

ONTWIKKELING VAN DE LANDELIJKE CONSULTVOERINGSTOETS VOOR DE BEROEPSOPLEIDING TOT HUISARTS

L.H.C. Tan
Socioloog
Uitvoerend Bureau landelijk
Samenwerkingsverband-IOH
Utrecht

C.H.G.M. Foolen
Huisarts
Voormalig medewerker
Uitvoerend Bureau

C.P.M. van der Vleuten
Psycholoog
Project Evaluatie Studieresultaten
Rijksuniversiteit Limburg

Correspondentieadres:
Uitvoerend Bureau landelijk
Samenwerkingsverband-IOH
Bijlhouwerstraat 6
3511 ZC Utrecht

In de beroepsopleiding tot huisarts worden huisartsen-in-opleiding onder meer getraind in de vaardigheden die nodig worden geacht voor het verrichten van een adequaat huisartsgeneeskundig consult. Om na te kunnen gaan in hoeverre twee belangrijke aspecten van deze vaardigheden, namelijk 'gericht en systematisch werken' en 'de arts-patiënt communicatie', aan het eind van de beroepsopleiding daadwerkelijk worden beheerst, heeft het Samenwerkingsverband van de acht universitaire huisartsinstituten (SV-IOH) een instrument ontwikkeld dat bestemd is voor de landelijke toetsing van deze vaardigheden.¹ In dit artikel wordt verslag gedaan van de wijze waarop de Landelijke Consultvoeringstoets (LACONTO) tot stand kwam en worden de resultaten gepresenteerd van de eerste experimentele toetsafname met het ontwikkelde instrument. Eerder zijn in dit tijdschrift artikelen verschenen over de activiteiten van het landelijk Samenwerkingsverband met betrekking tot toetsing op het gebied van medisch-technische vaardigheden en kennis en de evaluatie van het instituutsonderwijs.²⁻⁴

INSTRUMENTONTWIKKELING

Uitgangspunten

Bij de constructie van het landelijke consultvoeringsinstrument is gebruik gemaakt van de resultaten uit onderzoeks- en ontwikkelingswerk in Nijmegen, Utrecht en Maastricht. Daar werden ten behoeve van onderwijs en onderzoek drie meetinstrumenten ontwikkeld op het gebied van consultvoering: in Nijmegen de PREVARA-SETH, in Utrecht de UCEM en in Maastricht de MAAS-R.⁵⁻⁷

Voor LACONTO geldt, evenals voor alle andere toetsingsinstrumenten die door het Samenwerkingsverband ontwikkeld worden, dat deze is gebaseerd op het LHV-Basistakenpakket van de huisarts, dat door de beroepsgroep is vastgesteld.⁸ De toetsinstrumenten van het Samenwerkingsverband zijn steeds gericht op die aspecten van huisartsgeneeskundige competentie waarover landelijk overeenstemming bestaat. De landelijke consultvoeringstoets is op te vatten als een 'grootste gemene deler', waarin de kernelementen vervat zijn die in elke plaatselijke visie op consultvoering van belang worden geacht.

In het Structuurplan voor de beroepsopleiding tot huisarts, dat van het LHV-Basistakenpakket is afgeleid, worden aan de vaardigheden van de huisarts op het gebied van consultvoering vier dimensies onderscheiden: 'gericht en systematisch werken', 'hanteren van de arts-patiënt relatie', 'soma-tisch handelen' en 'psychosociaal handelen'.⁹ Bij bestudering van de drie beschikbare lokaal ontwikkelde instrumenten bleek dat deze alleen inhoudelijk overeenstemden op de eerste twee dimensies ('gericht en systematisch werken' en 'hanteren van de arts-patiënt relatie'). Uit onderzoek wordt steeds duidelijker dat voor toetsing van de laatste twee dimensies ('soma-tisch handelen' en 'psychosociaal handelen') scoringslijsten vereist zijn die specifiek op de inhoud van het te toetsen probleem zijn toegesneden.¹⁰ Er werd dan ook besloten om het landelijk toetsingsinstrument *alleen* te richten op 'het gericht en systematisch werken' en 'het hanteren van de arts-patiënt relatie'.

Bij de ontwikkeling van de eerste versie van het landelijk instrument werd *inhoudelijk* gebruik gemaakt van de drie bestaande consultvoeringsinstrumenten. Elementen die in elk van de drie instrumenten voorkwamen, werden per definitie opgenomen in het landelijk instrument; elementen die in twee van de drie

instrumenten voorkwamen, werden op een enkele uitzondering na opgenomen en elementen die slechts in één van de instrumenten voorkwamen, werden niet opgenomen in het nieuwe instrument.

De *meettechnische en vormkenmerken* van het landelijk instrument werden bepaald op basis van de drie bestaande instrumenten en de methoden beschreven in recente psychometrische literatuur.¹⁰

Gekozen werd voor scoringslijsten met vrij globale items. Voor deze keuze pleitte een aantal argumenten. Ten eerste bleek uit de literatuur dat scoringslijsten met globale items een voldoende betrouwbaarheid kunnen opleveren onder de voorwaarde dat de beoordelaars goed getraind zijn en er duidelijke scoringsafspraken zijn gemaakt.^{11 12} Ten tweede was bijerdere vaardigheidentoetsing van het Samenwerkingsverband gebleken dat zeer gedetailleerde scoringslijsten, door ervaren huisartsen, inhoudelijk als onvoldoende relevant voor het niveau van de praktiserende huisarts werden beoordeeld.¹³ Ten derde was uit de literatuur bekend dat gedetailleerde scoringslijsten het 'domweg uit het hoofd leren van handelingen' in de hand werken.¹⁴

Wat betreft de *beoordelingsschaal* werd gekozen voor een combinatie van een checklist ('niet gedaan'/'gedaan') en een rating-scale ('slecht'/'goed gedaan' op een vijfpuntschaal). Een dergelijke beoordelingsschaal maakt het mogelijk om enerzijds aan te geven of een handeling al dan niet is verricht, en anderzijds aan te geven in hoeverre de verrichte handeling goed of fout is uitgevoerd. Beide aspecten zijn van belang voor het geven van concrete aanwijzingen (feedback) voor verdere scholing.

