

TOUCHÉ?

EEN 'KLINISCHE LES' IN TOETSONTWIKKELING

G.M. VERWIJNEN
A.J.J.A. SCHERPBIER
L.H.C. TAN

In 1987 werd door het uitvoerend Bureau van het Samenwerkingsverband van het Interfacultair Overleg Huisartsgeneeskunde (SV-IOH) de Werkgroep Vaardighedenevaluatie ingesteld. De werkgroep had als taak om 'stations' te ontwikkelen voor de landelijke toetsing van medische vaardigheden van huisartsen-in-opleiding.¹ De term 'station' wordt gebruikt bij het toetsen van vaardigheden volgens het model van het 'Objective Structured Clinical Examination' (OSCE).² De toets bestaat uit een circuit van verschillende 'stations', waarbij de kandidaat in ieder 'station' een opdracht binnen een vastgestelde tijd dient uit te voeren. In Nederland spreekt men over stationsexamen, vaardighedentoets of gestructureerd praktisch examen. Inmiddels is er veel geschreven over deze toetsvorm.³⁻⁵ In dit artikel worden de perikelen rond het ontwikkelen van een toetsstation beschreven en aan het einde van het artikel wordt - als in een klinische les - een aantal kanttekeningen geplaatst bij het onderwijs en de toetsing van medische vaardigheden.

WAT MOET ER GETOETST WORDEN?

De eerste vraag waar de Werkgroep Vaardigheden zich mee geconfronteerd zag was: wat moet er getoetst worden?

Deze vraag kan beantwoord worden als duidelijk is wat een huisarts moet kennen en kunnen. De afgelopen jaren is er met ondermeer het Basistakenpakket van de huisarts en het Standaardenbeleid van het Nederlands Huisartsen Genootschap meer duidelijkheid gekomen over het kennen en kunnen van de huisarts.^{6,7} Ook zijn er voor de beroepsopleiding tot huisarts doelstellingen geformuleerd.⁸ Voor de vervaardiging van een toetsstation zijn bovengenoemde documenten echter te globaal. Uitgebreide beschrijvingen van concrete handelingen ontbreken. In het verleden hebben onderwijskundigen vaak aanbevolen om bovenal uitvoerig onderwijsdoelstellingen te specificeren, in de veronderstelling dat de concrete invulling van het onderwijs en de toetsing daar dan eenvoudig van afgeleid zouden kunnen worden. In de onderwijspraktijk is echter gebleken dat pogingen om deze route te volgen veel tijd en energie kosten, veel frustraties opleveren en bovendien vaak verzanden. Bender beschreef in zijn proefschrift dat er ook alternatieve

routes mogelijk zijn.⁹ Zo kan het - na het aangeven van globale aandachtsgebieden - vruchtbaarder zijn om meteen over te gaan tot de concretisering of operationalisering van impliciete doelstellingen in onderwijs of toetsing. De Werkgroep Vaardigheden koos deze laatste route. De productie van toetsstations kan in principe gebaseerd worden op de inhoud van het vaardigheidsonderwijs van de diverse huisartsinstituten door uit te gaan van aldaar beschikbare 'instructiestations'.¹⁰

HET INSTRUCTIESTATION 'RECTAAL TOUCHER'

Het vaardigheidsonderwijs krijgt voor huisartsen ondermeer gestalte via zogenaamde instructiestations. In de bijbehorende instructielijst staat gedetailleerd beschreven hoe een bepaalde vaardigheid uitgevoerd moet worden. Een van de eerste instructiestations was het instructiestation 'rectaal toucher', dat op het Nijmeegs Universitair Huisartsen Instituut tot stand is gekomen.¹¹ De werkgroep heeft dit station als basis genomen voor het ontwikkelen van een toetsstation

G.M. Verwijnen
Huisarts
Projectgroep Evaluatie
Studieresultaten
Rijksuniversiteit Limburg

A.J.J.A. Scherpbier
Arts
Skillslab
Rijksuniversiteit Limburg

L.H.C. Tan
Socioloog
Samenwerkingsverband
Interfacultair Overleg
Huisartsgeneeskunde (SV-IOH)
Utrecht

