Achtergronden bij de Richtlijn korte praktijkbeoordeling

1. Inleiding en context

De beoordeling van artsen in opleiding tot specialist (aios) is grofweg te verdelen in drie benaderingen, te vertalen in drie vragen:

- Wat *weet* de aios?

Het antwoord op deze vraag is goed te ontleen aan de resultaten van (allerlei vormen van) kennistoetsing, zoals een schriftelijke of computergestuurde toetsing, van een voortgangsmeting of van een mondelinge kennismeting zoals 'chart-stimulated recall' (naar aanleiding van patiëntendossiers) en 'case-based discussion'.¹

- Wat *kan* de aios?

Voor deze vraag leent zich de directe observatie van het handelen. Dat kan in de dagelijkse praktijk via KPB's en in gearrangeerde omstandigheden met simulatietechnieken zoals de Objective Structured Clinical Examination (OSCE) en de Objective Structured Assessment of Technical Skills (OSATS*).¹ Ook het beoordelen van producten zoals een status, protocol of wetenschappelijk verslag valt hieronder. In de snijdende beroepen kan een product overlappen met een vaardigheid, indien sprake is van een beoordeelbaar resultaat van een ingreep. Aan de OSATS is een afzonderlijke Richtlijn gewijd die in dit supplement is opgenomen.²

- Hoe *functioneert* de aios als professional in de praktijk?

Deze vraag wordt het best beantwoord via evaluatie van het functioneren over een periode in het verleden vanuit meerdere bronnen. Hiervoor is een multisource feedback procedure nuttig. Ook hieraan is een andere Richtlijn gewijd.³

De KPB is direct ontleend aan de zogeheten Mini Clinical Evaluation Exercise (miniCEX), ontwikkeld bij de American Board of Internal Medicine (ABIM). De systematische beoordeling van klinische competenties bij aios is vanaf 1972 in de VS op gang gekomen nadat de ABIM het mondelinge examen had afgeschaft.⁴ De Clinical Evaluation Exercise (CEX) werd geïntroduceerd als nieuw evaluatiesysteem.⁵ Hierbij observeert de supervisor een aios bij een uitgebreid patiëntcontact van begin tot eind en geeft hierover feedback in een gestructureerd examen van ongeveer twee uur. De betrouwbaarheid en validiteit van de CEX bleken laag te zijn.⁶⁻⁹ Na 25 jaar werd deze aanpak daarom herzien. De miniCEX gaat uit van meerdere korte beoordelingsmomenten van klinische consulten op de werkvloer. De methode is flexibel omdat diverse patiëntgebonden klinische situaties als beoordelingsmoment kunnen fungeren, de toepassing per keer weinig tijd vergt en de beoordelingssituatie natuurgetrouwer is dan bij een lang patiëntenexamen.

* In de Nederlandse context wordt de term OSATS niet gebruikt voor een objectief gestructureerde assessment, maar voor observatie van technisch handelen in de praktijk.

Er is voor deze Richtlijn systematisch gezocht naar literatuur over KPB methoden, onder verschillende benamingen. Hiervoor zijn de zoekmachines Pubmed en Cochrane gebruikt, met trefwoorden mini-CEX, CEX, Workplace-based assessment, internationale termen voor aios, gecombineerd met de KPB aspecten, en ingeperkt tot de medische vervolgopleidingen. Dit leverde in Pubmed 376 en in Cochrane 434 artikelen op. Vervolgens werd op titel, abstract en tekst gescreend. Na ontdebelling bleken 32 artikelen bruikbaar. Gaandeweg werd nieuwe literatuur toegevoegd via extra searches met nieuwe termen, opgedaan in de initiële search. Daarnaast werden van relevante artikelen gerelateerde verwijzingen gescreend en werd gebruik gemaakt van de Pubmed's service MyNCBI om nieuwe artikelen binnen de opgegeven zoekstrategie te signaleren. Tenslotte zijn relevante publicaties die in de periode van schrijven (2008-2010) verschenen handmatig toegevoegd.

2. Definitie en eisen van de KNMG

Een KPB bestaat uit een observatie van een specifieke werksituatie gedurende vijf tot vijftien minuten, gevolgd door feedback die schriftelijk wordt vastgelegd en mondeling wordt toegelicht.

Hiermee onderscheidt de KPB zich van andere beoordelingen, zoals bij examensituaties, bij simulaties, bij schriftelijke toetsingen en bij multisource feedback. De KPB is bij uitstek een instrument om op niveau vier van Miller (knows, knows how, shows how, en *does*) de competentie van de aios op de werkvloer steekproefsgewijs te beoordelen.¹⁰

De KPB is een 'snapshot' of steekproef uit de praktijk en heeft vooral waarde indien er meerdere KPB's worden uitgevoerd bij dezelfde aios. Het is dus niet een verslag van een periode (zoals een werkdag of langer, een dienst, een poli-ochtend,

e.d.), of van een product (zoals een status). De term 'KPB' wordt in de praktijk zowel gebruikt voor de beoordelingsprocedure als ook ter aanduiding van een ingevuld KPB formulier. Samenvattend kent de KPB vijf kenmerken, en kan daarmee gedefinieerd worden als een:

- directe observatie van het optreden van een aios,
- in een natuurlijke werksituatie in de (patiënten)zorg,
- gedurende vijf tot vijftien minuten,
- gevolgd door een korte schriftelijke vastlegging van het oordeel,
- en direct gevolgd door een kort feedbackgesprek.

Bij de KPB maakt men altijd gebruik van een formulier. Er zijn in korte tijd in Nederland veel varianten ontworpen van het basisformulier dat afgeleid is van het model uit de VS dat als voorbeeld heeft gediend. Om van een KPB te kunnen spreken zijn enkele kenmerken noodzakelijk aanwezig en enkele kenmerken optioneel.

Vaste kenmerken van ieder KPB formulier zijn:

- Identificatie/naam van de aios
- Identificatie/naam van de beoordelaar
- Datum
- Setting of situatieaanduiding
- Lijst met te beoordelen aspecten van het handelen
- Ruimte voor schriftelijke feedback, in het bijzonder voor suggesties voor verdere ontwikkeling

Optionele kenmerken van het KPB formulier zijn:

- Scoringsopties met schaalwaarden
- Opleidingsstadium van de aios
- Type patiënt of type handeling
- Toelichting per beoordelingsaspect
- Instructie voor invulling van het formulier

- Plaats voor een handtekening van beoordelaar en aios

In 2004 werd de Korte Klinische Beoordeling of KKB, een directe vertaling van het prototype van de miniCEX,¹¹ in Nederland door het Centraal College voor Medische Specialismen (CCMS) in een Kaderbesluit opgenomen en verplicht gesteld als instrument voor de beoordeling van aios.¹² In 2007 werd de term KPB vastgelegd in het CCMS protocol over toetsen en beoordelen van aios.¹³ De naamsverandering van KKB naar KPB weerspiegelt de opvatting dat de methode niet slechts geschikt is voor de beoordeling van de aios bij een patiëntgebonden klinisch consult, maar voor ieder observeerbaar moment waarop de aios relevant klinisch werk in de praktijk verricht. In het Kaderbesluit CCMS van 2009, artikel B11, is opgenomen dat in iedere opleiding bij iedere aios tenminste tien maal per jaar een KPB wordt uitgevoerd.¹⁴

3. Doel en functie van de korte praktijkbeoordeling

De KPB behoort tot de groeiende familie van 'workplace assessment' methoden, instrumenten om klinische competenties op de werkplaats te beoordelen, waarvan onlangs een uitstekend overzicht is verschenen.¹⁵ Hoewel de KPB mede is ontwikkeld om de betrouwbaarheid van eindoordelen over aios te vergroten, is de algemene opvatting thans dat de KPB vooral een onderwijsfunctie heeft en informatie aan de aios moet verschaffen over de voortgang.