Ten aanzien van de *aard* van de consulten werd besloten dat het toetsinstrument toepasbaar moest zijn op alle soorten consulten: initiële consulten (huisarts en patiënt zien elkaar voor het eerst), eerste consulten (de patiënt komt met een nieuwe klacht) en vervolgsconsulten.

Commentaarronden

De concept-versie van het landelijk instrument werd in twee schriftelijke rondes voorgelegd aan 27 commentatoren afkomstig van de acht huisartsinstituten en van de Stichting Deskundigheidsbevordering Huisartsen (SDH). De groep commentatoren bestond uit vijftien huisartsen en twaalf gedragswetenschappers. Deze hadden vrijwel allen ervaring met het geven van onderwijs op het gebied van consultvoering. Elke commentator gaf aan de hand van een vragenlijst schriftelijk commentaar op de inhoudelijke, meettechnische en vormkenmerken van het nieuwe instrument.

In de eerste commentaarronde was de respons 96%, in de tweede 67%. Op grond van het geleverde commentaar werd het concept-instrument bijgesteld. Deze bijstelling had voornamelijk betrekking op de inhoud van het instrument.

Landelijk instrument voor gericht en systematisch werken en arts-patiënt communicatie

Het landelijk instrument bestaat uit een scoringslijst van één pagina en een aparte toelichting. In de toelichting zijn de dertig items van de scoringslijst nader omschreven. Zeventien items hebben betrekking op de dimensie 'gericht en systematisch werken' en dertien op de dimensie 'arts-patiënt communicatie'. Om de herkenbaarheid van de twee dimensies te bewaren voor de gebruikers van het instrument, zijn de items met betrekking tot de arts-patiënt communicatie op de scoringslijst cursief weergegeven (figuur 1). De items zijn verdeeld over drie fasen van het consult: intake (tien items), diagnostiek (tien items) en beleid (acht items). De scoringslijst wordt afgesloten met twee items waarin naar een algemene indruk gevraagd wordt over de twee dimensies afzonderlijk.



Truets:
Inst.:
Huis:

Groep:
Patiënt:
Observator:

SCORINGSLIJST GERICHT EN SYSTEMATISCH WERKEN EN ARTS-PATIENT COMMUNICATIE

		nee					zeer					zeer					
		slecht					slecht					goed					
		0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	5
INTAKE																	
1.	Begroet patiënt, stelt zich zo nodig voor, opent consult	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.	Heeft uitnodigende houding naar patiënt (luistert, geeft aandacht)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.	Inventariseert klachten/problemen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.	Verheldert hulpvraag (reden van komst, klachtheleving, wensen/verwachtingen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.	Gebruikt bij vraagverheldering passende gesprekstijlen (doorvragen, samenvatting, geveel, reflectie)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6.	Besprekt behandeling (al nu toe (incl. zelfzorg))	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7.	Vat probleem en hulpvraag samen (teknopt, eigen woorden, inhoudelijk correct)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8.	Vraagt om reactie patiënt op samenvatting probleem en hulpvraag, besprekt reactie zo nodig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9.	Kondigt informatieverzameling aan (klachtgericht vragen, lichamelijk onderzoek)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10.	Vraagt instemming patiënt niet overgaan tot informatieverzameling, overlegt zo nodig	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
DIAGNOSTIEK																	
11.	Verzamelt informatie geordend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12.	Werkt op meer sporen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13.	Gebruikt bij klachtgericht vragen passende gesprekstijlen (gericht vragen, leiding geven)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14.	Geeft uitleg m.b.t. lichamelijk onderzoek (voor en tijdens)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15.	Bejegt patiënt zorgvuldig en respectvol tijdens lichamelijk onderzoek	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ALGEMENE INDRUK																	
Gericht en systematisch werken																	
Arts-patiënt communicatie																	

Figuur 1.
Scoringslijst LACONTO
Scoringslijst gericht en
systematisch werken en
arts-patiënt communicatie

De items worden gescoord op een zespuntschaal, die varieert van 'nee' (score nul) tot 'zeer goed' (score vijf). De twee uitersteschaalpunten (nul en vijf) zijn inhoudelijke omschreven ('verankerd') in de toelichting. Bij voorbeeld: bij item drie houdt een score vijf in dat 'de arts nagaat of er naast de spontaan genoemde klachten en problemen nog meer klachten of problemen zijn waarvoor de patiënt de arts bezoekt' en een score nul dat de arts de voor score vijf omschreven handeling in het geheel niet heeft verricht.

EXPERIMENTELE TOETSAFNAME

Doel

Voor het uittesten van het landelijk instrument werd een experimentele toetsafname georganiseerd. Het doel van deze toetsafname was drieledig: zicht krijgen op de *betrouwbaarheid*, de *validiteit* van het landelijk instrument en op de *uitvoerbaarheid* van een landelijke toetsafname.

Uit de literatuur is bekend dat voor de toetsing van consultvoeringsvaardigheden door gaans veel casus nodig zijn om een stabiel of betrouwbaar beeld van een kandidaat te verkrijgen.⁹ Bovendien wordt de betrouwbaarheid beïnvloed door de mate van overeenstemming tussen beoordelaars. Het onderzoek naar de betrouwbaarheid werd daarom gericht op het aantal casus en het aantal beoordelaars dat naar schatting nodig was voor het verkrijgen van een voldoende betrouwbaar toetsresultaat. De discipline van de beoordelaar (huisarts of gedragswetenschapper) werd tevens in de analyses betrokken.

Het onderzoek naar de validiteit werd in het experiment gericht op de beperkte vraag in hoeverre het landelijk instrument in staat is verschillen aan te tonen tussen beginnende en gevorderde huisartsen-in-opleiding. Hiermee zou een indruk worden verkregen van de constructvaliditeit van het instrument. De verwachting was dat de gevorderde huisartsen-in-opleiding hoger zouden scoren dan de beginnende huisartsen-in-opleiding. Aan de inhoudsvaliditeit van het instrument is door het raadplegen van de beschikbare inhoudsdeskundigen op het gebied van consultvoeringsvaardigheden, de grootst mogelijke aandacht besteed.