Correspondentieadres:
Projectgroep Evaluatie
Studieresultaten
J.G. van de Brink, secretaresse
Faculteit der Geneeskunde
Rijksuniversiteit Limburg
Postbus 616
6200 MD Maastricht

In dit instructiestation wordt het probleem gesimuleerd van een 66-jarige man die op het spreekuur van de huisarts verschijnt met mictieproblemen. Gesteld wordt dat het vermoeden bestaat op prostaatlijden en dat er besloten wordt om een rectaal toucher te verrichten ter nadere evaluatie van de klachten. Een simulatiepatiënt vertelt de klachten. Het rectaal toucher wordt niet bij de patiënt uitgevoerd, maar op een fantoom: een plastic model van het onderlichaam van een man, waarin zich een draaischijf met vijf verschillende 'prostaten' bevindt. Voor, tijdens en na het onderzoek op het fantoom geeft de huisarts-in-opleiding uitleg aan de simulatiepatiënt. In de instructielijst van in totaal 56 onderdelen is het gehele proces stap voor stap beschreven.

VERTALING VAN ONDERWIJS NAAR TOETS

Een vertaling van het instructiestation in een toetsstation zou betekenen dat de beheersing van alle aspecten van het rectaal toucher, zowel technische handeling als uitleg aan de patiënt, zouden worden getoetst. Veel van de items in het instructiestation zijn daarbij niet op het fantoom uit te voeren. Zo zit er geen ontlasting in het fantoom; het bekijken van de handschoen na het rectaal toucher om het aspect van de ontlasting te beoordelen is dan geen zinvolle handeling. Er zijn ook geen wondjes, fissuren of zwellingen te zien. Een groot deel van de handelingen kan zo worden tot het opdreunen van rijtjes. Het geven van uitleg aan een simulatiepatiënt terwijl het betreffende onderzoek niet op hem wordt uitgevoerd, is ook nogal gekunsteld. Het station zou realistischer worden als simulatiepatiënt en fantoom vervangen zouden worden door instructiepatiënten die de toets afnemen.

Instructiepatiënten verzorgen als docenten op een aantal faculteiten het onderwijs in ondermeer het rectaal toucher.¹² Hoewel het inzet-

ten van instructiepatiënten een goede manier zou zijn om een realistisch station te maken, was de werkgroep van mening dat deze vorm voor landelijke toetsing niet haalbaar was. Er werd gekozen voor de ontwikkeling van een station waarin alleen de vaardigheid van het rectaal toucher zou worden beoordeeld aan de hand van een fantoom. De kandidaat zou gevraagd worden een beschrijving te geven van de bevindingen bij rectaal toucher van vijf 'prostaten' en een diagnose te stellen. Een uitermate efficiënt station, waarmee in niet meer dan vijf minuten een indruk verkregen zou kunnen worden van het vermogen van de kandidaat om de prostaat te onderzoeken en te beoordelen.

De volgende vraag was of er gegevens waren over de beschrijving en de diagnose van de vijf 'prostaten'.

Het oorspronkelijke instructiestation was ontwikkeld om de techniek te oefenen. In de scoringslijst waren daarom geen beschrijvingen van de verschillende prostaten opgenomen. De fabrikant (CLA^R) van het fantoom geeft in de bijbehorende folder alleen de diagnoses van de verschillende prostaten.

Om meer informatie over het fantoom te krijgen werd een aantal urologen gevraagd om onafhankelijk van elkaar het fantoom te beoordelen, een beschrijving te geven van de bevindingen en een diagnose te stellen.

HOE DOEN UROLOGEN HET OP EEN PROSTAATFANTOOM?

In totaal werden twaalf urologen, werkzaam in het Academisch Ziekenhuis Maastricht of het Academisch Ziekenhuis Groningen, gevraagd een oordeel te geven over de kwaliteit van het fantoom, en vervolgens de verschillende prostaten te beschrijven en een diagnose te stellen.

