

Betrouwbaarheid van de beoordeling van verpleeghuisartsen-in-opleiding aan de Vrije Universiteit Amsterdam

J.F. Hoek, M.W. Ribbe, C.P.M. van der Vleuten, W.L.J.M. Devillé

Samenvatting

Om de betrouwbaarheid van het voortgangs-evaluatieformulier als instrument voor de beoordeling van verpleeghuisartsen-in-opleiding (VAIO's) aan de Vrije Universiteit Amsterdam te bepalen hebben verschillende beoordelaars met dit formulier een groep VAIO's beoordeeld. Negen maanden na het begin van hun opleiding zijn 34 VAIO's beoordeeld door hun verpleeghuisartsopleider (VAO) (n=33) en een instituutsbegeleider (n=3). De VAO herhaalde de beoordeling tien tot twintig dagen later. In achttien verpleeghuizen beoordeelde ook een tweede verpleeghuisarts de VAIO. De verschillen tussen: 1) de twee beoordelingen van de VAO's (intrabeoordelaarsbetrouwbaarheid), 2) de beoordeling van de VAO en de tweede verpleeghuisarts (interbeoordelaarsbetrouwbaarheid) en 3) de beoordeling van de VAO en de instituutsopleider zijn geanalyseerd. De intrabeoordelaarsbetrouwbaarheid was goed. Tussen de oordelen van de VAO en een tweede verpleeghuisarts werd onvoldoende overeenstemming gevonden. De instituutsbegeleiders scoorden lager, maar wel in hetzelfde patroon als de VAO. Het bleek dat de taakuitoefening van VAIO's te beoordelen is. Training zal mogelijk de oordelen van VAO en de instituutsopleider dichter bij elkaar brengen.

Inleiding

Nog maar enkele tientallen jaren geleden werd de duur van de meeste medische beroepsopleidingen bepaald door het talent van de

arts-assistent en de beoordelingsmaatstaf van de opleider. Welke taken de arts-assistent diende te beheersen zat voornamelijk in het hoofd van de opleider en het maakte als gevolg hiervan een groot verschil waar iemand was opgeleid. Uitgeschreven leerplannen en in het verlengde hiervan eindtermen, beoordelingsmaatstaven en -methoden waren in die tijd dan ook meer uitzondering dan regel.

Anno 1998 is deze situatie volledig gewijzigd. Met het gereed komen van het Raamplan 1994 Artsopleiding is er meer duidelijkheid ontstaan over de begintermen van een medische vervolgopleiding.¹ Voor de meeste medische beroepen is het takenpakket door de beroepsverenigingen beschreven.² De opleidingen hebben hieraan de eindtermen voor het betreffende vakgebied ontleend en een curriculum ontworpen. De voortgang en geschiktheid van arts-assistenten worden in de meeste opleidingen vastgesteld door regelmatige beoordelingen.

Waarom beoordelen?

De vrije uitwisseling van specialisten tussen landen maakt het nodig dat de certificaten van verschillende landen verwijzen naar eenzelfde kwaliteit van professionele standaarden.³ Opleidingsinstellingen dienen maatregelen te nemen om zich ervan te verzekeren dat de arts-assistenten over de kennis en vaardigheden beschikken om patiënten optimaal te behandelen. Gegeven de omvang van de investering in de medische vervolgopleiding door arts-assistenten, opleiders en de overheid zijn er ook economische redenen voor een goed beoorde-

lingssysteem, omdat een van de doelen het verbeteren van het onderwijs is.

Hoe beoordelen?