Toetsopzet

De experimentele toetsafname vond plaats bij veertig huisartsen-in-opleiding. Deze waren afkomstig van twee instituten die niet beschikten over een eigen consultvoeringsinstrument. De twee instituten leverden elk één groep huisartsen-in-opleiding die ongeveer vier maanden in opleiding waren ('beginners') en één groep huisartsen-in-opleiding die ongeveer zestien maanden in opleiding waren ('gevorderden' op twee derde van de opleidingstijd). Door deze toetsopzet kon worden nagegaan of het instrument in staat was verschillen tussen de beginners en gevorderden aan te tonen (constructvaliditeit).

Elke huisarts-in-opleiding deed vier consulten met vier verschillende simulatiepatiënten. De aangeboden patiëntenproblematiek was gebaseerd op de NHG-standaarden Diabetes Mellitus Type II, Urineweginfecties, Enkel-dorsie en Schouderklachten.¹⁵⁻¹⁸ De keuze van de NHG-Standaarden maakte het mogelijk om samen te werken met de Werkgroep Onderzoek Kwaliteit (WOK) van de Katholieke Universiteit Nijmegen en de Rijksuniversiteit Limburg, die scoringslijsten ontwikkelt voor het beoordelen van het medisch handelen.¹⁹ Over de ervaringen met deze WOK-lijsten is separaat gerapporteerd.²⁰

Gekozen werd voor simulatieconsulten in plaats van reële praktijkconsulten omdat elke huisarts-in-opleiding op deze wijze op dezelfde vier patiënten getoetst kon worden en maximaal vergelijkbare toetsresultaten verkregen zouden worden. Bovendien blijkt uit de literatuur dat simulatiepatiënten in toets-situaties - waarin het erom gaat kandidaten onder zo gelijk mogelijke omstandigheden hun kunnen te laten demonstreren - te prefereren zijn boven echte patiënten aangezien zij een grotere accuratesse vertonen en consistent in hun gedragingen zijn.¹⁰ Op voorhand was te verwachten dat een aantal van vier consulten onvoldoende zou zijn voor het verkrijgen van betrouwbare resultaten in dit experiment, maar dit aantal werd voldoende geacht om een schatting te kunnen maken van het benodigde aantal consulten voor toekomstige toets-situaties.

De consulten werden op videoband opgenomen. Ten opzichte van directe observatie had dit het voordeel dat de beoordelaars niet tegelijkertijd op één plaats aanwezig behoeften te zijn. Bovendien zouden de consulten zonodig meerdere malen beoordeeld kunnen worden.

Om een indruk te krijgen van het oordeel van de simulatiepatiënten over de wijze waarop de huisarts-in-opleiding met hen was omgegaan tijdens het consult, werd door hen na afloop van ieder consult een korte 'bejegeningsslijst' ingevuld. Deze bejegeningsslijst was afgeleid van een lijst die in 1988 in de landelijke experimentele vaardigheidentoets van het Samenwerkingsverband was gebruikt.²¹ Om de mening van de huisartsen-in-opleiding over de toets te verkrijgen werd door hen na afloop van de toetsafname een evaluatieformulier ingevuld.

Scoring van de consulten

De video-opnames van de in totaal 160 consulten werden met behulp van het landelijk instrument gescoord door acht huisartsen en acht gedragswetenschappers. Deze hadden allen deelgenomen aan beide commentaarronden die bij de samenstelling van het concept-instrument waren gehouden. In de groep van beoordelaars waren alle acht huisartsinstituten en de voormalige Stichting Deskundigheidsbevordering Huisartsen vertegenwoordigd. De beoordelaars kregen van tevoren een training van één dagdeel, waarin met het nieuwe instrument werd geoefend en aanvullende scoringsafspraken werden gemaakt. Elk consult werd door een beoordelaarspaar, bestaande uit hetzij twee huisartsen hetzij twee gedragswetenschappers, gescoord. De beoordelaars scoorden onafhankelijk van elkaar twintig consulten van beginners of van gevorderden. Zij waren niet op de hoogte van de opleidingsfase van de kandidaten.

Feedback

Elke deelnemende huisarts-in-opleiding ontving direct na de toetsafname een kopie van de door 'hun' patiënten ingevulde bejegeningsslijsten.

Na de verwerking van de gegevens kregen de huisartsen-in-opleiding een overzicht van de individuele toetsresultaten per consult, uitgesplitst naar consultfase, algemene indruk en totaalscore. Ter vergelijking ontvingen zij ook de gemiddelde resultaten van de eigen opleidingsgroep en die van de huisartsen-in-opleiding die in dezelfde opleidingsfase waren (beginners of gevorderden). De groepsbegeleiders kregen een kopie van de groepsoverzichten.

RESULTATEN

Scores huisartsen-in-opleiding

Per kandidaat werden scores berekend voor de consultfasen (intake, diagnostiek, beleid), de algemene indruk en voor de totale scoringslijst. Tevens werden scores berekend voor de twee onderscheiden dimensies 'gericht en systematisch werken' en 'arts-patiënt communicatie'. De scores werden weergegeven als percentage van de maximaal te behalen score. Bij de berekening werden de scores van het beoordelaarspaar gemiddeld.

In tabel 1 wordt een overzicht gegeven van de gemiddelde resultaten van de beginnende en de gevorderde huisartsen-in-opleiding per casus en voor de vier casus tezamen.

De gemiddelde scores per casus vertonen aanzienlijke verschillen, dit geldt ook voor de standaarddeviaties. Zo ligt voor de intake-fase de gemiddelde score het hoogst voor casus III (Enkeldistorsie) bij de gevorderden en het laagst voor casus IV (Schouderklachten) eveneens bij de gevorderden. Opvallend is dat zowel bij casus I (Diabetes) en bij casus IV (Schouderklachten) de beginners, tegen de verwachting in, op alle aspecten gemiddeld hoger scoren dan de gevorderden.