Over de kwaliteit van het fantoom werden de volgende opmerkingen gemaakt: "De werkelijkheid is heel anders, maar het fantoom en met name de 'prostaten' geven wel een redelijk acceptabele indruk van wat je in werkelijkheid voelt". "Het rectum is te mobiel en de 'prostaten' verschuiven gemakkelijk." "Het

materiaal waaruit rectum en anus zijn opgebouwd is te stug.”

Er bestond bijzonder weinig uniformiteit in de beschrijvingen van de verschillende prostaat door de urologen, hoewel de overeenstemming over de uiteindelijke diagnoses groter was.

In tabel 1 zijn de verschillende diagnoses weergegeven.

Weinig overeenstemming bestond er over de indeling van kwaadaardige aandoeningen van de prostaat (prostaat nr. 3 en 4 in tabel 1). Ook bleken prostaathypertrofie en prostatitis moeilijk van elkaar te onderscheiden. In de dagelijkse praktijk is dit laatste onderscheid uiteraard eenvoudiger omdat de patiënt bij een ontsteking ook pijn zal aangeven.

WAT BETEKENEN DE BEVINDINGEN VOOR HET TOETSSTATION?

De verschillen tussen de beschrijvingen die de experts gaven, waren aanzienlijk. Hieruit werd de conclusie getrokken dat als specialisten niet overeenstemmen in hun beschrijvingen, het onredelijk lijkt huisartsen-in-opleiding te toetsen op het geven van een beschrijving. Immers, om de beschrijving van een kandidaat te beoordelen is een norm, in dit geval een eenduidige beschrijving, een vereiste. Ook de overeenstemming over de diagnoses bij de urologen was volgens de werkgroep te gering om de vertaling naar een toets situatie te wettigen. Deze overwegingen waren voor de Werkgroep Vaardighedenevaluatie aanleiding te stoppen met de verdere ontwikkeling van het toetsstation ‘rectaal toucher’.

VERSCHILLEN IN OORDELEN TUSSEN DOKTERS

Zijn de gevonden verschillen te verklaren? Een verklaring voor de verschillen tussen de beschrijvingen van de urologen zou kunnen zijn dat het fantoom in kwalitatief opzicht tekortschiet.

De urologen hadden wel kritiek op het fantoom, maar hun opmerkingen waren niet van dien aard, dat de verschillen hierdoor ver-

klaard kunnen worden. Een andere verklaring zou kunnen zijn dat de urologen het bij de beschrijving van onderzoeksbevindingen bij echte patiënten ook niet met elkaar eens zijn. We hebben geprobeerd dit in de literatuur na te gaan. Meyhoff en Hald publiceerden een onderzoek waarin bij elke patiënt door drie verschillende artsen de omvang van de prostaat pre-operatief werd geschat.¹³

Post-operatief werd het gewicht van het verwijderde prostaatweefsel bepaald. In totaal namen vijftien artsen en veertig patiënten deel aan het onderzoek. Het schatten van de omvang bleek een zeer inaccurate methode te zijn; ook vonden de auteurs een aanzienlijke

diagnose	aantal urologen
<i>Prostaat 1 'normal prostate'</i>	
- normaal, benigne	10
- vergrote linkerkwab	1
- dubieuze rechterkwab	1
<i>Prostaat 2 'congestion of the prostate'</i>	
- benigne prostaathypertrofie	9
- prostatitis	2
- normaal	1
<i>Prostaat 3 'beginning of carcinoma'</i>	
- T2 carcinoom rechts	6
- T3 carcinoom links	1
- T1 carcinoom rechts	1
- T3 carcinoom rechts	1
- carcinoom	2
- diffuus carcinoom	1
<i>Prostaat 4 'widespread carcinoma'</i>	
- T3 carcinoom rechts	6
- beperkt carcinoom rechts	1
- nodulaire prostaat	1
- zeer maligne	1
- T3 carcinoom links en rechts	3
<i>Prostaat 5 'adenoma of the prostate'</i>	
- prostatitis	4
- benigne hypertrofie	5
- hypertrofie III-IV	3

Tabel 1.
De verschillende diagnoses, die door de urologen zijn gesteld (tussen aanhalingstekens staat de diagnose die de firma in de handleiding van het fantoom vermeldt)

interdoktersvariatie. De onbetrouwbaarheid van het schatten van de omvang van de prostaat werd ook in een ander onderzoek aangetoond.¹⁴

Is het rectaal toucher wellicht een moeilijk onderzoek, waarvan de bevindingen moeilijk zijn weer te geven?