KPB's hebben in de eerste plaats dus een formatief doel, c.q. het doel om van de beoordeling iets te leren, waarbij verbetering van het handelen op de voorgrond staat.¹⁶ Dit doel wordt bij een KPB bereikt doordat aan drie basisprincipes voor formatieve beoordelingen voldaan wordt: a)

er wordt beoordeeld wat men wil weten; b) de feedback wordt direct gegeven en c) de beoordelingen worden zo ingezet dat de beoordeelde in de richting van de gewenste uitkomst gestuurd kan worden.

Van belang is dat de KPB verweven wordt met zowel de opleidingsdoelen als met de klinische praktijk.¹⁶⁻¹⁷ Dit is belangrijk omdat toetsing het leergedrag stuurt. Als een aios op een valide manier wordt beoordeeld in de praktijk zal diens leergedrag zich richten op het zo goed mogelijk handelen in die praktijksituatie. Zo wordt leren voor het beoordelingsmoment gelijk aan leren om een betere arts of specialist te worden.¹⁸⁻¹⁹ Competentie is vooral goed te beoordelen bij samengestelde taken, waarbij het nodig is om zowel kennis, vaardigheid, probleemoplossend vermogen als attitude bij elkaar te brengen.^{17 20}

De KPB's kunnen behalve een formatief soms ook een summatief doel dienen. De termen formatief (ontwikkelingsgericht) en summatief (gericht op beslissingen) worden vaak als strikt gescheiden gezien. In de praktijk is er eerder sprake van een glijdende schaal, van een terloopse opmerking met het karakter van een tip tot een beslissing over slagen voor of stopzetten van de gehele opleiding. Hoe zwaarder de consequenties zijn, des te belangrijker is de betrouwbaarheid van het oordeel. Als uit een serie KPB's een duidelijke progressie naar voren komt, kan afhankelijk van andere informatie, bijvoorbeeld uit kennistoetsen, uit multisource feedback, uit de scores op CAT procedures en uit OSATS scores, tot een verantwoorde summatieve beoordeling besloten worden.

Samengevat hebben KPB's in de eerste plaats een formatieve functie, waarin feedback voorop staat, en een algemene educatieve functie, doordat observatie wordt afgedwongen. Tenslotte is de be-

langrijkste functie van de KPB misschien wel dat er observatie in de praktijk geschiedt, hetgeen vaak niet of nauwelijks spontaan gebeurt.²¹

4. Situaties waarvoor de korte praktijkbeoordeling geschikt is

Bij een KPB wordt de aios beoordeeld op het handelen in de dagelijkse praktijk.¹⁶ Het kan om uiteenlopende activiteiten gaan: een lichamelijk onderzoek of een deel daarvan, een anamnese, een slecht-nieuwsgesprek, een technische handeling, een ingreep of operatie, een presentatie tijdens een ochtendrapport of tijdens de overdracht, een referaat of een andere activiteit. De methode leent zich voor de koppeling van korte observaties aan die elementen die voor de opleiding als kritische activiteiten zijn aangemerkt, c.q. aan wat wel wordt genoemd 'entrustable professional activities', of aan kritische beroepssituaties: activiteiten die op een bepaald moment tijdens de opleiding aan een aios voor zelfstandige uitvoering moeten kunnen worden toevertrouwd.²² Observatie van een technische verrichting, van een presentatie of van het leiden van een vergadering, van het spreken met de familie van een patiënt, al of niet na overlijden, en van het uitvoeren van onderwijs, zijn voorbeelden van activiteiten die observeerbaar zijn en daarom geschikt voor de KPB, mits er een toepasselijk formulier is.

Het CCMS kaderbesluit van 2009 noemt de volgende situaties geschikt voor de KPB: het (poli-)klinische consult, een technische ingreep, het omgaan met medewerkers, het multidisciplinair overleg (MDO) en het beoordelen van een biops of beeldopname. Alle gevallen waarin door korte, directe observaties een goed beeld kan worden gekregen van de kwaliteit van handelen lenen zich in beginsel voor een KPB. Of 'omgaan met medewerkers' daar-

voor geschikt is moet betwijfeld worden, omdat een momentopname hiervoor waarschijnlijk een onevenwichtig beeld geeft. Multisource feedback procedures zijn daarvoor geschikter.

De toepassing van de mini-CEX is voor klinische situaties met verschillende categorieën patiënten beschreven. Onderscheid is wel gemaakt tussen opgenomen patiënten, ambulante patiënten op de poli en patiënten op de Spoedeisende Eerste Hulp (SEH). Ook wordt soms een onderscheid gemaakt tussen nieuwe en controlepatiënten.^{4 23-25}

De KPB is oorspronkelijk bedoeld als observatie-instrument voor het klinisch consult. Er zijn echter veel meer situaties waarin de KPB kan worden toegepast, hoewel niet iedere situatie zich leent voor de observatie van alle relevante competenties. In Tabel 1 zijn acht praktijksituaties aangegeven die geschikt zijn voor een KPB, gekoppeld aan de CanMEDS gebieden. Per situatie is het uiteraard niet altijd mogelijk om al deze aspecten te observeren, maar de tabel geeft houvast over de mogelijke te beoordelen competenties.

Er is onder verschillende benamingen gepubliceerd over uiteenlopende varianten op de KPB die vermeldenswaard zijn:

- De *Minicard* is een variant op de mini-CEX met vier groepen van beoordelingsaspecten met elk drie items:²⁶ anamnese (communicatie, data collectie/medische kennis en professioneel gedrag), lichamelijk onderzoek (medische kennis/lichamelijk onderzoek als diagnosticum, interpretatie van het lichamelijk onderzoek en professioneel gedrag), beoordeling van bevindingen (mondeling verslag, redenering en plan van aanpak), en plan/begeleiding/houding (communicatie, medische kennis en professionaliteit).

