

De toetsing van medische competentie

Cees P. M. van der Vleuten¹ · Jeroen J. G. van Merriënboer¹

Published online: 15 November 2016
© The Author(s) 2016. This article is available at SpringerLink with Open Access.

Samenvatting Toetsing van medische competentie kent een rijke historie van ontwikkeling en onderzoek. In dit artikel wordt eerst een kort overzicht gegeven van een aantal door onderzoek verworven inzichten en hun praktische implicaties. Vervolgens wordt er ingegaan op het moderne competentiegerichte opleiden en de implicaties voor toetsing. Ook de koppeling van competenties aan beter te begrijpen kritische beroepstaken, de *Entrustable Professional Activities* (EPA's), komt aan bod. Tot slot wordt ingegaan op een meer systematische aanpak van toetsing, met aandacht voor meer feedback om het leren te bevorderen en het op basis daarvan nemen van zwaarwichtige beslissingen. Mits de toetsing zorgvuldig wordt geïmplementeerd en wordt geaccepteerd in de klinische praktijk, kan toetsing bijdragen aan het verkrijgen van betere dokters.

Trefwoorden toetsing · competentiegericht opleiden · programmatische toetsing

The assessment of medical competence

Abstract The assessment of medical competence has a rich history of research and development. This paper provides a short overview of acquired insights and their practical implications. Attention is being paid to competency-based education and its consequences for assessment. The coupling of competencies to more easily understandable clinical tasks, the *Entrustable Professional Activities* (EPA's),

is included in this discussion. Finally, an explanation is given of a more systematic approach to assessment with more attention on feedback for learning and using this rich aggregated information for robust high-stake decision-making. Provided a proper implementation and with an appropriate buy-in from clinical professionals, assessment may contribute to training better doctors.

Keywords assessment · competency-based education · programmatic assessment

Introductie

Toetsing in de geneeskunde is een omvangrijk onderwerp waar heel veel over geschreven is [1]. Er is dan ook veel ontwikkeld en onderzocht, wat heeft geleid tot veel kennis over toetsing. We willen in dit artikel in vogelvlucht hierop ingaan. We beginnen met het uitleggen van enkele bouwstenen van toetsing, vervolgens bespreken we het moderne opleiden door middel van competentiegericht onderwijs en de gevolgen hiervan voor toetsing. We eindigen met een beschrijving van een moderne integrale aanpak van toetsing.

Bouwstenen, methoden en inzichten

Onder toetsing verstaan we alle activiteiten die leiden tot een uitspraak over een vaardigheid van een individu of van een team van individuen. Toetsing kent twee doelen. Het bekendste doel is de summatieve functie: het nemen van een beslissing of de beoordeelde aan de maat is. Deze functie van toetsing legt tegelijkertijd aan de maatschappij verantwoording af van het feit dat een opleiding ergens toe

✉ prof. dr. Cees P. M. van der Vleuten
c.vandervleuten@maastrichtuniversity.nl

¹ Vakgroep Onderwijsontwikkeling en Onderwijsresearch, Universiteit Maastricht, Maastricht, Nederland

geleid heeft. Het laat ook zien dat de opleiding effectief is geweest. Daartegenover staat de formatieve functie van toetsing. Bij deze functie ligt het accent op de vormende werking van toetsing: de toets als spiegel van het leren en het gebruik van toetsinformatie om het leren te bevorderen. Toetsing kan een krachtig middel zijn om leren te bevorderen. Onderzoek toont systematisch aan dat onderwijs met toetsing veel effectiever is dan onderwijs zonder toetsing [2].

Het doel van een toets *Summatieve toetsing*. De maat nemen: bijvoorbeeld een landelijk theoretisch examen met einduitslag in maat en getal, vergelijkbaar met mede-aios.

Formatieve toetsing. Het accent ligt op vormende werking: bijvoorbeeld feedback geven tijdens of na een operationele ingreep, eventueel met behulp van *objective structured assessment of technical skill* (OSATS).