Guinan et al. vergeleken een aantal methoden voor screeningsonderzoek naar prostaatacarcinoom.¹⁵ Zij zijn van mening dat voor screening het rectaal toucher het meest efficiënte onderzoek is. Chodak et al. meldden dat het rectaal toucher een lage sensitiviteit heeft.¹⁶ In een vergelijkend onderzoek tussen het rectaal toucher en transrectaal echografisch onderzoek van de prostaat bleek dat de sensitiviteit van beide onderzoeken vergelijkbaar was, maar dat de specificiteit van het rectaal toucher laag was.¹⁷

Opvallend was dat in de drie laatstgenoemde artikelen over het rectaal toucher in het geheel geen aandacht werd besteed aan verschillen in beoordeling tussen dokters.¹⁵⁻¹⁷ De suggestie wordt hierdoor gewekt dat er helemaal geen probleem is. Het artikel van Meyhoff en Hald en de eerder in dit artikel beschreven bevindingen van de urologen zouden dan gewoon toevalsbevindingen zijn. Of ruimer geformuleerd, meestal beoordelen dokters met een hoge mate van overeenstemming. Helaas blijkt uit de literatuur dat dit laatste niet waar is.

Zo beschrijft Koran in twee artikelen een aantal studies, waarin de mate van overeenstemming tussen dokters is nagegaan.¹⁸⁻¹⁹ Hoewel de studies vaak kleine aantallen betreffen en het vaak om de beoordeling van geïsoleerde gegevens gaat, zoals bijvoorbeeld een ECG-beoordeling zonder verdere klinische gegevens, is Koran van mening dat er toch reden tot enige ongerustheid is. Zijns inziens moet geprobeerd worden om de betrouwbaarheid van waarnemingen te verhogen; om dit te bevorderen zouden volgens hem studenten en assistenten-in-opleiding vaker geobserveerd en beoordeeld moeten worden tijdens het afnemen van de anamnese en het verrichten van lichamelijk onderzoek.

Praktiserende artsen moeten volgens Koran in de gelegenheid worden gesteld om hun klinische bevindingen en interpretaties te kunnen blijven toetsen.

Ook in de Nederlandstalige literatuur is gepubliceerd over variatie tussen waarnemers.^{20,21} Veldhuyzen van Zanten en Hydra schrijven: "Er zijn goede redenen om de betrouwbaarheid van al ons dagelijks werk, en vooral van die waarnemingen die in de besluitvorming een cruciale betekenis hebben, niet voetstoots aan te nemen zonder deze eerst aan een grondig onderzoek te onderwerpen".²⁰ Ook schrijven zij dat er in de dagelijkse praktijk weinig rekening met variatie tussen waarnemers wordt gehouden. Lamberts ziet interdoktersvariatie als een "onmisbare professionele prikkel tot meer bescheidenheid en minder geneigdheid om bepaalde waarden snel tot norm te verheffen en daaraan vervolgens strak vast te houden".²¹ Hij pleit ervoor dat huisartsen met elkaar over interdoktersvariatie overleggen, proberen een norm vast te stellen en vervolgens onderzoek doen of de nieuwe norm in de praktijk leidt tot verbetering. Dit proces komt volgens Lamberts de kwaliteit van het huisartsgeneeskundig handelen ten goede.²¹

Wat betekenen de in het voorgaande beschreven perikelen en de informatie uit de literatuur voor de opleiding? Op deze vraag zal in de slotparagraaf worden ingegaan.