- Het *In-Training Evaluation Report* (ITER), de verplichte workplace beoordeling die in de Canadese vervolgopleidingen al sinds medio jaren zeventig bestaat.²⁷ Veel voorbeelden zijn op het internet eenvoudig terug te vinden. Het formulier lijkt op dat van de KPB, maar is in het gebruik wezenlijk verschillend. De ITER betreft het oordeel over het functioneren gedurende een periode. De ITER heeft een summatief karakter en moet worden vastgelegd voor de 'Royal College of Physicians and Surgeons of Canada'.
- De *Professionalism Mini Evaluation Exercise* (P-mex) is een voorbeeld van een KPB, gericht op één competentiedomein. Het formulier kent vierentwintig items en meet vier eigenschappen van professionaliteit in korte observaties in uiteenlopende situaties.²⁸
- De *Ophthalmic Clinical Evaluation Exercise* (OCEX) heeft de ACGME-competenties uit de VS als leidraad; de competenties patiëntenzorg, professionaliteit, communicatievaardigheden en medische kennis zijn opgenomen op het beoordelingsformulier.²⁹⁻³⁰ Er worden vier aspecten onderscheiden: anamnese, lichamelijk onderzoek, communicatie/professionaliteit en casuspresentatie. Per aspect worden zes tot twaalf concrete items genoemd die beoordeeld worden.
- De *Objective Structured Assessment of Technical Skills* (OSATS) wordt in Nederland gebruikt voor de beoordeling van een technische vaardigheid in de praktijk, in het bijzonder tijdens operaties.³¹ Deze procedure is in feite een KPB, zij het dat de duur van de observatie veel langer kan zijn dan de tien tot vijftien minuten uit de definitie van de KPB. Strikt genomen is de term OSATS niet juist. De OSATS is een in Toronto ontwikkelde stationstoets met simulatietechnieken voor de beoordeling van technische vaardigheden en is daarom 'objective' en 'structured'.³² De Nederlandse 'OSATS' is niet vooraf gestructureerd, maar is een beoordeling in de praktijk, die een meer subjectief karakter heeft en waarvan de psychometrische kenmerken vanzelfsprekend minder sterk zijn dan de echte OSATS.³³ In het Engels wordt *Direct Observation of Procedural Skills* (DOPS) wel gehanteerd voor een op vaardigheden gerichte KPB.¹⁶ Omdat in Nederland de OSATS een belangrijke, in het Kaderbesluit CCMS vastgelegde, procedure is voor de snijdende vakken is er een beknopte afzonderlijke In VIVO Richtlijn aan gewijd die in relatie tot de onderhavige Richtlijn moet worden gelezen. Veel van de achtergrondinformatie van de KPB Richtlijn is ook van toepassing voor de OSATS Richtlijn.²
- De *Global Rating Assessment of Skills in Intraocular Surgery* (GRASIS) is een beoordelingsformulier dat speciaal is ontwikkeld voor directe observatie bij oogoperaties.³⁴ Hierbij wordt op een vijf-puntsschaal antwoord gegeven op specifieke aandachtspunten zoals omgang met instrumenten, kennis van procedure, gebruik van de niet-dominante hand, omgang met (oog)weefsel, etc. Het is te beschouwen als een oogheekundige variant op het OSATS formulier. In de anesthesie zijn KPB's beschreven voor specifieke situaties zoals het aanleggen van een epiduraalcatheter, een pre-operatief consult, een spoedinductie en een ronde op de Intensive Care Unit (ICU).³⁵ In de dermatologie is de KPB beschreven voor het beoordelen van een patch test³⁶ en in de chirurgie voor cholecystectomieën.³⁷
- De *Objective Structured Assessment of Laparoscopic Salpingectomy* (OSA-LS) is beschreven als een methode met een

onderverdeling in algemene vaardigheden en specifieke vaardigheden.³⁸ De scoringssystemen voor deze technische vaardigheden kennen soms twee lijsten, één met punten voor technische zaken waaraan de aiOS voldaan moet hebben en één met fout-punten.³⁷

Een systematische en onderwijskundig verantwoorde, alternatieve benadering is om het dagelijks werk van het specialisme te analyseren, de kritische activiteiten te identificeren die in de loop van de opleiding moeten kunnen worden toevertrouwd aan de aiOS en vervolgens voor de belangrijkste hiervan een KPB formulier te ontwerpen.³⁹

5. Betrouwbaarheid, validiteit en aantal korte praktijkbeoordelingen

De KPB wordt beschouwd als een valide beoordelingsmethode – validiteit in de zin van authenticiteit – met een beperkte be-

trouwbaarheid, afhankelijk van het aantal KPB's waarop een oordeel is gebaseerd en van de zorgvuldigheid waarmee de KPB's zijn ingevuld.

Summatief beoordelen met KPB's stelt hoge psychometrische eisen, die niet altijd makkelijk zijn te vervullen. Beoordelingen zijn onder meer afhankelijk van opleiders en opleidingen en kunnen per opleidingssituatie wezenlijk verschillen.¹⁶ Om deze reden zijn meerdere KPB beoordelingen door getrainde beoordelaars nodig voor een betrouwbare beslissing over voldoende of onvoldoende. De KPB is een minder betrouwbare beoordelingsmethode dan examinering met simulatietechnieken.⁴⁰ Een directe praktijkobservatie heeft echter ook duidelijke voordelen: het is gemakkelijker te organiseren en goedkoper en de beoordeling is meer valide in de zin van authentiek.⁴¹ Simulatiepatiënten of -hulpmiddelen zijn niet altijd voorhanden. Minstens zo belangrijk is dat echte patiënten in echte situaties zorgen

Tabel 1. CanMEDS competentiegebieden en situaties voor de KPB.

	Medisch handelen	Communicatie	Samenwerking	Kennis & wetenschap	Organisatie	Maatschappelijk handelen	Professionaliteit
Klinisch consult bij opname of poli	X	X			X		X
Zaavisite	X	X	X		X		X
Technische verrichting en functieonderzoek	X		X		X		X
Technisch handelen binnen de OK (OSATS)	X	X	X		X		X
Patiëntpresentatie bij rapport of overdracht	X	X		X			
Specifieke gesprekken met patiënten of familie*		X				X	X
Intercollegiale consultatie	X	X	X				
Voeren van mono- of multidisciplinair overleg**	X	X	X		X	X	X

* slechtnieuwsgesprek, ontslaggesprek, gesprekken met familie rond sterven (niet-reanimatiegesprek, verzoek om obductie of orgaandonatie).

** met een collegae, consulenten, maatschappelijk werk, fysiotherapie, verpleging.

voor dynamiek. Patiënten met echte gevoelens, echte pijn en echte karakters geven de aios een veel breder en meer valide spectrum van mogelijke handelingen, waarbij het brede competentiebegrip beter wordt benaderd.