Er bestaan heel veel toetsmethoden en het is ondoenlijk om deze in kort bestek te behandelen. Een handig overzicht wordt gegeven door Epstein [3].

Het is gemakkelijk om toetsmethoden te classificeren. Dat kan gedaan worden door een eenvoudig competentie-model te beschouwen, de piramide van Miller (fig. 1 en [4]).

Aan de basis van het model van Miller ligt feitenkennis (*kennen*). Als met de feitenkennis gewerkt kan worden tijdens klinisch redeneren en probleemoplossing, wordt gesproken van ‘*kunnen*’. Het derde niveau is wanneer deze vaardigheden kunnen worden gedemonstreerd in een gesimuleerde situatie (*laten zien*). Dit kan een geheel consult of gehele handeling zijn, of een deelvaardigheid, zoals omgaan met laser of elektrochirurgie. Op het vierde niveau vindt een beoordeling plaats in de echte praktijk (*doen*). ‘Kennen’ en ‘kunnen’ zijn cognitieve vaardigheden, ‘laten zien’ en ‘doen’ zijn gedragsvaardigheden. De cognitieve vaardigheden worden typisch getoetst met schriftelijke toetsen (meerkeuze, open vragen, computertoetsen, mondeling, etc.).



Voor de toetsing van klinisch redeneren wordt vaak gebruikgemaakt van (klinische) scenario's die aan de toepas-

sing van kennis appelleren op het specifieke scenario. Soms wordt zo'n scenario gecombineerd met een open-boekstrategie. Boeken mogen worden geraadpleegd tijdens de toetsing. Kenmerkend voor de beoordeling van gedragsvaardigheden is de observatie door een professionele beoordelaar, vaak een supervisor. Het verschil tussen de derde en de vierde laag is de setting. In de derde laag is deze setting gesimuleerd en in de vierde laag is de setting de praktijk-situatie. De eerste drie lagen betreffen gestandaardiseerde toetsvormen. Hier worden condities geschapen die voor elke beoordeelde gelijk zijn. De vierde laag omvat ongestandaardiseerde toetsvormen. De condities van beoordeling zijn afhankelijk van verschillende patiënten, fysieke context en beoordelaar en verschilt van situatie tot situatie.

Millers piramide is ook bekritiseerd. Het is bijvoorbeeld geen didactisch model dat voor zou schrijven hoe we zaken in een bepaalde volgorde moeten leren. De vorm van de piramide is ook bekritiseerd. Voor een vervolgonderwijs tot medisch specialist zou de piramide eigenlijk op zijn kop moeten staan, omdat in die situatie, voor dat leerproces, het formatief toetsen in de praktijk centraal staat.

Toetsmethoden onderzocht

Gesteld kan worden dat in de afgelopen decennia in de geneeskunde de piramide ‘beklommen’ werd en dat voor alle lagen toetsmethoden ontwikkeld en gevalideerd zijn. Tijdens deze beklimming is veel onderzoek gedaan. Tab. 1 geeft een overzicht van deze onderzoeksbevindingen en hun praktische implicaties, en destilleert daarmee een aantal consistenties in deze bevindingen. Voor een wetenschappelijke verantwoording van deze onderzoeksconsistenties wordt in het kader van dit artikel verwezen naar Van der Vleuten et al. [5].