CASUS I: DIABETES MELLITUS

	Beginners (N=20)		Gevorderden (N=20)	
	<i>gemid.</i>	<i>sd</i>	<i>gemid.</i>	<i>sd</i>
Intake	35.40	8.49	31.75	9.06
Diagnostiek	41.44	8.39	36.55	8.85
Beleid	58.00	12.46	46.88	9.29
Algemene indruk	61.50	14.52	57.25	10.82
Totale scoringslijst	45.45	8.48	39.29	7.17
Gericht en systema- tisch werken	45.34	8.28	38.50	6.91
Arts-patiënt commu- nicatie	45.58	9.67	40.36	9.01

CASUS II: URINEWEGINFECTIE

	Beginners (N=20)		Gevorderden (N=20)	
	<i>gemid.</i>	<i>sd</i>	<i>gemid.</i>	<i>sd</i>
Intake	22.71	10.72	29.35	7.60
Diagnostiek	22.44	5.70	31.20	7.33
Beleid	42.69	10.94	49.67	10.95
Algemene indruk	37.75	8.19	59.47	11.41
Totale scoringslijst	29.45	6.63	38.00	6.53
Gericht en systema- tisch werken	30.54	7.20	38.59	6.15
Arts-patiënt commu- nicatie	28.00	7.91	37.06	8.89

CASUS III: ENKELDISTORSIE

	Beginners (N=20)		Gevorderden (N=20)	
	<i>gemid.</i>	<i>sd</i>	<i>gemid.</i>	<i>sd</i>
Intake	32.25	8.48	41.03	13.86
Diagnostiek	35.22	8.65	46.69	9.59
Beleid	39.13	7.77	47.70	6.76
Algemene indruk	45.50	9.72	67.50	14.19
Totale scoringslijst	35.98	6.43	46.18	8.27
Gericht en systema- tisch werken	38.84	7.28	49.60	9.02
Arts-patiënt commu- nicatie	32.46	6.21	41.92	7.83

CASUS IV: SCHOUDERKLACHTEN

	Beginners (N=20)		Gevorderden (N=20)	
	<i>gemid.</i>	<i>sd</i>	<i>gemid.</i>	<i>sd</i>
Intake	29.65	10.02	18.11	7.92
Diagnostiek	28.80	5.60	24.48	5.55
Beleid	28.76	8.70	27.69	10.56
Algemene indruk	44.25	9.77	36.58	9.58
Totale scoringslijst	30.10	5.67	24.03	5.05
Gericht en systema- tisch werken	32.45	7.02	24.69	5.05
Arts-patiënt commu- nicatie	27.04	5.68	23.17	5.64

VIER CASUS TEZAMEN

	Beginners (N=20)		Gevorderden (N=20)	
	<i>gemid.</i>	<i>sd</i>	<i>gemid.</i>	<i>sd</i>
Intake	30.00	6.54	30.30	7.71
Diagnostiek	31.97	4.30	34.67	5.11
Beleid	42.14	6.10	42.72	5.36
Algemene indruk	47.25	7.46	55.49	7.81
Totale scoringslijst	35.25	4.70	36.88	5.16
Gericht en systema- tisch werken	36.79	4.49	37.84	5.14
Arts-patiënt commu- nicatie	33.27	5.73	35.63	5.74

Tabel 1.
Scores beginners en
gevorderden per casus en
voor de vier casus
tezamen
(gemiddelden en
standaarddeviaties)

Betrouwbaarheidsstudie

Voor het experiment werd zowel de inter-beoordelaarsbetrouwbaarheid (waarbij verschillen tussen beoordelaars als voornaamste foutenbron worden beschouwd) als de totale toetsbetrouwbaarheid (waarbij verschillen tussen casus als foutenbron worden beschouwd) berekend. Voor de schatting van deze laatste betrouwbaarheid werd gebruik gemaakt van de generaliseerbaarheidstheorie, waarin de betrouwbaarheid van een instrument benaderd wordt vanuit de combinatie

van variantiebronnen die voor de betrouwbaarheid bepalend zijn.²²⁻²⁴ De generaliseerbaarheidsanalyse biedt de mogelijkheid om de bijdrage van afzonderlijke variantiebronnen (zowel foutenbronnen als gewenste variantiebronnen) aan de totale variantie te bestuderen. Doorgaans wordt voor toetsdoeleinden een generaliseerbaarheidscoëfficiënt van .80 als minimaal gewenste waarde voor de betrouwbaarheid beschouwd.

Tabel 2.

Interbeoordelaarsbetrouwbaarheid per casus naar discipline van beoordelaarspaar en voor de totale toets (intra-class correlatie-coëfficiënt met zowel gemiddelde als rangorde verschillen als error)

Per casus			Per discipline	
Casus I	Diabetes Mellitus	.00	Huisartsen	.29
Casus II	Urinaryweginfectie	.38	Gedragwetenschappers	.56
Casus III	Enkeldistorsie	.34		
Casus IV	Schouderklachten	.37		
Totale toets		.39		

Interbeoordelaarsbetrouwbaarheid

De interbeoordelaarsbetrouwbaarheid werd bepaald per casus, naar discipline van het beoordelaarspaar en voor de totale toets. De gegevens zijn weergegeven in tabel 2.

De interbeoordelaarsbetrouwbaarheid voor de casus Diabetes Mellitus blijkt extreem laag te zijn, die voor de drie overige casus (Urinaryweginfectie, Enkeldistorsie en Schouderklachten) middelmatig tot laag. De casus Diabetes Mellitus betrof als enige een vervolgsconsult, waarbij de huisarts-in-opleiding een diabetes-controle moest uitvoeren bij een patiënt die tot dan toe steeds door de huisarts-opleider was gecontroleerd.

De interbeoordelaarsbetrouwbaarheid, uitgesplitst naar de discipline van het beoordelaarspaar, ligt voor de gedragwetenschappers hoger dan voor de huisartsen: .56 versus .29. Het betrouwbaarheidsniveau voor de gedragwetenschappers benadert de waarde die gewoonlijk wordt gevonden in dit soort onderzoek (variërend van .50 tot .77).²⁵ De interbeoordelaarsbetrouwbaarheid voor de huisartsen ligt een stuk lager dan de empirisch gebruikelijke waarde. Voor de totale toets ligt de interbeoordelaarsbetrouwbaarheid op .39.