TEN SLOTTE

Als praktiserende artsen moeten leren omgaan met interdoktersvariatie, dan moet daar in hun opleiding voldoende aandacht aan worden besteed. In hun opleiding moeten huisartsen dus enerzijds leren elk onderzoek goed uit te voeren, dat wil zeggen, zo dat de betrouwbaarheid maximaal is. Anderzijds moeten zij echter ook de grenzen van deze betrouwbaarheid kennen, dus onderwezen krijgen. Hetzelfde geldt natuurlijk ook voor de basisopleiding. Ook studenten moeten enerzijds leren om zo betrouwbaar mogelijk informatie te verzamelen en anderzijds moeten zij de grenzen van deze betrouwbaarheid kennen. Wanneer moeten ze dat leren? In de co-assistentenschappen, is een voor de hand

liggend antwoord. Daar komen de studenten immers intensief in contact met de echte praktijk. Als het onderwijs voorafgaand aan de co-assistentenschappen ondermeer een voorbereidende functie heeft, dan moet interdoktersvariatie ook daar al aan de orde komen.

Zo wordt aan alle faculteiten, als voorbereiding op de co-assistentenschappen, vaardigheidsonderwijs gegeven.²² In dit onderwijs moet de betrouwbaarheid van waarnemingen uiteraard aan de orde komen. Maar ook bij onderwijs als patiëntendemonstraties en anatomie in vivo vergt het verschijnsel van de interdoktersvariatie de nodige aandacht.

Voor het onderwijs hebben bovenbeschreven perikelen ons inziens dus consequenties. Ook voor de toetsing, een niet onbelangrijk onderdeel van de opleiding, heeft een en ander consequenties.

In de toetsontwikkeling zijn we terecht bezorgd over variatie tussen examinatoren en casusspecificiteit.²³ Minstens even bezorgd moeten we zijn als artsen het kennelijk niet eens zijn over vele aspecten van het medisch handelen.¹⁸⁻²⁰ Voor het ontwikkelen van toetsen is het daarom dan ook van belang dat steeds de vraag wordt gesteld: *is het reëel?* Met andere woorden, worden er reële eisen aan de kandidaten gesteld. Het heeft immers geen zin om iets te willen toetsen waar dokters in de praktijk het niet met elkaar over eens zijn. Of anders geformuleerd: bij de ontwikkeling van toetsen moet rekening worden gehouden met de resultaten van onderzoek naar interdoktersvariatie.

Een andere belangrijke vraag, die steeds gesteld moet worden bij de ontwikkeling van toetsen, is de vraag: *wat moet er getoetst worden?* Als bijvoorbeeld de essentie van de te-toetsen-vaardigheid voornamelijk kennis is, kan beter gekozen worden voor een schriftelijke toets.

Ook kan - door het herhaaldelijk stellen van de vraag naar de essentie -, het ontwikkelen en uitvoeren van een onnodig ingewikkeld

station worden voorkomen. Helaas kon het hier besproken station de toets der realiteit niet doorstaan, hetgeen illustreert dat steeds beide vragen (*'Is het reëel?'* en *'Wat is de essentie?'*) gesteld moeten worden om kwalitatief goede toetsen te waarborgen.

Voor het goed beantwoorden van deze vragen zijn discussies tussen deskundigen met verschillende achtergrond nodig.

Dokters kunnen beoordelen of de toets relevant is. *Toetsontwikkelaars* kunnen aangeven of de gekozen opzet haalbaar is. *Psychometrici* kunnen aangeven of met de gekozen toetsopzet betrouwbare resultaten kunnen worden verkregen. Een dergelijke interdisciplinaire samenwerking geeft garanties dat er kritisch naar een toets wordt gekeken en dat komt de kwaliteit van de toets ten goede.

WIJS ONDERWIJS

In het onderwijs behoort van meet af aan aandacht besteed te worden aan de resultaten van onderzoek naar interdoktersvariatie en de empirische basis van diagnostiek en behandeling. Zo leren studenten beseffen dat lang niet alle 'zekerheden' een even deugdelijk fundament hebben en kan 'onwijze' toepassing van kennis en vaardigheden voorkomen worden. Het getuigt bijvoorbeeld niet van een wetenschappelijke benadering van het onderwijs als we studenten wijsmaken dat ze via het rectaal toucher de grootte van de prostaat kunnen vaststellen, terwijl we weten dat experts daar in de praktijk al weinig van terecht brengen.