In de literatuur worden verschillende wijzen van beoordelen beschreven, zoals (multiple-choice-)kennistoetsen, kennis-over-vaardigheidentoetsen, vaardigheidstoetsen (OSCE), gebruik van simulatiepatiënten, observatie in de praktijkwerksituatie, et cetera.³⁻⁸ Geen van deze beoordelingen is voor 100% valide en betrouwbaar. Wel is aangetoond dat er een verband is tussen de uitkomsten van al deze vormen van beoordelen. Elk instrument verschaft een beeld van een onderdeel van de competentiepiramide van op elkaar steunende deelaspecten van klinische competentie.⁹

Bij de keuze van een beoordelingsinstrument zijn ook het tijdstip en de status van een beoordeling van belang. Zo kan met selectie vooraf wel onder andere het aantal specialisten gereguleerd worden, maar het is de vraag of een inschatting gemaakt kan worden of een opleiding met goed gevolg voltooid kan worden. Voor de keuze van het instrument maakt het bijvoorbeeld ook uit of het de bedoeling is om de voortgang van de klinische competentie te meten of het eindresultaat. Samenvattend, moeten bij de keuze voor een instrument vele afwegingen gemaakt worden. Het is belangrijk om deze afwegingen bewust te maken. Een hulpmiddel hiervoor is de 'utiliteitsformule'.¹⁰

Het beoordelingssysteem van de verpleeghuisartsopleiding van de Vrije Universiteit

Het beoordelingssysteem van de verpleeghuisartsopleiding van de Vrije Universiteit (VU) Amsterdam is gebaseerd op een meting van de ontwikkeling van de klinische competentie op meerdere momenten gedurende de vervolgopleiding. De klinische competentie is geoperationaliseerd als de mate waarin de standaard-

taken, zoals beschreven door de Nederlandse Vereniging van Verpleeghuisartsen (NVVA), in de praktijk worden verricht.¹¹

Gedurende de tweejarige vervolgopleiding tot verpleeghuisarts vindt op vier momenten een beoordeling van de verpleeghuisarts-in-opleiding (VAIO) plaats door de praktijkopleider (VAO) en de instituutsbegeleiders (verpleeghuisartsbegeleider en gedragswetenschappelijk begeleider), te weten na 3, 9, 15 en 22 maanden. Als beoordelingsinstrument is hiervoor in 1996 een voortgangsevaluatieformulier ontwikkeld, waarin 50 taken worden opgesomd. Deze taken zijn vrijwel letterlijk afgeleid van de 33 taken van een verpleeghuisarts, zoals die door de NVVA zijn geformuleerd, en die tevens de eindtermen van de opleiding zijn. Een aantal voor beoordeling te omvangrijke taken is opgesplitst in deeltaken. Om een betere aansluiting te verkrijgen bij de structuur van het theoretisch deel van het curriculum, zijn de taken vervolgens herschikt volgens de indeling van het onderwijs in veertien blokken (zoals Revalidatie, Psychogeriatric, Palliatieve zorg, Methodische werkwijze, Management, Wetenschappelijk onderzoek).¹² In figuur 1 is een deel van het voortgangsevaluatieformulier weergegeven.

Achter ieder item moet op een ordinale vijf-puntsschaal een rangscore ingevuld worden, die staat voor de mate waarin de VAIO de betreffende taak beheerst: 1 = <40%, 2 = 40-50%, 3 = 50-60%, 4 = 60-80% en 5 = >80%. Score 4 (60-80%) staat voor de mate waarin een geregistreerd verpleeghuisarts de taak geacht wordt te beheersen en is daarmee de norm die aan het eind van de opleiding volgens de beroepsgroep gehaald moet worden. Het oordeel van de VAO is gebaseerd op observatie van de VAIO in de praktijkwerksituatie. Indien de opleider een taak (nog) niet kan beoordelen, wordt een score '?' (weet ik niet/niet te beoordelen) gegeven. De scores geven op ieder meetmoment de mate aan waarin een taak beheerst wordt. Bij een normaal opleidingsverloop en een goed functionerend beoordelings-