Betrouwbaarheid

De betrouwbaarheid is de mate waarin een meting of beoordeling reproduceerbaar is, c.q. bestand is tegen verstoringen die de meting beïnvloeden. Een betrouwbare KPB moet eenzelfde uitslag opleveren bij een direct herhaalde uitvoering. Bij de vraag hoe reproduceerbaar KPB uitslagen zijn, moet de uitslag eerst worden gedefinieerd. Er zijn in principe drie soorten uitslagen: scores op individuele items, een overall score die soms is toegevoegd en feedbackopmerkingen. Daarnaast is voor de uitslag belangrijk wat de norm is waartegen de scoring wordt afgezet. Die kan a) absoluut zijn, dat wil zeggen een vergelijking zijn met een uiteindelijk te bereiken niveau, b) relatief zijn, dat wil zeggen een vergelijking betreffen met het gewenste of verwachte opleidingsstadium waarin de aios zich bevindt, c.q. een vergelijking met collega-aios betreffen, of c) een ontwikkelingsnormering betreffen, dat wil zeggen een vergelijking met vorige KPB's bij dezelfde aios. De betrouwbaarheid is niet zozeer afhankelijk van welke normering wordt gehanteerd, maar wel van de vraag of alle beoordelaars dezelfde normering gebruiken en of ze de KPB consciëntieus invullen.

De kans is groot dat een tweede afname van een KPB bij eenzelfde aios niet precies gelijk is, zeker als de KPB wordt afgenomen door een ander staflid. Scores zullen verschillen en feedbackopmerkingen zullen verschillen. Dat kan zeer waardevol zijn omdat de aios veel kan leren van verschillende beoordelaars, maar in termen van betrouwbaarheid (reprodu-

ceerbaarheid) voldoet één KPB niet om een 'harde' uitspraak te doen over de competentie van de aios. De reproduceerbaarheid van scores op individuele items en van feedbackopmerkingen op de KPB is dus laag; die van een overall-score of van een somscore is iets hoger. Ook somscores zijn niet altijd voldoende betrouwbaar voor een summatief oordeel. Het is ook de vraag of een gemiddeld oordeel nuttig is, omdat juist het onderscheid tussen de items waardevol is: een aios is misschien goed in organisatie en in efficiëntie, maar niet in communicatieve vaardigheden. Er gaat dus ook waardevolle informatie verloren bij middeling van scores.²⁰

Open opmerkingen en feedbackadviezen weerspiegelen de observaties van de clinicus die erbij staat; die kunnen anders zijn dan de observatie van een andere clinicus. Opmerkingen worden vaak waardevoller gevonden dan scores. Of ze ook betrouwbaarder zijn is moeilijk te zeggen. Uitrekenen daarvan is lastig, omdat ze daarvoor eerst gecategoriseerd en getransformeerd moeten worden naar een score, hetgeen eigenlijk nooit gebeurt.

De betrouwbaarheid van een score op een bepaald aspect die gebaseerd is op meerdere KPB's is evident beter dan die van enkelvoudige KPB's. Dat komt doordat ruis, door toevallige omstandigheden of door het verschil tussen personen, over meerdere metingen uitmiddelt; de gemiddelde score komt bij de toename van het aantal metingen steeds dichterbij de score die de 'waarheid' weerspiegelt (de 'ware vaardigheid'). Hoeveel metingen nodig zijn voor een betrouwbare uitspraak over de competentie hangt af van het gewicht dat men hecht aan de nauwkeurigheid van die uitspraak. Zijn de consequenties groot, dan moet hij nauwkeurig zijn en dan zijn er veel metingen nodig. Wordt er alleen feedback verwacht, dan zijn de eisen veel minder hoog. Dan is één KPB al waardevol.

Men kan concluderen dat KPB's alléén niet geschikt zijn om een harde uitspraak te doen die ernstige consequenties heeft voor de voortgang van de opleiding, ook niet als er tien KPB's zijn afgenomen. Daarvoor is altijd meer informatie nodig. Daar komt bij dat de kwaliteit van een serie KPB's afhankelijk is van de kwaliteit van de afzonderlijke KPB's ervan. KPB's die slordig worden ingevuld of te gunstig worden ingevuld om een aiOS ter wille te zijn, verstoren het algemene beeld dat ontstaat. Training van stafleden in het toepassen van KPB's kan de betrouwbaarheid belangrijk verbeteren, maar voorzichtigheid blijft geboden.

Aantal korte praktijkbeoordelingen voor een betrouwbaar oordeel

Cronbach's alpha (range 0.0 – 1.0) wordt dikwijls gebruikt om de betrouwbaarheid van toetsen mee te schatten. Bij schriftelijke studietoetsen wordt een alpha van 0.80 meestal als minimumcriterium gehanteerd voor betrouwbaarheid, hoewel deze vaak lager ligt. Voor de mini-CEX zijn veel afnames vereist voor een redelijke alpha; Norcini noemt vier tot veertien;⁴ Durning et al. vonden een goede betrouwbaarheid bij zeven KPB's.²³ Alves de Lima et al. vonden lagere betrouwbaarheden en schatten dat tot vijftig KPB's nodig zijn voor betrouwbare summatieve uitspraken.¹⁹ Williams et al. schatten in hun review dat een minimumaantal van zeven tot elf KPB's genoeg is om een betrouwbaar oordeel te vellen over de 'overall clinical competence', en dat het dubbele nodig is om een betrouwbaar beeld van specifieke competenties te krijgen.⁴²

De tien KPB's per jaar die het CCMS in Nederland vereist, lijkt een redelijk aantal om bij een 'normaal' functionerende aiOS de competentieontwikkeling te kunnen volgen. Als er twijfels bestaan over de competentie van een aiOS, dan is het aan te bevelen dit aantal te vergroten.

Het aantal beoordelaars is mede bepalend voor de betrouwbaarheid van het oordeel dat gebaseerd is op de somscore van een serie KPB's en het is daarom van belang om zoveel mogelijk verschillende beoordelaars in te schakelen.⁴³

Contextspecificiteit

Een belangrijke waarneming uit onderzoek is dat medische competentie contextspecifiek is. Artsen die goed een bepaalde aandoening kunnen diagnosticeren zijn niet per definitie kundig in het diagnosticeren van andere (zelfs gerelateerde) aandoeningen.⁴² Het is daarom verstandig om observatiemomenten te spreiden over verschillende situaties, ziektebeelden en typen patiënten. Dat verkleint de contextspecificiteit van de scores. Wat men wil is dat, indien een aiOS voor een aantal KPB's een voldoende beoordeling heeft gekregen, men mag aannemen dat hij of zij in andere praktijksituaties ook adequaat handelt. Anders gezegd: het oordeel moet te generaliseren zijn.