In examens waarmee zak/slaag beslissingen worden genomen, behoren zelfs in het geheel geen onderwerpen voor te komen waarvoor een deugdelijk wetenschappelijk fundament ontbreekt!

LITERATUUR

1. Pollemans MC, Tan LHC. Toetsing van kwaliteit. Rapport Samenwerkingsverband Interfacultair Overleg Huisartsgeneeskunde, 15. Utrecht: Samenwerkingsverband-IOH, 1990.
2. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; 13: 41-54.
3. Hart IR, Harden RM, Walton HJ, eds. Assessment of clinical competence. Montreal: Heal Publications, 1986.
4. Hart IR, Harden RM, eds. Newer developments in assessment of clinical competence. Montreal: Can Heal Publications, 1988.
5. Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, eds. Teaching and assessing clinical competence. Groningen: BoekWerk, 1990.
6. Springer MP, ed. Basistakenpakket voor de huisarts. Utrecht: Landelijke Huisartsen Vereniging, 1986.
7. Tielens V. Om de kwaliteit van de huisarts. *Huisarts en Wetenschap* 1987; 30: 166-8.
8. Pollemans MC, Verheij Th. Onderwijsdoelstellingen voor de meerjarige beroepsopleiding tot huisarts. Utrecht, 1986; Bijlage bij rapport CCBOH-08.
9. Bender W. Studeergedrag van medisch studenten en didactische mogelijkheden voor docenten. Rijksuniversiteit Groningen, 1985. Dissertatie, 74 pp.
10. Tan LHC, Van Geldorp G, Foolen CHGM. Handleiding ter ontwikkeling van instructiestations medisch technische vaardigheden. Utrecht: Samenwerkingsverband IOH, 1989.
11. De Quay M. Instructiestation rectaal toucher. Nijmegen: Nijmeegs Universitair Huisartsen Instituut, 1989.
12. Andringa H. Het andrologisch onderzoek; een overzicht van praktische vaardigheden. *Bulletin Medisch Onderwijs* 1988; 7(3): 5-8.
13. Meyhof HH, Hald T. Are doctors able to assess prostatic size? *Scand J Urol Nephrol* 1978; 12: 219-21.
14. Meyhoff HH, Ingeman L, Nordling J, Hald T. Accuracy in preoperative estimation of prostatic size. *Scand J Urol Nephrol* 1981; 15: 45-51.
15. Guinan P, Bush I, Roy V, Vieth R, Roo R, Bhatti R. The accuracy of the rectal examination in the diagnosis of prostate carcinoma. *N Engl J Med* 1980; 303(9): 499-503.
16. Chodak GW, Keller P, Schoenberg H. Routine screening for prostate cancer using the digital rectal examination. EORTC genitourinary group monographs: progress and controversies in oncological urology 1988; 87-95.
17. Clements R, Griffiths GJ, Peeling WB, Roberts EE, Evans KT. How accurate is the index finger? A comparison of digital and ultrasound examination of the prostatic nodule. *Clinical Radiology* 1988; 39: 87-9.
18. Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975; 293: 642-6.
19. Koran LM. The reliability of clinical methods, data and judgments. *N Engl J Med* 1975; 293: 695-701.
20. Veldhuyzen van Zanten SJO, Hydra A. Onderzoek naar de variatie tussen waarnemers met behulp van kappa. *Ned Tijdschr Geneesk* 1988; 132(5): 199-202.
21. Lamberts H. Interdokersvariatie en de kwaliteit van huisartsgeneeskundig handelen. *Huisarts en wetenschap* 1986; 29: 146-152.
22. Metz JCM, Scherpbier AJJA. Systematische vaardigheidstraining als voorbereiding op de co-assistentenschappen. *Ned Tijdschr Geneesk* 1989; 133: 361-4.
23. Van der Vleuten CPM. Toetsing van medische competentie: stand van zaken en ontwikkelingen. In: Scherpbier AJJA, Metz JCM, Van der Vleuten CPM, eds. *Gezond Onderwijs. Speciale uitgave Bulletin Medisch Onderwijs* 1991; 10(2): 52-63.