BEH
3=50
Blok
4.1
4.2
Blok
5.1
5.2
5.3
Blok
6.1
6.2
6.3
De g
(eind

Figuur 1
de v

syst
volg
zier
I
teer
doo
gee
inst
dee
leid
VA
bes
tijd
I
van
me
VA
ant
zitt

BEHEERST DEZE TAAK VOOR: 1=<40%, 2=40-50%, 3=50-60%, 4=60-80%=EINDTERM, 5=>80%, ?=weet ik niet.	1	2	3	4	5	?
<i>Blok 4: Psychogeriatric</i>						
4.1 Het verrichten van een probleemgericht psychogeriatrisch onderzoek en het vaststellen van prioriteiten voor zorg, behandeling en begeleiding.						
4.2 Diagnostiek van angst, onrust en inadequaat gedrag bij een psychogeriatrische patiënt en het formuleren van een zorg/behandelplan hiervoor.						
<i>Blok 5: Speciële ziektebeelden en chronische aandoeningen</i>						
5.1 Behandelen van chronische ziekten, speciaal in de latere fases van het ziektebeeld en met aandacht voor de vaak specifieke presentatie bij ouderen.						
5.2 Afwegen van zinvolheid bij het instellen of continueren van therapie.						
5.3 Het actief opsporen van mogelijke interacties van verschillende therapieën in geval van multipathologie.						
<i>Blok 6: Palliatieve zorg</i>						
6.1 Bepalen in welke gevallen curatieve behandeling medisch niet zinvol is en dit bespreken met patiënt, familie en behandelteam.						
6.2 Behandelen van pijn en andere symptomen.						
6.3 Actief opsporen van en begeleiden bij psychologische en sociale problemen en zingevingsaspecten.						
De gearceerde kolom is het niveau van taakbeheersing dat aan het eind van de opleiding gehaald moet worden (eindterm).						

Figuur 1. Gedeelte van het voortgangsevaluatieformulier voor de beoordeling van een verpleeghuisarts-in-opleiding door de verpleeghuisartsopleider.

systeem moeten de scores op de vier achtereenvolgende meetmomenten een stijgende lijn te zien geven.

Het hierboven beschreven beoordelingssysteem gaat uit van een beoordeling van de VAIO door de VAO en de instituutsbegeleiders. Eerst geeft de VAO een oordeel en zendt dit naar de instituutsbegeleiders, die vervolgens hun oordeel geven. Het oordeel van de instituutsbegeleiders is gebaseerd op het oordeel van de VAO, aangevuld met de resultaten van beschikbare toetsen en hun eigen observaties tijdens het cursorisch gedeelte.

In dit onderzoek staat de betrouwbaarheid van het voortgangsevaluatieformulier als meetinstrument voor de beoordeling van VAIO's centraal. Daarnaast wordt getracht antwoord te geven op de vraag of er verschillen zitten in het oordeel van de VAO en de insti-

tuutsbegeleiders. Immers zij moeten tot een gezamenlijk oordeel komen. De VAO heeft de beste mogelijkheden om de VAIO te beoordelen, aangezien hij direct zicht heeft op de taakuitvoering. De instituutsbegeleiders beoordelen de taken in een 'laboratoriumsituatie', die mogelijk niet overeenkomt met de praktijk. De vraagstellingen zijn:

1. Hoe is de intrabeoordelaarsbetrouwbaarheid van de beoordeling van de VAIO door de opleider (VAO) met gebruikmaking van het voortgangsevaluatieformulier?
2. Is er een verschil in beoordeling van de VAIO door de VAO en door een collega-verpleeghuisarts uit het opleidingshuis (interbeoordelaarsbetrouwbaarheid)?
3. Is er een verschil in beoordeling van de VAIO door de VAO en door de verpleeghuisartsbegeleider op het instituut?

Methode

Het onderzoek naar de betrouwbaarheid van het voortgangsevaluatieformulier vond plaats in juni 1997; dit is na negen maanden opleiding van het cohort 1996-1998, bestaande uit 37 VAIO's. Dit moment is gekozen omdat de opleider na negen maanden geacht wordt voldoende zicht te hebben op het functioneren van de VAIO. Bovendien wisselt een aantal VAIO's na een jaar van opleidingsplaats en dus van opleider, zodat een later tijdstip het nadeel heeft dat er verschillen ontstaan in de mate waarin de opleider zijn VAIO kent. Van de 37 VAIO's hadden er in juni 1997 drie voortijdig de opleiding beëindigd (één wegens een ernstige ziekte, één wegens aanvaarden van een opleidingsplaats elders en één op grond van een onvoldoende beoordeling). Voor het onderzoek resteerden dus 34 VAIO's.