Het onderzoek levert verrassende en soms contra-intuïtieve inzichten op. Objectiviteit en subjectiviteit zijn kernbegrippen rondom toetsing. Zo hebben we de neiging om subjectieve invloeden uit te sluiten door beoordelingen te standaardiseren en te objectiveren (bijv. door gedetailleerde checklijsten te gebruiken). Uit het onderzoek komt echter duidelijk naar voren dat betrouwbaarheid (de reproduceerbaarheid van de scores) minder een kwestie is van standaardisering of van objectivering, maar veeleer van steekproeftrekking. Als een uitspraak over de competentie van een individu wordt gebaseerd op één enkele situatie, of één enkele beoordelaar, zal een uitspraak nooit betrouwbaar zijn, ongeacht hoe objectief zo'n beoordeling ook plaatsvindt. Hebben we beoordelingen van een individu over meerdere situaties, momenten of beoordelaars, dan wordt de uitspraak betrouwbaar. Subjectieve informatie kan dus heel goed gebruikt worden, mits we voldoende oordelen verzamelen. Voor het beoordelen van de top van de piramide is dit goed

Figuur 1 Een eenvoudig competentiemodel voor het classificeren van toetsingsmethoden (naar Miller) [4].



nieuws. Hier berusten oordelen op subjectieve beoordelingen van professionele beoordelaars.

Genoeg is genoeg

Verzamelen we genoeg informatie in verschillende situaties bij verschillende beoordelaars, dan wordt het eindoordeel robuust. Een magisch getal van een achttal beoordelingen komt uit veel onderzoek naar voren als voldoende om tot een betrouwbaar oordeel te komen, mits deze beoordelingen betrekking hebben op meerdere situaties en beoordelaars [6]. Een korte klinische beoordeling is een directe observatie van een klinisch moment of consult, waarna een beoordeling op een formulier wordt aangegeven [7]. Na het verzamelen van acht van deze beoordelingen wordt een tamelijk betrouwbaar beeld verkregen van de competentie van een beoordeelde. 360 graden feedback of *multi-source feedback* is een beoordeling van een persoon over een langere periode van functioneren, waarbij verschillende beoordelaars worden uitgenodigd een oordeel te geven [8]. Dit is de beoordeelde zelf die een zelfoordeel uitspreekt en het zijn anderen die de beoordeelde tijdens het werken hebben meegemaakt (collega's, supervisors, verpleegkundigen, de persoon van de administratie, patiënt). Ook hier geldt weer dat acht oordelen een tamelijk robuust totaaloordeel opleveren (met uitzondering van patiëntoordelen; daar moeten er veel meer van verzameld worden). Als we informatie van verschillende toetsmethoden met elkaar combineren (bijvoorbeeld informatie van de korte klinische beoordeling met informatie uit de 360 graden feedback) hebben we zelfs minder beoordelingen nodig binnen elke afzonderlijke methode [9].

Narratieve feedback

Om de formatieve functie van toetsing te bevorderen is feedback van groot belang. Er bestaat geen krachtigere impuls voor leren dan informatieve feedback. Wat het onderzoek in recente jaren laat zien, is dat bij de beoordeling van complexe vaardigheden, cijfers of scores op Likert-schalen een beperkte betekenis hebben en woorden des te meer [10]. Beschrijvende of narratieve informatie bevat veel meer aanknopingspunten voor verbetering en uiteindelijk ook voor beoordeling. In onze elektronische portfolio's is de knop die de narratieve informatie samenvat, een belangrijke geworden. De woorden vertellen het 'verhaal' over de betrokken persoon en is zeer bruikbaar voor het nemen van een eindoordeel of een beslissing. Eigenlijk zijn alle zachte vaardigheden complex en zijn we gebaat bij meer narratieve feedback. Sommigen stellen zelfs voor om alle cijfers af te schaffen [11].

Het moderne opleiden

Onderwijs en opleiden veranderen in een grotere beweging van inputsturing (tijd) naar outputsturing (definitie van uitkomsten). Voorheen werd opleiden vooral in tijd uitgedrukt (bijvoorbeeld de opleiding tot uroloog duurde zes jaar, tegenwoordig in uitkomsten (bijvoorbeeld met individuele verkorting duurt de opleiding voor de ene aios 5 jaar en voor de andere 5,5 jaar)). Om deze uitkomsten te definiëren zijn competentieraamwerken gedefinieerd. Een competentie is een vaardigheid om een complexe beroepssituatie tot een goed einde te brengen [12].