Totale toetsbetrouwbaarheid

Voor de bestudering van de totale toetsbetrouwbaarheid werden ten behoeve van de generaliseerbaarheidsstudie variantiecomponenten geschat.^{22,24} Om de niveau-effecten van beginners en gevorderden binnen de betrouwbaarheidsanalyse buiten beschouwing te laten, werden voor de twee onderscheiden opleidingsfasen eerst afzonderlijke variantiecomponenten geschat, die vervolgens werden gemiddeld. Deze gemiddelden zijn weergegeven in tabel 3.

In een toetssituatie gaat het om het onderscheid tussen kandidaten. De H-component (Huisartsen-in-opleiding) in tabel 3 geeft dit onderscheid weer. Deze factor is klein ten opzichte van de totale variantie, maar niet afwijkend van wat doorgaans wordt gevonden.¹⁰

De factor Casus (C) is daarentegen betrekkelijk groot. Dit duidt er op dat de gemiddelde scores per casus aanmerkelijke verschillen vertonen. De vier casus verschilden dus nogal in moeilijkheidsgraad. Component B:C omvat zowel het hoofdeffect B (dit is het effect van de strengheid in beoordelen van de Beoordelaars) als het interactie-effect BxC (de consistentie van de Beoordelaars over de vier Casus). Dit houdt in dat het niet mogelijk is om binnen het design, waarin per casus andere beoordelaarsparen gebruikt werden, het effect van de strengheid van de beoordelaars (B) en de consistentie van de beoordelaars over de vier casus (BxC) van elkaar te onderscheiden. Beide effecten zijn gering in dit ex-

Tabel 3.

Geschatte variantiecomponenten voor de totale toets gemiddeld over beide opleidingsniveaus

Variantiebron	Geschatte variantiecomponent	% van de totale variantie
H	11.80	7.62
C	48.14	31.07
B:C	2.11	1.36
HxC	7.70	4.97
HxB:C	85.17	54.98

H: Huisarts-in-opleiding

C: Casus

B: Beoordelaar

periment, aangezien component B:C slechts een fractie vormt van de totale variantie. De beoordelaars verschilden blijkbaar weinig in strengheid van beoordelen en de inhoud van de casus had blijkbaar weinig invloed op de consistentie in scoring van de beoordelaars. Component HxC geeft de mate aan waarin de huisartsen-in-opleiding op de verschillende casus hetzelfde rangnummer krijgen (anders gezegd: de mate waarin de casus 'het eens zijn' over de competentie van een kandidaat). Doorgaans is deze component, die ook wel met casus- of inhoudsspecificiteit wordt aangeduid, (één van) de grootste. Met nog geen tiende van de totale variantie is de component in het experiment relatief laag te noemen.

Component HxB:C weerspiegelt zowel de mate waarin de beoordelaars consistent zijn in hun beoordeling van de huisartsen-in-opleiding (HxB) als de algemene error (alle overige mogelijke foutenbronnen die van invloed zijn geweest op de meting). Beide effecten zijn niet te ontwarren in dit design. De algemene error is doorgaans vrij hoog. In combinatie met de tamelijk lage interbeoordelaarsbetrouwbaarheden (zie tabel 2) moet geconcludeerd worden dat de HxB component (consistentie van beoordelaars over huisartsen-in-opleiding) hiervan een aanzienlijk deel uitmaakt, omdat de overige variantiebronnen waarin de beoordelaars zijn betrokken relatief laag zijn en er niettemin lage interbeoordelaarsbetrouwbaarheden gevonden zijn. Dat impliceert dat de beoordelaars in hun scoring blijkbaar nogal inconsistent waren in hun oordeel over de verschillende kandidaten.

Concluderend kan gezegd worden dat de gewenste variantie in het experiment niet afwijkt van de resultaten bij andere metingen met behulp van gedragsobservatie, maar dat de foutenvariantie een andere opbouw heeft dan gebruikelijk.¹⁰ Met name blijkt de inhoudsspecificiteit een geringere rol te spelen dan gewoonlijk terwijl de beoordelaarsvariatie veel groter is. De relatief lage inhoudsspecificiteit wijst erop dat het bij het gebruikte instrument om een (relatief) casus-onafhankelijk meetinstrument gaat. De grote beoordelaarsvariatie doet vermoeden dat de huidige versie van het instrument, in combinatie met de beperkte training van één dagdeel, onvoldoende steun biedt aan de beoordelaars, met name aan de huisartsen.

Tabel 4.

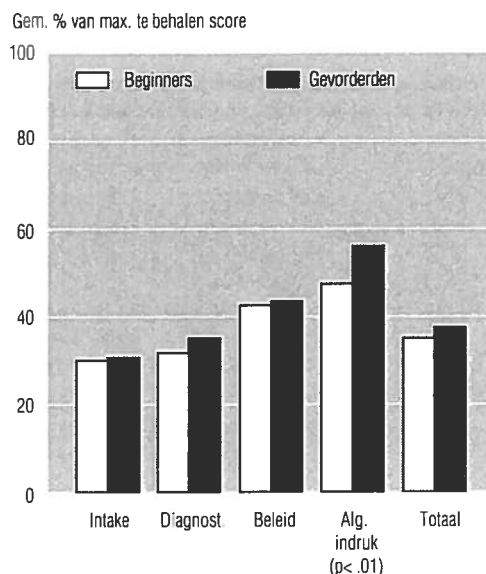
Generaliseerbaarheidscoëfficiënt voor de totale toets, als functie van het aantal casus en van aantal en discipline van de beoordelaars

Aantal casus	Generaliseerbaarheidscoëfficiënten		
	Twee beoordelaars	Eén beoordelaar <i>huisarts</i>	<i>gw-er</i>
4	.48	.36	.56
8	.65	.53	.72
12	.74	.63	.79
16	.79	.69	.84
20	.82	.74	.86

Om na te kunnen gaan hoeveel casus noodzakelijk zijn om een adequate totale toetsbetrouwbaarheid te bereiken, werden de generaliseerbaarheidscoëfficiënten als functie van het aantal casus berekend. In tabel 4 zijn de gegevens opgenomen voor zowel twee beoordelaars per consult, als voor één beoordelaar per consult (huisarts of gedragswetenschapper).