De intrabeoordelaarsbetrouwbaarheid (onderzoeksvraag 1) werd bepaald met de test-hertestmethode. Van de 34 opleiders hebben 33 het voortgangsevaluatieformulier tweemaal ingevuld met een tussentijd van ongeveer twee weken (spreiding 10-20 dagen): eenmaal in het kader van de standaard negenmaandsbeoordeling en een tweede keer tijdens de in die maand geplande reguliere terugkomdag voor opleiders op de universiteit. Voor de periode van twee weken is gekozen omdat is aangenomen dat deze enerzijds zo kort is dat het niveau van de VAIO niet wezenlijk is veranderd en anderzijds voldoende lang is om te voorkomen dat de VAO nog precies weet wat hij de eerste maal heeft ingevuld. De opleiders ontvingen vooraf een training van een middag in het invullen van het voortgangsevaluatieformulier.

De interbeoordelaarsbetrouwbaarheid (onderzoeksvraag 2) werd bepaald door aan een tweede verpleeghuisarts (tweede VPHA) uit het opleidingsverpleeghuis te vragen eveneens een voortgangsevaluatieformulier in te vullen. De tweede VPHA diende geregistreerd te zijn, bereid te zijn aan het onderzoek deel te nemen en voldoende observatiemogelijkheden te

hebben om een oordeel over de VAIO te kunnen geven (bijvoorbeeld door een afdeling met de VAIO te delen of er regelmatig waar te nemen). Van de 34 verpleeghuizen beschikten er achttien over een tweede VPHA die aan deze criteria voldeed. De deelnemende tweede VPHA's ontvingen alleen een schriftelijke invulinstructie en geen training. Het oordeel van de tweede verpleeghuisarts werd vervolgens vergeleken met het eerste oordeel van de opleider (VAO). Voor de interbeoordelaarsbetrouwbaarheid waren dus achttien paren beoordelingen beschikbaar.

Op eenzelfde wijze is nagegaan of er een verschil is in de beoordeling van de VAIO door de opleider en de instituutsbegeleider (onderzoeksvraag 3). De instituutsbegeleiders (n=3) hebben hiertoe, eveneens in juni 1997, een voortgangsevaluatieformulier ingevuld van de VAIO's uit hun groep (respectievelijk twaalf, elf en elf VAIO's).

De studieopzet waarbij elke beoordelaar (buiten de instituutsbegeleider) slechts één VAIO beoordeelt, beperkt de mogelijkheden van de analyse: de variatiebron beoordelaar kan niet onderscheiden worden van de variatiebron VAIO. Hierdoor beperkt de analyse zich tot het toetsen van het verschil tussen de mediane scores (en de range) met de Wilcoxon Matched-Pairs Signed Test. Er is gecorrigeerd voor meervoudig toetsen volgens Bonferroni. Ook is het percentage overeenkomst met het 95% betrouwbaarheidsinterval (BI) berekend voor alle afzonderlijke items. Wegens de gepaarde items binnen een blok is voor blokken met meerdere items geen percentage overeenkomst berekend.

Resultaten

Bij 48 van de 50 taken werden geen significante verschillen ($p < 0.05$) gevonden tussen de eerste en tweede beoordeling door de opleider (tabel 1). De statistische verschillen op de twee resterende taken verdwijnen na correctie voor meervoudig toetsen. Als er verschil was tussen de

Tabel
ter) (

Item

—
Blok
Blok
Blok
Blok
en cl
Blok
Blok
Blok
Blok
Blok
Blok
Blok
Blok

* p=
** G
Het p
is in

eers
32 :
hog
ove
55-:

de t
take
sco
ove
24-:

beg
bij
ove
26-
war
dan
taal
zoe
opl

Tabel 1. Beoordelingsscores van de opleiders bij de eerste beoordeling (VAO1) en de tweede beoordeling (10-20 dagen later) (VAO2).