Diverse landen hebben eigen raamwerken ontwikkeld. Ze zijn allemaal met veel stakeholderinput tot stand gekomen om tegemoet te komen aan gesignaleerde problemen in de gezondheidszorg in dat land. Nederland heeft het raamwerk van Canada overgenomen (met de compe-

Tabel 1 Toetsinzichten uit onderzoek naar toetsing verkregen en hun praktische implicaties.

<i>Gestandaardiseerde toetsing van 'kennen, kunnen, en laten zien'</i>	<i>Praktische implicaties</i>
Competentie is sterk contextgebonden.	Zorg dat een toets bestaat uit veel items. Korte toetsen leveren veel foute beslissingen op. Combineer informatie uit verschillende bronnen. Vermijd hele zware beslissingen te nemen op een enkele toets.
Objectiviteit is niet hetzelfde als betrouwbaarheid. Subjectieve toetsen kunnen betrouwbaar zijn. Dit alles is helemaal afhankelijk van de grootte van de steekproef aan contexten (items) en aantallen beoordelaars. Vele subjectieve oordelen maken een hard oordeel.	Gebruik het menselijk oordeel voor het beoordelen van complexe vaardigheden. Gebruik veel oordelen in combinatie.
De inhoud van een toets is belangrijker dan de vorm van een toets.	Elke methode – ook meerkeuzevragen – kan hogere cognitieve vaardigheden meten, geheel afhankelijk van de formulering van de vraag. Maak vragen zo authentiek mogelijk voor de klinische praktijk. Gebruik klinische scenario's.
De kwaliteit van een toets wordt bepaald door de zorgvuldigheid van samenstelling, toetsafname en scoring.	Organiseer kwaliteitszorg rondom productie van toetsen. Laat meerdere deskundigen vragen beoordelen. Doe een psychometrische analyse na toetsafname. Gebruik input van degenen die de toetsing ondergaan.
<i>Ongestandaardiseerde toetsing van 'Doen'</i>	
Bias is inherent aan professionele oordelen	<i>Praktische implicaties</i> Gebruik een brede steekproef aan beoordelaars om systematische variatie te beperken. Gebruik procedurele maatregelen ter vermindering van onsystematische variatie, die bijdragen aan de geloofwaardigheid van het oordeel.
De waarde van de instrumenten wordt meer bepaald door de gebruikers van de instrumenten dan door de kenmerken van de instrumenten zelf.	Train de beoordelaars en beoordeel in het doel van de beoordeling. Creëer een werkomgeving waarin toetsing een onderdeel wordt van de routine.
Bij de beoordeling van complexe vaardigheden is kwalitatieve/narratieve informatie waardevoller dan scores of 'checks'.	Gebruik woorden en tekst. Pas op voor onbedoelde neveneffecten van kwantitatieve informatie.
Het geven van feedback is niet genoeg voor het gebruik van feedback.	Creëer een dialoog rondom feedback. Creëer feedback follow-up. Creëer betekenisvolle relaties tussen beoordelaar en beoordeelde.
<i>Algemeen</i>	
Geen enkele methode is perfect.	<i>Praktische implicaties</i> Varieer de verschillende methoden voor toetsing. Combineer informatie uit verschillende methoden.