Interbeoordelaarbetrouwbaarheid

Beoordelingen kunnen per beoordelaar aanzienlijk verschillen. Twee factoren lijken hiervoor verantwoordelijk. Ten eerste vinden verschillende beoordelaars verschillende dingen belangrijk; ten tweede verschillen de verwachtingen die beoordelaars vooraf hebben van een aiOS. Uit onderzoek van Margolis et al. blijkt dat de strengheid van de beoordelaar meer invloed heeft op het cijfer dan de aiOS zelf.⁴³ Er worden ook hoge correlaties tussen items gevonden: als het ene item streng wordt beoordeeld, dan wordt het andere dat ook. Dat is een halo-effect. Norcini et al. vonden dat de interbeoordelaarbetrouwbaarheid van de mini-CEX noch door de setting (opgenomen patiënten, patiënten bij de polikliniek of SEH) noch door de typen patiënten (nieuwe of con-

trolepatiënten) sterk beïnvloed wordt.⁴⁴ Uit recenter onderzoek bleek echter wel dat beoordelaars compenseren voor de moeilijkheidsgraad van de casus.⁴⁵

De betrouwbaarheid en objectiviteit kan vergroot worden door gebruik te maken van gerichte video-opnames. Bij gynaecologische laparoscopieën³⁸ en laparoscopische cholecystectomieën³⁷ werd alleen het operatiegebied in beeld gebracht, waardoor de aios onherkenbaar was. De objectiviteit werd hiermee vergroot.

In de kliniek is er vaak weinig onderlinge afstemming tussen stafleden over beoordelingscriteria. De ene opleider geeft een voldoende omdat hij iets goeds heeft gezien terwijl een ander dat wellicht ook vond maar een onvoldoende scoort omdat het niet volledig was. Verder scoren sommige opleiders een wat chaotische anamnese als een onvoldoende anamnese, terwijl anderen dit rekenen onder organisatie/ efficiëntie omdat het zoveel tijd kost. Een training vooraf kan dit verschil in ieder geval reduceren.

Er zijn auteurs die kritiek hebben op de traditionele benadering van betrouwbaarheid als interne consistentie en interbeoordelaarsovereenstemming bij dit soort beoordelingen, juist omdat experts moeten kunnen verschillen van inzicht.⁴⁶⁻⁴⁷ Schuwirth en Van der Vleuten stellen de vraag of die diversiteit zo erg is en of men dus wel moet streven naar homogeniteit onder de beoordelaars.²⁰ De werkelijkheid is ook divers. Patiënten en collega's zijn verschillend en denken en doen niet gelijk. Dat vergt ook van de aios flexibiliteit; hij/zij moet leren om het beste te halen uit beoordelingen en feedback van een grote diversiteit van beoordelaars. Dat staat ver af van de wens tot gestandaardiseerde beoordelingen, waarin iedereen gelijk wordt behandeld, alsof het een schriftelijk examen betreft. Competentie kan ook

worden gezien als een wisselwerking tussen het individu en de omgeving en betreft dan de vraag hoe goed de aios in *deze* omgeving met *deze* collega's kan werken.⁴⁷⁻⁴⁸

Validiteit

De inhoudsvaliditeit wordt bepaald door de aspecten die op het KPB formulier staan, en dus beoordeeld worden, en door de vraag of deze dekkend zijn voor wat er gemeten moet worden. Daarnaast is de zorgvuldigheid waarmee er beoordeeld wordt van groot belang. Dat beïnvloedt de betrouwbaarheid maar ook de validiteit, omdat er systematische vertekening kan optreden. Albanese heeft gewezen op het gevaar dat de toenemende één-op-één beoordelingen op de klinische werkvloer de normen van onderwijs compromitteren.⁴⁹⁻⁵⁰ Er worden namelijk vaak te gunstige beoordelingen gegeven. Gemakzucht bij de beoordeling leidt snel tot te hoge scores. Dat pleziert de aios en vaak ook de beoordelaar, maar maskeert zwakte. Een te positieve documentatie, vooral als het een serie KPB's met handtekeningen van beoordelaars betreft, maakt het de opleider later zeer moeilijk om werkelijke problemen aan de orde te stellen. Het is ook daarom essentieel dat beoordelaars goed geïnstrueerd en getraind worden en consciëntieus te werk gaan.

Als via KPB resultaten een beeld ontstaat van de wijze waarop de competentie van een aios in de loop van de tijd toeneemt, pleit dit voor de constructvaliditeit. Blijkbaar wordt er dan iets gemeten (een 'construct') waarvan je verwacht dat het toeneemt en dat wordt dan gevonden. Holmboe et al. concludeerden dat stafleden over het algemeen goed het onderscheid kunnen maken tussen een onvoldoende, voldoende en een goede prestatie.⁴¹ Zij lieten veertig stafleden naar video's kijken waarin aios duidelijk verschil-

lende prestatieniveaus lieten zien. Hoewel het juiste beoordelingsniveau significant vaker werd gekozen, hetgeen pleit voor de constructvaliditeit, was de spreiding wel groot.

Durning et al. vonden steun voor de predictieve en convergerende validiteit van de KPB door hoge correlaties met andere maten voor klinische competentie.²³

6. Welke competenties worden beoordeeld met een korte praktijkbeoordeling?

Diversiteit in settings en te beoordelen activiteiten hebben geleid tot het ontstaan van uiteenlopende KPB formulieren. Oorspronkelijk bestond de mini-CEX uit vijf items: anamnese, lichamelijk onderzoek, klinische beoordeling, omgang/professioneel handelen en een totaalindruk.^{1 4 23} In de meeste beschrijvingen van de mini-CEX gaat het om een formulier met zeven items: anamnese, lichamelijk onderzoek, omgang/professioneel handelen, klinische beoordeling, begeleiding, organisatie/efficiëntie en een 'overall competence'.^{18 19 24 51}

De KKB (Korte Klinische Beoordeling) zoals die oorspronkelijk in Nederland werd geïntroduceerd, en waarschijnlijk thans het meest wordt gebruikt, bevat de items anamnese, lichamelijk onderzoek, professioneel handelen, probleemanalyse en klinisch redeneren, vervolgonderzoek en begeleiding, communicatie met de patiënt en tot slot organisatie en efficiëntie.

Mede op basis van de literatuur en de beschrijving van KPB formulieren in nieuwe opleidingsplannen voor de vervolgopleidingen⁵² is in Tabel 2 een overzicht gegeven van ruim vijftig beoordelingsaspecten die opgenomen zouden kunnen worden in KPB formulieren in de meest voorkomende praktijksituaties. De lijst is niet limitatief; sommige specialismen zullen situaties kunnen

formuleren waarvoor een KPB formulier met nog andere beoordelingsaspecten nuttig is.

Het is aan te raden om niet alle mogelijke beoordelingsaspecten op één KPB formulier op te nemen, maar alleen die aspecten die voor een bepaalde praktijksituatie relevant zijn. Een KPB beoogt geen volledigheid. Anderzijds wordt het ook zeer onoverzichtelijk als er gewerkt wordt met zeer veel verschillende formulieren. Een beperkt aantal verschillende formulieren (niet meer dan drie of vier) per opleiding wordt aangeraden. Deze formulieren moeten dan algemeen toepasbaar zijn voor meerdere situaties. Een formulier voor technische verrichtingen moet voor alle relevante technische verrichtingen geschikt zijn. De consequentie kan zijn dat sommige items niet altijd kunnen worden ingevuld. Voor de formatieve doeleinden waarvoor de KPB vooral geschikt is, is dat geen enkel bezwaar. Dat ligt anders wat betreft sterk summatieve doeleinden. Dan is inhoudsvaliditeit van het oordeel belangrijk en is het gewenst om volledige, zorgvuldig ingevulde formulieren te hebben. Als richtlijn kan dienen dat drie tot maximaal zeven aspecten worden opgenomen op een KPB formulier.