Item	VAO1 mediaan (range)	VAO2 mediaan (range)	Percentage overeenkomst** (95% BI)
Blok 1: Acute pathologie	4.0 (2-5)	4.0 (2-5)	79 (65-93)
Blok 2: Geriatrische syndromen	4.0 (2-5)	4.0 (2-4)	79 (65-93)
Blok 3: Revalidatie	4.0 (2-4)	3.5 (2-4.5)	
Blok 4: Psychogeriatric	3.7 (3-4.3)	3.3 (2.7-4.3)*	
Blok 5: Specifieke ziektebeelden en chronische aandoeningen	3.5 (2-5)	3.5 (2.5-5)	
Blok 6: Palliatieve zorg	4.0 (3-5)	4.0 (2-5)	78 (64-92)
Blok 7: Gericht werken	3.6 (2-6-4.3)	3.7 (2.6-4.6)	
Blok 8: Gesprekstechniek	4.0 (2-5)	4.0 (2-5)	72 (56-87)
Blok 9: Samenwerken	4.0 (2-5)	3.0 (2-5)	75 (60-90)
Blok 10: Ethiek/ Recht	3.5 (1-5)	3.0 (2-5)	63 (45-81)
Blok 11: Onderwijs	3.8 (2-5)	4.0 (2.3-5)	
Blok 12: Management	3.3 (1.7-4.3)	3.0 (2-4)	
Blok 13: Extramuraal	2.5 (1-4)	2.5 (1.5-4)	
Blok 14: Wetensch. onderzoek	2.0 (1-5)	3.0 (1-5)*	70 (50-90)

* $p < 0.05$.

** Gemiddelde overeenkomst: 74% (95% BI 71-76); range 55-86.

Het percentage overeenkomst is per item berekend. Aangezien de blokken 3, 4, 5, 7, 11, 12 en 13 uit meerdere items bestaan, is in de tabel geen percentage overeenkomst vermeld.

eerste en tweede score van de VAO, was dat in 32 gevallen een lagere en in 11 gevallen een hogere score. Er is een gemiddelde totaalovereenkomst van 74% (95% BI 71-76); range 55-86.

Vergelijking van de scores van de VAO en de tweede VPHA liet op geen van de vijftig taken een significant verschil zien in mediane scores (tabel 2). De gemiddelde totaalovereenkomst is 49% (95% BI 46-53); range 24-81.

Vergelijking met de score van de instituutsbegeleider leverde significante verschillen op bij 33 van de 50 scores bij een gemiddelde overeenkomst van 43% (95% BI: 40-45); range 26-63. De mediane scores van de opleiders waren op 37 van de 50 taken een vol punt hoger dan die van de instituutsbegeleiders, en op één taak (verrichten van wetenschappelijk onderzoek) een punt lager. De mediaanscore van de opleiders was op 39 taken 4, op twee taken 3.5,

op acht taken 3 (voornamelijk uit de blokken Management en Extramuraal verpleeghuiszorg) en op één taak 2 (Wetenschappelijk onderzoek).

Score 1 (beheerst de taak voor <40%) werd door alle beoordelaars weinig gegeven en dan vooral op de taken uit de blokken 3, 4 en 11-14. Score '?' (weet ik niet/niet te beoordelen) werd vaker gegeven door de tweede VPHA (37/50 items) dan door de opleiders (24/50 items) en de instituutsbegeleiders (19/50 items). De meeste '?'-scores waren te vinden in alle items van de blokken 3, 4, 7 (7.1.3.) en 10-14.

Beschouwing

De intrabeoordelaarsbetrouwbaarheid van het voortgangsevaluatieformulier is goed. Er werden geen verschillen gevonden in de mediane scores van de eerste en tweede beoordeling door de opleider. Gezien het tijdstip zijn

Tabel 2. Beoordelingsscores van de opleiders (VAO) en de tweede verpleeghuisartsen (tweede VPHA).