Uitgaande van twee beoordelaars per consult zou bij zestien casus een aanvaardbare betrouwbaarheid bereikt worden. Gezien de feitelijke tijd die gebruikt is door de huisartsen-in-opleiding per consult (nog geen tien minuten) zou dit een toetstijd van bijna drie uur per kandidaat inhouden. Wordt echter uitgegaan van één beoordelaar per casus, dan zou één *huisarts*-beoordelaar meer dan twintig casus noodzakelijk maken (waarmee een toetstijd van bijna drieëneenhalf uur per kandidaat gemoeid zou zijn). Bij één beoordelaar met *gedragswetenschappelijke* achtergrond wordt daarentegen een acceptabele betrouwbaarheid bereikt bij twaalf casus (dit is twee uur toetstijd per kandidaat).

Figuur 2.
Gemiddelde scores over
vier casus



Validiteit

Om de constructvaliditeit van het landelijk instrument te onderzoeken werd nagegaan in hoeverre het instrument in staat was verschillen tussen beginnende en gevorderde huisartsen-in-opleiding aan te geven. De gemiddelde scores van beginners en gevorderden over de vier casus tezamen zijn in figuur 2 grafisch weergegeven per consultfase (intake, diagnostiek en beleid), voor de algemene indruk en de toets als geheel.

Met uitzondering van de algemene indruk liggen de gemiddelde scores voor zowel beginners als gevorderden onder de 45% van de maximaal te behalen score. Opmerkelijk is dat alle verschillen in de gewenste richting gaan: de gemiddelde scores van de gevorderden zijn hoger dan die van de beginners. Alleen bij de algemene indruk is sprake van een statistisch significant verschil, waarbij de gevorderden hoger scoren dan de beginners ($F = 12.16$, $p = .001$, $df = 1$).

De verschillen tussen de twee instituten zijn, met uitzondering van beleid, statistisch significant ($p < .01$). Deze bevinding zou kunnen worden toegeschreven aan steekproef-fluctuaties (twee toevallig slechte groepen van Instituut 2) of aan een verschil in onderwijsaanbod op de twee instituten. Om de verschillen te kunnen interpreteren moet men informatie hebben over het onderwijs in consultvoering en de praktijkervaring voorafgaand aan de toetsafname. Deze informatie werd binnen het experiment niet verzameld.

De inhoudsvaliditeit werd gewaarborgd door de gekozen constructieprocedure waarin deskundigen op het gebied van consultvoering, afkomstig van alle universitaire huisartsinstituten alsmede van de Stichting Deskundigheidsbevordering Huisartsen, bij de instrumentontwikkeling betrokken waren.²⁶ Het feit dat vrijwel alle aangezochte deskundigen bereid waren om deel te nemen aan de commentaar- en beoordelingsrondes, evenals de kwaliteit en omvang van het geleverde commentaar, en de waardering van de commentatoren voor het uiteindelijke product, ondersteunen de inhoudsvaliditeit van het ontwikkelde instrument.

Uitvoerbaarheid landelijke toetsafname

De experimentele toetsafname vond om praktische redenen (beschikbare menskracht en middelen) plaats op twee instituten bij in totaal veertig huisartsen-in-opleiding. De organisatie verliep redelijk soepel; er traden, met uitzondering van één groep huisartsen-in-opleiding die onvoldoende bleek te zijn voorgelicht over de toets, geen noemenswaardige problemen op.

Uit de generaliseerbaarheidsanalyse blijkt dat naar schatting twee uur toetstijd per kandidaat nodig is om voldoende betrouwbare toetsresultaten te verkrijgen. Of een dergelijke toetstijd landelijk haalbaar is, daarover is op grond van het experiment geen uitsluitsel te geven. Ten aanzien van de beoordeling van de videobanden bleek dat het centraal trainen van beoordelaars, afkomstig uit het hele land (in groepjes van vier gedurende één dagdeel) niet steeds te verwezenlijken was. De dubbele beoordeling per consult bleek haalbaar in het experiment, maar bij uitbreiding van het aantal consulten en deelnemers aan de toets, zal dit zeker leiden tot grote problemen wat betreft het benodigde aantal beoordelaars.

Inpassing van de toetsing van consultvoeringsvaardigheden in het landelijk evaluatiesysteem lijkt goed mogelijk in relatie tot de toetsing van medisch-technische vaardigheden. Verwacht mag worden dat toetsing van consultvoerings- en medisch-technische vaardigheden gelijktijdig of in ieder geval aan de hand van eenzelfde consult kan plaatsvinden. Deze geïntegreerde vorm van toetsing van verschillende vaardigheden behoeft geen extra toetstijd te vergen.

SIMULATIEPATIËNT-OORDELEN

De vier simulatiepatiënten beoordeelden de consulten van de kant van de patiënt. Daartoe vulden zij een beoordelingslijst in met zes stellingen over de bejegening van de patiënt (tabel 5).

Over het geheel genomen blijken de simulatiepatiënten in ten minste 55% van de consulten de bejegening door de huisartsen-in-opleiding positief te beoordelen. Met name over de ruimte die men kreeg om zijn of haar verhaal te vertellen (stelling 2) waren de simulatiepatiënten, onafhankelijk van het opleidingsniveau van de arts, in 90% van de consulten tevreden. Getoetst aan het gemiddeld totaaloordeel (uitgedrukt in het percentage van de maximaal te behalen score) blijken de gevorderde huisartsen-in-opleiding statistisch significant hoger te scoren op de bejegeninglijst dan de beginners ($F = 7.92$, $p = .008$, $df = 1$). De oordelen van de simulatiepatiënten vertonen geen statistisch significante verschillen tussen de instituten in tegenstelling tot de consultvoeringsvaardigheden.