De relatie tussen beoordelingsaspecten en CanMEDS competentiegebieden roept soms ook vragen op. Het is zinvol om waar dat mogelijk is de CanMEDS competentiegebieden als kader te vermelden op het formulier zoals in Tabel 3 is aangegeven.

De beoordelingsaspecten moeten goed begrepen worden door de beoordelaars. Het voert te ver om bij alle mogelijke beoordelingsaspecten in deze Richtlijn een definitie of beschrijving te geven. Op veel formulieren staat separaat (ernaast of op de achterkant) een korte definitie van elk aspect. Dat is aan te raden, en verdient de voorkeur boven het opnemen van een be-

Tabel 2. Beoordelingsaspecten bij de meest voorkomende praktijksituaties waarvoor de KPB inzetbaar is.

Klinisch consult bij opname of poli <ul style="list-style-type: none"> • Anamnese • Lichamelijk onderzoek • Professioneel handelen • Probleemanalyse en klinisch redeneren • Vervolgonderzoek en begeleiding • Communicatie met de patiënt • Organisatie en efficiëntie 	Zaalvisite <ul style="list-style-type: none"> • Anamnese • Lichamelijk onderzoek • Professioneel handelen • Probleemanalyse en klinisch redeneren • Communicatie met de patiënt • Organisatie en efficiëntie
Technische verrichting en functieonderzoek <ul style="list-style-type: none"> • Organisatie vooraf • Uitleg en instructie aan patiënt • Instrumentgebruik en zorgvuldigheid • Communicatie met assistentie • Communicatie met patiënt tijdens handeling • Afsluiting • Efficiëntie 	Technisch handelen binnen de OK <ul style="list-style-type: none"> • Weefselgevoel • Tijd en beweging • Hanteren van instrumenten • Instrumentkennis • Gebruik van assistentie • Voortgang operatie • Kennis van procedure • Communicatie met assistentie • Communicatie en samenwerking met collega's
Patiëntpresentatie bij rapport of overdracht <ul style="list-style-type: none"> • Voorbereiding • Ordening diagnostische gegevens (anamnese, onderzoek, diagnostiek) • Probleemstelling; hoofd- en bijzaken • Differentiële diagnose • Behandelplan, prognose, complicaties • Decursus • Relevante vakkennis • Opbouw voordracht • Spreken (taal, stijl, dictie) en houding • Kwaliteit visuele hulpmiddelen • Tempo en tijdsbewaking 	Voorzitten mono- of multidisciplinair overleg <ul style="list-style-type: none"> • Organisatie vooraf • Leiding nemen • Vergader- en discussietechniek • Formulering probleemstelling en prioritering hoofd- en bijzaken • Discussiesamenvatting maken • Actieplan formuleren
Specifieke gesprekken met patiënten of familie <ul style="list-style-type: none"> • Organisatie en voorbereiding • Professionele distantie en betrokkenheid • Omgaan met sterke emoties • Communicatie met patiënt en familie 	Intercollegiaal overleg <ul style="list-style-type: none"> • Beknopte, heldere vraagstelling formuleren • Communicatie met collega's • Formulering van voorstellen en adviezen • Professionele intercollegiale opstelling

Tabel 3. Voorbeeld van KPB items met het CanMEDS raamwerk als basis.

Competentiegebied	Beoordelingsaspect	Schaal							
Medisch Handelen	Blijk van kennis en vaardigheid bij deze activiteit	4	5	6	7	8	9	10	
Communicatie	Communicatie met patiënt, collega's of anderen	4	5	6	7	8	9	10	
Samenwerking	Samenwerking met assistentie tijdens deze activiteit	4	5	6	7	8	9	10	
Organisatie	Organisatie en efficiëntie bij deze activiteit	4	5	6	7	8	9	10	
Kennis en wetenschap	De kennisoverdracht tijdens deze activiteit	4	5	6	7	8	9	10	
Maatschappelijk handelen	Begrip van de bredere maatschappelijke, juridische en gezondheidszorgcontext bij deze activiteit	4	5	6	7	8	9	10	
Professionaliteit	Integriteit en ethische opstelling bij deze activiteit	4	5	6	7	8	9	10	

schrijving op het formulier. Dat leidt tot een niet makkelijk leesbare uitleg in kleine letters die in de praktijk van de tien minuten observatie toch niet gelezen wordt. Heldere trefwoorden zijn daarom aan te raden. Als een uitleg gewenst is, is training met aanvullende uitleg een goede optie.

Zoals eerder is aangegeven moet de KPB primair als feedbackinstrument worden gebruikt. Ongeïnformeerde of niet getrainde beoordelaars neigen nogal eens de KPB als examen te zien. Dat is een verkeerde toepassing. Een ander verschijnsel is dat aios KPB's verzamelen en een keus maken welke in het portfolio worden opgenomen. Zwakke KPB's zouden zo achterwege kunnen blijven. In beide gevallen is een open, eerlijk en veilig opleidingsklimaat het antwoord. Dan hoeft er geen angst te bestaan dat de KPB wordt misbruikt.

7. De beoordelaars

Opleiders, superviserende stafleden en afdelingshoofden kunnen bij een KPB als beoordelaar fungeren. Belangrijk is dat een aios door verschillende beoordelaars wordt beoordeeld. Clinici verschillen onderling en juist dat kan leerzaam zijn voor de aios.

Training van beoordelaars vergroot de betrouwbaarheid van het beoordelen.

Ook kan een ouderejaars aios een KPB uitvoeren. De MSRC zal een aios mogelijk niet als beoordelaar erkennen, maar wel als de beoordeling gesuperviseerd wordt. De aios kan dan oefenen in het intercollegiaal beoordelen en feedback geven. Dat past bij het competentiegebied 'kennis en wetenschap'. Intercollegiale beoordeling binnen de beroepsgroep is van groot belang en het is daarom een zinvolle gedachte dat aios hier in de opleiding al in getraind wordt. Voor specifieke vaardigheden kunnen ook niet-medici soms goede beoordelaars zijn, zoals ervaren verpleegkundigen of physician assistants.

8. Praktische aspecten

Wanneer in de opleiding

KPB's zijn geschikt in alle opleidingsjaren in de vervolgopleiding en ook in de klinische fase van de medische basisopleiding. De situaties waarin aios worden beoordeeld kunnen echter verschillen. Larsen et al. pleiten er in de chirurgische opleiding voor om later in de opleiding toenemend complexe situaties voor observatie op te

nemen,³⁸ een advies dat ook voor andere specialismen waardevol lijkt.