Item	VAO1 mediaan (range)	Tweede VPHA mediaan (range)	Percentage overeenkomst * (95% BI)
Blok 1: Acute pathologie	4.0 (2-5)	4.0 (2-5)	37 (15-59)
Blok 2: Geriatrie syndromen	4.0 (2-5)	3.0 (2-5)	37 (15-59)
Blok 3: Reactivering	4.0 (2-4)	3.5 (1.5-4)	
Blok 4: Psychogeriatric	3.7 (3-4.3)	3.4 (2.7-4.3)	
Blok 5: Specieële ziektebeelden en chronische aandoeningen	3.5 (2-5)	3.5 (2.5-4)	
Blok 6: Terminale zorg	4.0 (3-5)	3.0 (2-5)	44 (21-67)
Blok 7: Gericht werken	3.6 (2.6-4.3)	3.6 (2.7-4.3)	
Blok 8: Gesprekstechniek	4.0 (2-5)	3.0 (2-5)	29 (8-51)
Blok 9: Samenwerken	4.0 (2-5)	3.0 (2-4)	32 (11-52)
Blok 10: Ethiek/ Recht	3.5 (1-5)	4.0 (2-5)	24 (3-44)
Blok 11: Onderwijs	3.8 (2-5)	3.3 (2-4.7)	
Blok 12: Management	3.3 (1.7-4.3)	3.0 (2-4)	
Blok 13: Extramuraal	2.5 (1-4)	2.5 (1-3)	
Blok 14: Wetensch. onderzoek	2.0 (1-5)	3.0 (1-4)	63 (29-96)

*Gemiddelde overeenkomst: 49% (95% BI 46-53); range 24-81.

Het percentage overeenkomst is per item berekend. Aangezien de blokken 3, 4, 5, 7, 11, 12 en 13 uit meerdere items bestaan, is in de tabel geen percentage overeenkomst vermeld.

er (te?) hoge scores: 78% van de taken krijgen na negen maanden opleiding een mediaanscore van 4, het niveau dat pas aan het eind van de opleiding gehaald hoeft te worden. De lager scorende taken zijn vrijwel allemaal terug te vinden in de blokken Management, Extramuraal verpleeghuiszorg en Wetenschappelijk onderzoek. Samen met Revalidatie zijn dit de onderdelen van de opleiding, die voornamelijk in het tweede jaar aan bod komen, hetgeen de lagere scores kan verklaren. We hebben een aantal mogelijke oorzaken voor de hoge scores overwogen.

a. De beoordelaars waren onvoldoende op de hoogte van het voortgangskarakter van de beoordeling. Bij het weergeven van het oordeel is de beoordeling onvoldoende afgezet tegen het eindniveau van de opleiding. De score is dus geen rapportcijfer, zoals dat op scholen wordt gebruikt en waarbij juist het tijdstip in de opleiding bepalend is (een 6 voor Frans in de eerste klas

houdt een lager niveau in dan een 6 voor Frans in de vijfde klas).

- b. De opleider scoort hoog, omdat een lage score door hem gezien wordt als een falen als opleider of de relatie tussen hem en de VAIO onder druk zet.³
- c. Beoordelen is meten, wegen en beslissen.¹⁰ Het voortgangsevaluatieformulier is het meetinstrument waarin aangegeven wordt in welke mate de taken beheerst worden. Pas hierna wordt gewogen welke omstandigheden van invloed zijn geweest op het gemeten niveau (de VAIO is ziek geweest, er was sprake van een hoge werkdruk door ziekte van collega's, de VAIO is nog maar kort in opleiding, enzovoorts). Tot slot volgt een beslissing over de interpretatie van de gevonden meetgegevens. De mogelijkheid bestaat ook dat de opleider verzachtende omstandigheden tijdens het invullen van het formulier heeft laten meewegen en daarom dus te hoog scoort.

De h
hoge
midd
voor:
versc
ders
dan e
aan c
korte
meer
hoog
form
mind
betro
een b
sing:
situat
aties

Tu
oplei
waar
maar
erg l:
oplei
signi
die z
lager
een v
ming
VAIO
de in
tert.

Liter.

1. Me
Br.
art
me
ver
2. He
Eis
tac

De hierboven genoemde oorzaken van een te hoge score zijn mogelijk weg te nemen door middel van een betere training van de opleiders vooraf. Een aanwijzing hiervoor is dat verschillen tussen de twee scores van de opleiders vaker een lagere tweede score betroffen dan een hogere, omdat er direct voorafgaande aan de tweede invulling abusievelijk nog een korte extra instructie is gegeven, waarbij onder meer verteld is dat er bij de eerste invulling erg hoog gescoord was. Mogelijk zal een andere formulering van de scoringsmogelijkheden, die minder een waardeoordeel inhoudt, tot betrouwbaarder scores leiden (bijvoorbeeld een beschrijving van de mate van taakbeheersing: 'beheerst deze taak alleen in eenvoudige situaties', 'beheerst deze taak in complexe situaties en/of onder tijdsdruk').