tenties: medisch handelen, communicatie, samenwerking, kennis en wetenschap, maatschappelijk handelen, professionaliteit). Opvallend is dat veel van deze competenties de zogenaamde zachte vaardigheden omvatten. Alle medische kennis en chirurgische vaardigheden vallen onder 'medisch handelen'. De overige competenties zijn de niet-technische vaardigheden, die men als 'zachter' zou kunnen omschrijven. Ze zijn moeilijker te definiëren en moeilijker te beoordelen. Uit onderzoek komt echter duidelijk naar voren dat, als het misgaat in de klinische praktijk, juist deze vaardigheden in het spel zijn [13]. Klachten van patiënten in het ziekenhuis zijn geassocieerd met het gebrek aan deze vaardigheden [13]. Er is onderzoek dat laat zien dat problemen in de zachte vaardigheden tijdens de opleiding voorspellend zijn voor latere incidenten in de beroepspraktijk [14]. Er is evidentie dat zowel falen als succes op de arbeidsmarkt te maken heeft met deze vaardigheden [15]. Kortom, deze zachte vaardigheden zijn uiterst belangrijk. Om die reden besteden we tegenwoordig in de opleiding meer aan-

dacht aan deze vaardigheden en liefst willen we ze ook beoordelen. Het is volstrekt duidelijk dat deze vaardigheden slecht met een gestandaardiseerde toetsaanpak kunnen worden beoordeeld. Een betere aanpak is gebruik te maken van de methoden uit de top van de piramide: directe observatie van daadwerkelijk handelen in de praktijk, feedback gericht op verdere ontwikkeling en coaching. Het geven van informatieve feedback is noodzakelijk, maar niet voldoende voor het leren van zachte vaardigheden [16]. Het creëren van een dialoog rond feedback en het opvolgen van feedback zijn belangrijke elementen om het leren te stimuleren. Coaching of *mentoring* is een krachtig middel om reflectie, feedbackgebruik en het leren te bevorderen [17].

Het geven van feedback is een vaardigheid die opleiders moeten (leren) beheersen. Goede feedback is van groot belang. Is de bron van feedback niet geloofwaardig – wat bijvoorbeeld het geval is bij slecht gegeven feedback – dan wordt deze feedback genegeerd [18]. De *'do's and don'ts'* van feedback zijn goed beschreven in een overzicht van

Lefroy et al. [19] en in een Nederlandse richtlijn door Boor et al. [20].

Hoewel competentiegericht onderwijs nobele doelen heeft, wordt het niet altijd even goed ontvangen in de praktijk [21]. Dat heeft geleid tot een beweging om uitkomsten klinisch relevanter te gaan beschrijven. Vaardigheden als ‘communicatie’, ‘samenwerken’ of ‘professionaliteit’ blijven vage begrippen waar we allemaal wel een idee bij hebben, maar ze zijn ook lastig concreet te maken. Voor de kliniek zijn kritische beroepstaken, ook wel *Entrustable Professional Activities* (EPA’s) genoemd, eenvoudiger te begrijpen. Elke uroloog weet bijvoorbeeld wat een laposcopische/robotgeassisteerde chirurgische procedure is. Vanuit deze beroepstaak is het onmiddellijk duidelijk met wie moet worden gecommuniceerd, hoe er moet worden samengewerkt en wat professioneel handelen is. In zo’n geval kan op grond van beroepsinhoudelijke informatie in een matrix duidelijk gemaakt worden welke competenties en vaardigheden relevant zijn in welke beroepssituaties. Veel opleidingen zijn momenteel bezig deze EPA’s in kaart te brengen en nader te omschrijven.

Dikwijls worden de EPA’s op meerdere niveaus van complexiteit beschreven [22]. Toetsing wordt dan gebruikt om te bepalen of een lerende de competenties die behoren bij die beroepssituaties op een bepaald niveau van complexiteit beheerst. Het gaat dan om het identificeren van kritische beroepssituaties die een supervisor toevertrouwt (*entrust*) aan een lerende [23]. De supervisor bepaalt het bekwaamheidsniveau van een aios voor elke kritische beroepssituatie en voor steeds hogere niveaus van complexiteit [24].

Voor de bekwaamheidsniveaus wordt vaak een classificatie gebruikt waarbij de betrokkene:

1. de taak mag observeren;
2. de taak mag uitvoeren onder directe supervisie;
3. de taak mag uitvoeren met supervisie op afstand;
4. de taak zelfstandig mag uitvoeren;
5. supervisie aan junioren mag geven ten aanzien van de taak.