Tijdsduur

De tijd die nodig is om te observeren is in het algemeen afhankelijk van de situatie die geobserveerd wordt en van het doel dat men voor ogen heeft. De adviestijd voor een KPB observatie is tien tot vijftien minuten. Voor de feedback kan vijf tot tien minuten gerekend worden. Onderzoek laat echter zien dat, ondanks de uitleg die gegeven wordt met betrekking tot de tijd, de gemiddelde beoordeling langer duurt. Bij onderzoek (n=388) kwam het gemiddelde op 31.5 minuten.⁴ Vooral als niet de nadruk wordt gelegd op de tijdsbeperking van een beoordeling, blijkt hier meer tijd aan op te gaan. In ander onderzoek, waarbij was aangegeven dat de beoordelingen in totaal tussen de vijftien en twintig minuten mochten duren, werd voor de observatie een gemiddelde van achttien minuten gevonden (SD 12.1 min.) en voor het feedbackgesprek gemiddeld 7.6 minuten (SD 5.3 min).⁴⁵

Sommige methoden hebben bewust een langere observatietijd. De Ophthalmic Clinical Evaluation Exercise bij een nieuwe patiënt duurt dertig tot vijfenveertig minuten²⁹⁻³⁰ en de OSATS soms zo lang als de uitgevoerde operatie. Voor de 'Palliative Care Clinical Evaluation Exercise' wordt gemiddeld 49.5 minuten uitgetrokken (SD 24.1).⁵³ Een kwart van de tijd is voor de voorbespreking tussen aios en opleider, de helft voor de observatie en een kwart voor de feedback. Bij KPB's waarbij de observator van nature aanwezig is, zoals bij het ochtendrapport en de overdracht, en bij operaties, kan de observatie langer zijn dan de gedefinieerde vijf tot vijftien minuten. Er zijn ook KPB's van korte duur. Een voorbeeld is de Minicard procedure waarin vier tot acht minuten geobserveerd wordt.²⁶

De langere observatietijd lijkt vaak te zijn ingegeven door de angst niet volledig of objectief te zijn bij de beoordeling. Het is echter beter om meer korte observaties uit te voeren dan een enkele lange observatie. Het lijkt soms tegen-intuïtief voor opleiders dat een reeks korte subjectieve beoordelingen bijdraagt aan een objectief oordeel. Toch is dat meer het geval dan een lange intensieve en gedetailleerde observatie.

Er worden vaak hoge intercorrelaties voor items op het beoordelingsformulier gevonden.^{43 54} Omdat er kennelijk een halo effect bestaat wordt een kort beoordelingsformulier aangeraden. Beoordelaars maken weinig onderscheid tussen de verschillende competenties bij de beoordeling en lijken het verspillen van tijd en moeite te vinden om een lange vragenlijst in te vullen.⁴² Het moment is vaak kort en de aios is in de regel het meest gebaat bij een beperkt aantal tips voor ontwikkeling.

Uitvoering van de observatie

Irby en Wilkerson noemen drie simpele uitvoeringsvoorwaarden voor succesvolle korte observaties in het klinisch onderwijs: 1) de aios moet weten met welk doel de observatie plaatsvindt en de procedure kennen, 2) de aios moet de patiënt uitleggen waarom een staflid komt meekijken en 3) het staflid moet buiten het gezichtsveld van de patiënt blijven.⁵⁵ Het staflid moet min of meer geruisloos kunnen komen en gaan, zonder het contact te verstoren. Dat neemt niet weg dat ook de patiënt geïnformeerd moet worden en moet kunnen weigeren.⁵⁶ De beoordelaar moet voorts voorbereid zijn op de focus van deze observatie, zorgen dat hij of zij niet gestoord kan worden en moet zelf ook in geen geval het contact tussen aios en patiënt interromperen.¹⁵

Toenemend wordt gebruik gemaakt van video-observatie voor KPB's. Dit heeft verscheidene voordelen:

- de beoordelaar kan op een voor hem of haar geschikt moment de beoordeling uitvoeren,
- de video kan teruggespoeld worden om de beoordeling te valideren,
- de feedback wordt veel krachtiger als de aios zichzelf kan zien,
- de aios kan ook voorafgaand aan de feedbackfragmenten selecteren voor de bespreking, bijvoorbeeld een waarover hij of zij tevreden is en een waarover hij of zij minder tevreden is. Dit is op zich al een reflectie- en leermoment.

Behalve dat er uiteraard geschikte apparatuur moet zijn is het wel van belang te voorzien in een standaard toestemmingsprocedure en belofte tot wisselen. Hoewel met moderne webcamtechniek veel van de oorspronkelijke bezwaren tegen video-opnames zijn opgeheven, zijn er wel enkele aandachtspunten. De privacy van de geobserveerden moet bewaakt worden, de KPB kan soms niet direct afgehandeld worden en er moet voor een bespreking gezocht ('gespoeld') worden naar een geschikt fragment, hetgeen enerzijds leerzaam is, maar anderzijds ook tijdrovend. Als poliruimtes met goed ingestelde vaste camera's worden ingericht, er in een standaard toestemmingsprocedure en belofte tot wisselen is voorzien en digitale video via een netwerk op de computer van de supervisor kan worden bekeken, wegen de voordelen zeker tegen de nadelen op.

Uitvoering van de feedback

Het geven van feedback is een vaardigheid die eenvoudig lijkt maar vaak lastig is.⁵⁷ Feedback in het klinisch onderwijs is niet hetzelfde als het geven van commentaar, omdat bij feedback met een opleidings-

doel het effect ervan op de verbetering van vaardigheden en gedrag voorop staat. Zonder dit doel zou eigenlijk niet van feedback gesproken moeten worden.⁵⁸

Dat in het algemeen negatieve feedback lastig is om te geven, beschrijven ook Norcini en Burch.¹⁶ Een valkuil voor de beoordelaar die de aios kent is dat men in plaats van louter de geobserveerde verrichting of handeling te beoordelen, men ook de intenties van de aios gaat interpreteren. Dit is niet de bedoeling. De beoordelaar zou primair feedback moeten geven over wat hij of zij waarneemt. Goede feedback is specifiek, persoonlijk en wordt gegeven in een 'veilige' setting.

Er zijn diverse specifieke methoden om feedback te geven. Bekend is de methode volgens de regels van Pendleton,⁵⁹ waarin achtereenvolgens aan de aios wordt gevraagd wat er goed ging, waarna de observator aangeeft in welk opzicht de handeling goed werd uitgevoerd. Daarna volgt de vraag wat de aios een volgende keer beter of anders zou willen doen en hoe, waarna als afsluiting de verbeterpunten van de observator volgen. Starten met positieve observaties is belangrijk om de aios te tonen dat het gaat om een positief te labelen interventie. Het streven is dat de aios graag wil gaan horen wat er beter kan. Holmboe et al. vonden in een onderzoek van 107 feedbacksessies dat 80% hiervan een duidelijke aanbeveling voor de aios bevatte.²⁴ In 61% van de gevallen vroeg de beoordelaar aan de aios om te reageren op de feedback, en in maar 34% vroeg de beoordelaar naar een zelfbeoordeling van de aios. Een actieplan werd na de beoordeling in slechts 8% van de gevallen door aios en beoordelaar opgesteld. De onderzoekers concludeerden dan ook dat er in de praktijk te weinig gebruik wordt gemaakt van het geven van goede feedback in al zijn verschillende facetten.