Tussen de mediane oordelen van de opleider en de tweede VPHA werden weliswaar geen significante verschillen gevonden, maar het percentage overeenkomst is met 49% erg laag. De vergelijking van de oordelen van opleider en instituutsbegeleider toont een significant verschil op 33 van de 50 items, in die zin dat de instituutsbegeleider significant lager scoort. Het betreft in de meeste gevallen een vast verschil. Er is dus wel overeenstemming over de sterke en zwakke kanten van de VAIO. Nagegaan zal worden of door training de interbeoordelaarsbetrouwbaarheid verbeterd.

Literatuur

1. Metz JCM, Pels Rijcken-Van Erp Taalman Kip EH, Brand-Valkenburg BMW van den. Raamplan 1994 artsopleiding. Eindtermen van de artsopleiding. Nijmegen: Universitair Publikatiebureau Katholieke Universiteit Nijmegen, 1994.
2. Hoek JF, Hoog JC de. Medische vervolgoopleidingen. Eisen, die gesteld worden aan opleiders. *Medisch Contact* 1997;52:1067-9.
3. Karle H, Nystrup J. Comprehensive evaluation of specialist training: an alternative to board examinations in Europe? *Med Educ* 1995;29:308-16.
4. Leeuwen YD van. Growth in knowledge of trainees in general practice: figures on facts [proefschrift]. Maastricht: Universitaire Pers Maastricht, 1995.
5. Luijk SJ van. Al doende leert men. Enkele studies naar aspecten van betrouwbaarheid en validiteit over de toetsing van vaardigheden [proefschrift]. Maastricht: Universitaire Pers Maastricht, 1994.
6. Newble D, Jolly B, Wakeford R. The certification and recertification of doctors. Issues in the assessment of clinical competence. Cambridge: Cambridge University Press, 1994.
7. Pollemans MC. Kennistoetsing bij huisartsen [proefschrift]. Maastricht: Universitaire Pers Maastricht, 1994.
8. Sloan DA, Donnelly MB, Schwartz RW, Strodel WE. The objective structured clinical examination. The new gold standard for postgraduate clinical performance. *Ann Surg* 1995;222(6):735-42.
9. Miller GE. The assesment of clinical skills/competence/performance. *Acad Med* 1990;65(9 Suppl): S63-7.
10. Vleuten CPM van der. Toetsing van medische competentie. In: Metz JCM, Scherpbier AJJA, Vleuten CPM van der, redactie. *Medisch onderwijs in de praktijk*. Assen: Van Gorcum, 1995:152-64.
11. Nederlandse Vereniging van Verpleeghuisartsen. Nota functieomschrijving en takenpakket van de verpleeghuisarts. Utrecht, 1992.
12. Hoek JF, Geertsema H. Leerplan vervolgoopleiding tot verpleeghuisarts, versie 3.2. Amsterdam: Verpleeghuisartsopleiding/VU, 1997:16-21.
13. Hoek JF, Geertsema H. Medische vervolgoopleidingen. Didactische vaardigheden van opleiders. *Medisch Contact* 1997;52:1064-6.

DE AUTEURS

*J.F. Hoek is verpleeghuisarts en hoofd Verpleeghuisartsopleiding aan de Vrije Universiteit Amsterdam.
Prof. dr. M.W. Ribbe is hoogleraar verpleeghuisgeneeskunde aan de Vrije Universiteit Amsterdam.
Prof. dr. C.P.M. van der Vleuten is hoogleraar aan de Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch van de Universiteit Maastricht.
W.L.J.M. Devillé is arts-epidemioloog aan de Vakgroep Epidemiologie en Biostatistiek van de Vrije Universiteit Amsterdam.*

Correspondentieadres:

Frank Hoek, Verpleeghuisartsopleiding Vrije Universiteit, Van der Boechorststraat 7, 1081 BT Amsterdam.