Vermey A, eds. Spieghel Chirurgicael. Assen: Van Gorcum, 1986: 78-81.
4. Metz JCM. Medische competentie. Proefschrift. Nijmegen, 1984.
5. Hubbard JP, Lenit EJ, Schumacher CF, Schnobel TG. On objective evaluation of clinical competence. N Engl J Med 1965; 272: 1321-8.