Om na te gaan of er een verband bestond tussen een positief hetzij negatief simulatiepatiëntoordeel en een hoge respectievelijk lage score op de LACONTO werd de correlatie berekend tussen de totaalscores op beide lijsten. Deze correlatie (gecorrigeerd voor attenuatie) bleek middelmatig tot laag te zijn (.36). Omdat de simulatiepatiënten vooral ook het communicatieve (bejegening-)aspect van de consulten beoordeelden, werd tevens gekeken naar de correlatie tussen simulatiepatiëntscore en de dimensie 'arts-patiënt communicatie' van de LACONTO. Ook deze correlatie (gecorrigeerd voor attenuatie) bleek laag te zijn (.26). Dit geeft een aanwijzing dat een 'patiënt' als deelnemer aan het consult vanuit een ander referentiekader oordeelt over de communicatie tussen arts en patiënt, dan een observator-buitenstaander.

Tabel 5.

Simulatiepatiënt-oordelen per opleidingsniveau

(in percentage van het aantal beoordeelde consulten: 79 voor beginners en 80 voor gevorderden)

	oneens	neutraal	eens
<i>De wijze waarop de dokter het contact begon beviel mij goed</i>			
Beginners	9	24	67
Gevorderden	-	17	83
<i>Ik kreeg voldoende ruimte om mijn verhaal te vertellen</i>			
Beginners	4	9	87
Gevorderden	2	9	89
<i>De informatie en uitleg die de dokter mij gaf was voldoende duidelijk en begrijpelijk</i>			
Beginners	14	31	55
Gevorderden	9	12	79
<i>Ik had voldoende mogelijkheid mee te kunnen beslissen over wat er ging gebeuren</i>			
Beginners	9	37	54
Gevorderden	4	40	56
<i>Ik voelde mij in goede handen bij deze dokter</i>			
Beginners	15	28	57
Gevorderden	5	29	66
<i>Ik ben tevreden over dit consult</i>			
Beginners	15	29	56
Gevorderden	6	28	66

CONCLUSIES

Het is mogelijk gebleken om met medewerking van alle acht universitaire huisartsinstituten en de Stichting Deskundigheidsbevordering Huisartsen, in een periode van twee jaar, een eerste uitgeteste versie van een landelijk toetsingsinstrument op het gebied van consultvoering gereed te hebben. Een instrument dat gezien de intensiteit van de medewerking van de 27 inhoudsdeskundigen en de kwaliteit van het door hen geleverde commentaar een hoge mate van inhoudsvaliditeit bezit.

Wat betreft de betrouwbaarheid en het discriminerend vermogen (constructvaliditeit) heeft het landelijk toetsingsinstrument in de huidige fase van ontwikkeling zeker niet zijn optimum bereikt. De resultaten uit de experimentele toetsafname wijzen erop dat het huidige instrument, bestaande uit globale items plus inhoudelijke toelichting en een beoordelingsschaal, tot voldoende betrouwbare en valide toetsresultaten zal kunnen leiden wanneer aan een aantal voorwaarden is voldaan. De scoringslijst en toelichting zouden inhoudelijk moeten worden bijgesteld op basis van de kritische commentaren van de beoordelaars en de uitkomsten van de itemanalyses, met name ten aanzien van de beleidsfase. Alle zes schaalpunten van de beoordelingsschaal zouden nader moeten worden omschreven of verankerd, om tot consistentere scoring te kunnen komen. Het instrument zou als instructiemethode in het onderwijs moeten worden ingevoerd om zowel huisartsen-in-opleiding als onderwijsgevers (groepsbegeleiders en huisarts-opleiders) de gelegenheid te geven om ervaring op te doen met de methode.²⁷ Om het verschil tussen huisartsen en gedragswetenschappers als beoordelaars te minimaliseren zou de beoordelaarstraining van één dagdeel naar drie dagdelen moeten worden uitgebreid. Het aantal casus in de

toets tenslotte zou, volgens de generaliseerbaarheidsstudie, minstens twaalf moeten bedragen om tot betrouwbare uitspraken te leiden.

Het feit dat in het experiment de inhoudspecificiteit naar verhouding een geringer probleem vormt dan uit de literatuur over observatietoetsing van medische competentie naar voren komt, lijkt erop te wijzen dat het bij het landelijke instrument gaat om een casus-onafhankelijk meetinstrument. Deze bevinding doet vermoeden dat het bij een adequate interbeoordelaarsbetrouwbaarheid mogelijk moet zijn om de twee consultvoeringsvaardigheden 'gericht en systematisch werken' en 'arts-patiënt communicatie' te toetsen met behulp van relatief korte toetsen. Deze vooronderstelling zal nader moeten worden onderzocht.

Het patiëntenoordeel over de bejegening bleek significant positiever te zijn ten aanzien van de gevorderde huisartsen-in-opleiding dan voor de beginners.

Het feit dat het oordeel van de patiënt over de bejegening weinig verband vertoont met het oordeel van de observator over het communicatieve (bejegenings-)aspect van het consult, kan beschouwd worden als een empirische ondersteuning van het verschil in referentiekader van een patiënt als deelnemer aan het consult en een observator als beoordelende buitenstaander.

Uit een vervolgonderzoek in samenwerking met het NIVEL waarbij het landelijk consultvoeringsinstrument werd gebruikt om videoconsulten van ervaren huisartsen te scoren, bleek de LACONTO in staat om onderscheid te maken tussen praktijkconsulten van 'goede' en van 'onvoldoende' kwaliteit.²⁸ Gezien de ontwikkelingstijd die nodig is om te komen tot een adequaat toetsinstrument lijkt het instrument voldoende potentie te hebben om, na de relatief korte ontwikkelingsperiode van twee jaar, uit te groeien tot een volwaardig, landelijk toepasbaar toetsinstrument.