Daarbij is het gebruikelijk dat de betrokkene de beroepstaken op een laag niveau van complexiteit al zelfstandig mag uitvoeren of zelfs superviseren, terwijl beroepstaken op een hoger niveau van complexiteit nog onder supervisie moeten worden uitgevoerd.

Bij competenties gaat het duidelijk om vaardigheden van de persoon. Bij kritische beroepssituaties gaat het om de klinische taken die moeten worden beheerst. De competenties maken het mogelijk om de taken te kunnen uitvoeren. Goede opleidingsprogramma’s zijn gebaseerd op een matrix die beroepstaken koppelt aan zowel de benodigde competenties als aan de mijlpalen om stapsgewijs op het eindniveau te komen. Zie bijvoorbeeld de rapportage van de urologen in de Verenigde Staten [25]. Dit rapport de-

finieert zowel het opleidingsprogramma als de manier van beoordeling. Tussentijdse feedback wordt gegeven op de taak en de competenties, tot het moment waarop een bekwaamheidsverklaring kan worden afgegeven voor het uitvoeren van kritische beroepstaken op een bepaald niveau van complexiteit (EPA). Als alle taken zijn voltooid, kan een opleiding worden afgesloten.

Programmatisch toetsen

Tot slot is een recente ontwikkeling om toetsing holistischer en systematischer te organiseren [26, 27]. Programmatisch toetsen vertrekt vanuit het makkelijk te bewijzen standpunt dat geen enkele toets perfect kan zijn, maar dat het geheel van toetsen in een programma alle eisen van kwaliteit kan doorstaan. Elk individueel toetsmoment wordt gezien als een leermoment. Dat betekent dat de focus ligt op het geven van feedback, liefst zo rijk mogelijk.

Een toetsmoment wordt slechts beschouwd als één datapunt, vergelijkbaar met een pixel van een foto. Beslissingen van enige zwaarte – bijvoorbeeld doorgaan naar de volgende fase in een leertraject – kunnen slechts worden genomen als er voldoende datapunten zijn. Wanneer door de veelheid aan pixels de foto een duidelijk beeld geeft, pas dan, kan een besluit genomen worden. Meestal wordt de zak/slaagbeslissing bij één toetsmoment achterwege gelaten. Dit bevrijdt de beoordelaar van de druk om sociaal-wenselijke hoge beoordelingen te geven [28]. De beoordelaar kan zich concentreren op het geven van rijke feedback. Om feedback goed te kunnen gebruiken en om reflectie te stimuleren, worden mentoren of coaches ingezet. De uiteindelijke zwaarwegende beslissing wordt door een commissie genomen. Het commissiebesluit kent allerlei spelregels om de beslissing robuust te maken en verantwoorde besluiten te kunnen nemen [29].

Programmatisch toetsen combineert eigenlijk een aantal toetsinzichten zoals die in tab. 1 zijn samengevat tot een geheel. De aanpak maakt het mogelijk zowel de formatieve als de summatieve functie van toetsing te optimaliseren. Verschillende medische opleidingen in de wereld, zowel in de basisartsopleiding als in de vervolgoopleiding, zijn bezig programmatisch toetsen in te voeren. In Nederland is het toetsprogramma van de huisartsopleiding sterk programmatisch van karakter. Andere opleidingen volgen. Het onderzoek en de wetenschappelijke verantwoording van deze aanpak zijn in volle gang [30, 31].