Dat het ontwikkelde instrument ook buiten de huisartsgeneeskunde van waarde kan zijn, blijkt uit het feit dat inmiddels op basis van de LACONTO een observatielijst voor 'systematisch logopedisch handelen' wordt ontwikkeld aan de Hogeschool Midden Nederland.²⁹

LITERATUUR

1. Foolen CHGM, Tan LHC. Eindrapportage project evaluatie consultvoering. Ontwikkeling en eerste versie van een landelijk instrument voor het toetsen van consultvoeringsvaardigheden van huisartsen(-in-opleiding). Utrecht: Samenwerkingsverband-IOH, april 1991; Rapport SV-IOH-17.
2. Tan LHC, Düsman H, Foolen CHGM, Pollemans MC. Vaardigheden aan het begin van de beroepsopleiding tot huisarts. In: Essed GGM, Metz JCM, Scherpbier AJJA, eds. Basisarts: bevoegd, maar waartoe bekwaam? Speciale uitgave Bulletin Medisch Onderwijs 1990; 9(4): 155-8.
3. Pollemans MC, Eekhof JAH, Kramer AWM, Van Leeuwen YD. Kennistoetsing bij aanstaande huisartsen. In: Scherpbier AJJA, Metz JCM, Van der Vleuten CPM, eds. Gezond Onderwijs. Speciale uitgave Bulletin Medisch Onderwijs 1991; 10(2): 76-80.
4. Stolk J, Pollemans MC. Instituutsonderwijs in de beroepsopleiding tot huisarts. Bulletin Medisch Onderwijs 1991; 10(3): 94-9.
5. Grol R, Mokkink H, Tielens V. PREVARA-SETH: scoringslijst voor evaluatie en onderlinge toetsing van het huisartsgeneeskundig handelen. Deel 1: algemene aspecten van consultvoering. Nijmegen: Nijmeegs Universitair Huisartsen Instituut, 1988.
6. Pieters HM. De Utrechtse Consult Evaluatie Methode. Vaardigheden in consultvoering van huisartsen-in-opleiding getoetst. Utrecht: Rijksuniversiteit Utrecht, 1991. Dissertatie.
7. Van Thiel J, Kraan HF, Van der Vleuten CPM. Reliability and feasibility of measuring medical interviewing skills: the revised Maastricht History-Taking and Advice Checklist. Medical Education 1991; 25: 224-9.
8. Springer MP, ed. Basistakenpakket van de huisarts. Bijlage Medisch Contact 1983; 38.
9. Grol R, Pollemans M, Verheij Th. De meerjarige beroepsopleiding tot huisarts. II: Structuurplan en onderwijsdoelen. Medisch Contact 1986; 41: 539-44.
10. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. Teaching and Learning in Medicine 1990; 2: 58-76.
11. Stillman P, Regan M, Swanson D. A diagnostic fourth year performance assessment. Arch Intern Med 1987; 147: 1981-5.
12. Van Luyk SJ, Van der Vleuten CPM. A comparison of checklists and rating scales in performance-based testing. In: Hart JR, Harden RM, eds. More developments in assessing clinical competence. Montreal: Can-Heal. In press.
13. Tan L. Rapportage Landelijke Experimentele Vaardigheidentoets (LEV). Utrecht: Samenwerkingsverband-IOH, 1988. Rapport SV-IOH-06b.
14. Dochy FJ, Van Luyk SJ, eds. Handboek vaardigheidsonderwijs. Lisse: Swets & Zeitlinger, 1987.
15. NHG-Standaard Diabetes Mellitus Type II. Huisarts en Wetenschap 1989; 32(1): 15-8.
16. NHG-Standaard Urineweginfecties. Huisarts en Wetenschap 1989; 32(11): 439-43.
17. NHG-Standaard Enkeldistorsie. Huisarts en Wetenschap 1989; 32(5): 182-5.
18. NHG-Standaard Schouderklachten. Huisarts en Wetenschap 1990; 33(5): 196-202.
19. Hofstra M, Zwaard A, Grol R, Mokkink H, Tielens V. Kwaliteitsbewaking in de huisartsgeneeskunde. Richtlijnen en procedures voor toetsing van het medisch handelen in de huisartspraktijk. Nijmegen: Werkgroep Onderzoek Kwaliteit Huisartsgeneeskunde Katholieke Universiteit Nijmegen-Rijksuniversiteit Limburg, 1991.
20. Schers HJ. De betrouwbaarheid en bruikbaarheid van een toetsinstrument voor medisch handelen in de beroepsopleiding tot huisarts. Nijmegen: Nijmeegs Universitair Huisartsen Instituut, maart 1991.
21. Tan L. Landelijke stations medisch-technische vaardigheden. Stand van zaken april 1989. Utrecht: Samenwerkingsverband-IOH, 1989.
22. Cronbach L, Gleser G, Nanda H, Rajaratnam N. The dependability of behavioral measurements: generalizability for scores and profiles. New York: John Wiley and sons, 1972.
23. Brennan RL. Elements of generalizability theory. Iowa: American College Testing Program, 1983.
24. Van der Vleuten CPM, Wijnen WHFW. Niets praktischer dan een goede theorie: generaliseerbaarheidstheorie als instrument voor betrouwbaarheidsstudies. Bulletin Medisch Onderwijs 1991; 10(1): 2-14.
25. Van der Vleuten CPM. Naar een rationeel systeem voor toetsing van studieprestaties in probleem-gestuurd onderwijs. Studies naar betrouwbaarheid en validiteit van toetsen voor praktische vaardigheden. Maastricht: Rijksuniversiteit Limburg, 1989. Dissertatie.
26. Ebel RL. De praktische validering van vaardigheidstoetsen. Paspoort Onderwijspraktijk 1986; 1(5): 3-15.
27. Everwijn SEM, Bomers GBJ, Knubben JA. Een competentiegericht managementcurriculum. De overbrugging van de kloof tussen kennisverwerving en het toepassen van kennis. Tijdschrift voor Hoger Onderwijs 1990; 8(2): 55-66.
28. Bensing JM, Tan LHC, De Boer SA, Mol SSL. National assessment of communication skills in everyday general practice. Paper presented at the workshop Doctor-patient communication and the quality of care. Utrecht: NIVEL, juni 1991.
29. Beyaert EMM, De Bruijn A. Observatielijst Systematisch Logopedisch Handelen. Utrecht: Faculteit Gezondheidszorg Opleidingen, Hogeschool Midden Nederland, 1991.