Conclusie

‘Krijgen we hier nu betere dokters door?’ is een niet te missen vraag. Dat antwoord is lastig wetenschappelijk



verantwoord te geven. Wij neigen ernaar om deze vraag voorzichtig bevestigend te beantwoorden, op basis van wetenschappelijke informatie uit een diversiteit aan bronnen. Onderzoek laat zien dat afgestudeerden van opleidingen die aandacht besteden aan zachte vaardigheden ook over meer van deze vaardigheden beschikken [32–34], zonder dat het ten koste gaat van de kennis [15, 35, 36]. Een recente studie vergeleek een vervolgopleiding waarin veel aandacht wordt besteed aan communicatie (huisartsopleiding) met een opleiding waar minder aandacht aan communicatie wordt besteed (chirurgie) [37]. Interessant is dat door deze communicatieaandacht communicatievaardigheden internaliseren en personaliseren [38]. Dat gebeurt door een proces van confrontatie (o. a. door feedback), reflectie, veilige oefening, nieuwe feedback en rolmodellen. Nadat communicatievaardigheden internaliseren en personaliseren worden ze flexibel ingezet om klinische doelen beter te bereiken. Dit zijn allemaal aanwijzingen dat aandacht geven aan te ontwikkelen vaardigheden ook daadwerkelijk helpt bij het ontwikkelen van die vaardigheden. Gelukkig, opleiden werkt!

Helder is ook dat het niet makkelijk is om het moderne opleiden en de toetsing daarvan te implementeren, zeker niet in een drukke klinische praktijk. Er wordt al zo veel gevraagd van de praktiserend clinicus. Wie deze veranderingen als een bureaucratische tijger ziet die zich op ons stort, kan er beter maar mee stoppen. Voor die persoon zal het niet functioneren. Onderzoek in de vervolgopleiding laat zien dat, daar waar draagvlak bestaat voor toetsing op de werkplek, er betere implementaties ontstaan [39]. Goede toetsing kan wat ons betreft bijdragen aan betere dokters en een betere gezondheidszorg.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Literatuur

- Wass V, Vleuten C van der, Shatzer J, Jones R. Assessment of clinical competence. *Lancet*. 2001;357(9260):945–9.
- Agarwal PK, Karpicke JD, Kang SH, et al. Examining the testing effect with open-and closed-book tests. *Appl Cogn Psychol*. 2008;22(7):861–76.
- Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356(4):387–96.
- Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9):63–7.
- Vleuten CP van der, Schuwirth LW, Scheele F, et al. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol*. 2010;24(6):703–19.
- Wilkinson JR, Crossley JG, Wragg A, et al. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ*. 2008;42(4):364–73.
- Ten Cate TJ, Fluit C. Richtlijn korte praktijkbeoordeling. *Tijdschr Medisch Onderwijs*. 2010;29(5):105–34.
- Horsman M, Ten Cate TJ. Richtlijn multisource feedback voor de aios. *Tijdschr Medisch Onderwijs*. 2010;29:3–7.
- Moonen-van Loon J, Overeem K, Donkers H, et al. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv Health Sci Ed*. 2013;18(5):1087–102.
- Ginsburg S, Eva K, Regehr G. Do in-training evaluation reports deserve their bad reputations? A study of the reliability and predictive ability of ITER scores and narrative comments. *Ac Med*. 2013;88(10):1539–44.
- Hanson JL, Rosenberg AA, Lane JL. Narrative descriptions should replace grades and numerical ratings for clinical performance in medical education in the United States. *Front Psychol*. 2013;4:668.
- Merriënboer JGG van, Klink MR van der, Hendriks M. Competenties: van complicaties tot compromis. Over schuifjes en begrenzers. Den Haag: Onderwijsraad; 2002.
- Mook WN van, Gorter SL, Kieboom W, et al. Poor professionalism identified through investigation of unsolicited healthcare complaints. *Postgrad Med J*. 2012;88(1042):443–50.
- Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med*. 2005;353(25):2673–82.
- Meng C. Discipline-specific or academic? Acquisition, role and value of higher education competencies [proefschrift]. Maastricht: Universiteit Maastricht; 2006.
- Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bull*. 1996;119:254–84.
- Driessen EW, Overeem K. Mentoring. In: Walsh K (redactie). *Oxford textbook of medical education*. Oxford: Oxford University Press; 2013. pag. 265–84.
- Watling C, Driessen E, Vleuten CP van der, Lingard L. Learning from clinical work: the roles of learning cues and credibility judgments. *Med Educ*. 2012;46(2):192–200.
- Lefroy J, Watling C, Teunissen PW, Brand P. Guidelines: the do's, don'ts and don't knows of feedback for clinical education. *Perspect Med Educ*. 2015;4(6):284–99.
- Boor K, Teunissen P, Brand P. Richtlijn feedback in de medische vervolgopleiding. *Tijdschr Medisch Onderwijs*. 2011;30:37–42.
- Ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 2007;82(6):542–7.
- Merriënboer JGG van, Kirschner PA. Ten steps to complex learning, 2e druk. New York: Routledge; 2013.
- Ten Cate O. Nuts and bolts of entrustable professional activities. *J Grad Med Educ*. 2013;5(1):157–8.
- Lips J, Scheele F, Dijksterhuis M, et al. Richtlijn bekwaamverklaringen in de specialisten-opleiding. *Tijdschr Medisch Onderwijs*. 2010;29:55–7.
- ACGME. The urology milestone project 2016. <https://www.acgme.org/Portals/0/PDFs/Milestones/UrologyMilestones.pdf>. Geraadpleegd op 7 september 2016.
- Vleuten CP van der, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205–14.
- Vleuten C van der, Schuwirth L, Driessen E, et al. Twelve tips for programmatic assessment. *Med Teach*. 2015;37(7):641–6.
- Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med*. 2005;80(10 Suppl):S84–S87.
- Driessen E, Vleuten C van der, Schuwirth L, et al. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ*. 2005;39(2):214–20.

30. Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ.* 2013;13(1):123.
 31. Heeneman S, Oudkerk Pool A, Schuwirth LW, et al. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ.* 2015;49(5):487–98.
 32. Prince KJ, Eijs PW van, Boshuizen HP, et al. General competencies of problem-based learning (PBL) and non-PBL graduates. *Med Educ.* 2005;39(4):394–401.
 33. Schmidt HG, Molen HT van der. Self-reported competency ratings of graduates of a problem-based medical curriculum. *Acad Med.* 2001;76:466–8.
 34. Frambach JM, Manuel BA, Fumo AM, et al. Students' and junior doctors' preparedness for the reality of practice in sub-Saharan Africa. *Med Teach.* 2015;37(1):64–73.
 35. Schmidt HG, Molen HT van der, Te Winkel WWR, Wijnen WHFW. Constructivist, problem-based learning does work: a meta-analysis of curricular comparisons involving a single medical school. *Educ Psychol.* 2009;44(4):227–49.
 36. Vleuten CP van der, Schuwirth LW, Muijtjens AM, et al. Cross institutional collaboration in assessment: a case on progress testing. *Med Teach.* 2004;26(8):719–25.
 37. Van den Eertwegh V, Dalen J van, Van Dulmen S, Vleuten CPM van der. Residents' perceived barriers to communication skills learning: comparing two medical working contexts in postgraduate training. *Patient Educ Couns.* 2014;95:91–7.
 38. Van den Eertwegh V, Vleuten C van der, Stalmeijer R, et al. Exploring residents' communication learning process in the workplace: a five-phase model. *PLOS ONE.* 2015;10(5):e0125958.
 39. Fokkema JP, Teunissen PW, Westerman M, et al. Exploration of perceived effects of innovations in postgraduate medical education. *Med Educ.* 2013;47(3):271–81.
- prof. dr. Cees P.M. van der Vleuten** psycholoog, hoogleraar onderwijskunde
- prof. dr. Jeroen J.G. van Merriënboer** psycholoog, leren en instructie