Het is verstandig om feedback direct na de observatie te laten plaatsvinden. Daarmee wordt voorkomen dat aspecten van de observatie vergeten worden en daarmee de feedback minder accuraat is.⁴²

9. Het formulier

Er zijn veel typen KPB formulieren in omloop. Dit is mede afhankelijk van het vakgebied waarin ze gebruikt worden. Zeker voor disciplines als radiologie en medische microbiologie zijn aanpassingen nodig. In het laatste vak is bijvoorbeeld het telefonisch consult een kritische activiteit waarvoor een KPB zeer zinvol is maar uiteraard een ander formulier vergt.

Soms zijn er sterke voorkeuren van individuele opleiders, bijvoorbeeld over de vraag of er veel of weinig ruimte voor opmerkingen moet zijn, en zelfs of er überhaupt scores moeten worden gegeven. Er is eigenlijk weinig evidentie voor één type formulier, maar een fundamentele vraag is wel of men scores van aios wil vergelijken over meerdere observatiemomenten, bijvoorbeeld om betrouwbaarheden te berekenen of om trends te signaleren. Dan zijn scores op een uniform formulier zeker nuttig.

De literatuur geeft voorbeelden van formulieren^{16 60-62} maar er zijn zeker veel meer varianten in praktijk. Bij de vraag wat een 'goed' formulier is spelen behalve psychometrische kwaliteiten, zoals de te gebruiken schaal, ook praktische zaken mee. Het formulier moet aantrekkelijk zijn om snel in te vullen. De hoeveelheid tekst moet beperkt en 'to the point' zijn. Trefwoorden voor items moeten direct begrijpelijk zijn, de uitnodiging tot het schrijven van punten van feedback moet dwingend maar vriendelijk zijn. Zoals mondelinge feedback volgens Pendleton positieve en verbeterpunten bevat, kan het verstandig zijn aparte omkaderde

ruimtes in het formulier op te nemen met als opdracht 'noem een of twee sterke punten' respectievelijk 'noem een of twee punten voor verdere ontwikkeling'. Formuleringen kunnen soms weerstand oproepen. Kies liever niet voor: 'wat ging goed' en 'wat ging slecht'. Zelfs het woord 'verbeterpunten' kan minder uitnodigen tot tips dan 'punten voor verdere ontwikkeling' omdat deze laatste formulering niet iets negatiefs hoeft te impliceren. Dat maakt het makkelijker om bij goede aios toch een tip te geven. Laat het formulier bij voorkeur niet anoniem invullen, maar ondertekenen en dateren door de observator.

Uiteraard is het niet verstandig als iedereen die KPB's afneemt eigen formulieren ontwerpt. Dat beperkt de vergelijkbaarheid over locaties en disciplines. Er zijn momenteel veel verschillen tussen disciplines. Het is in ieder geval wenselijk om binnen een opleidingscluster gelijke formulieren te gebruiken. In Nederland wordt de KPB nog pas enkele jaren gebruikt en het is zeer wel denkbaar dat er na verloop van tijd een zekere uniformering zal optreden.

Alhoewel de KPB in de regel nog een papieren formulier is, zien we dat ze steeds vaker digitaal wordt uitgevoerd, als onderdeel van een elektronisch portfolio.⁶³ Voorwaarde is wel dat er een PC aanwezig moet zijn in de ruimte waarin geobserveerd wordt. KPB methoden met behulp van mobiele elektronica, zoals een PDA of een geïntegreerde mobiele telefoon met zakcomputer, waarbij men niet meer afhankelijk is van de aanwezigheid van een PC, zijn in opkomst.⁶⁴ De elektronische uitvoering betekent niet alleen minder papieren rompslomp, maar ook de mogelijkheid om gegevens van achtereen volgende KPB's met elkaar te combineren en zicht te krijgen in competentieontwikkeling.

Checklists en schalen

Ringsted et al. vonden dat een gedragsgebaseerde checklist beter werkt dan globale beoordelingen. Zij onderzochten dit bij aios anesthesie.³⁵ In 25 opleidingsziekenhuizen in Denemarken beoordeelden 32 beoordelaars vier verschillende video's van verschillende handelingen. De helft beoordeelde de video's met een checklist, de andere helft moest drie globale items scoren. De auteurs concludeerden 1) dat beoordelaars de voorkeur geven aan een checklist, 2) dat zij hogere beoordelingen geven met de checklist dan met het globale beoordelingsformulier, 3) dat fouten van de aios op de checklist vaker teruggezien werden dan bij het globale beoordelingsformulier en 4) dat de interbeoordelaarsovereenstemming gering is, ongeacht het type beoordelingsformulier. Beoordelaars blijven mild, of ze fouten van aios nu wel of niet opmerken. Ook andere auteurs menen dat checklists zorgen dat beoordelaars eerder op specifieke punten gaan letten, waarbij fouten eerder gesignaleerd worden.^{16 23 42} Dit zou, naast meer passende feedback, voor een hogere interbeoordelaarbetrouwbaarheid zorgen.

Iedere belangrijke variant op de KPB hanteert een eigen schaalverdeling. De mini-CEX heeft van origine een negenpuntsschaal. In de meeste gevallen wordt hierbij de volgende onderverdeling gemaakt: 1-3 = onvoldoende, 4-6 = voldoende en 7-9 = goed. Soms wordt er nog een extra onderverdeling gemaakt waarbij 4 marginaal betekent en 5-6 voldoende. Bij elk item staat ook het kopje 'insufficiënt contact to judge' of 'not observed'. Het is evident dat een negenpuntsschaal, met '5' als laagste voldoende, niet past in de Nederlandse context, omdat iedereen is opgegroeid met tienpuntsschalen.

De P-mex en de OCEX hanteren een vierpuntsschaal (1 = does not meet expectations, 2 = meets some expectations, 3 =

meets all expectations, 4 = exceeds expectations) plus 'not applicable'.²⁸⁻²⁹ Ringsteds globale schaal gebruikt vijf waarden: 'clear fail', 'borderline fail', 'borderline pass', 'clear pass', 'excellent' en een aanduiding of de aios gezakt of geslaagd is.³⁵ Soms worden op een vijfpuntsschaal alleen labels gegeven aan 1, 3 en 5 zoals bij de GRASIS, de OSA-LT en de OSATS.^{2 34 38}

De Minicard kent vier schaalpunten: excellent, goed, marginaal, slecht. De Minicard en de Mini-CEX zijn wel vergeleken op hun onderscheidend vermogen naar onvoldoende en voldoende 'handelingen' (als anamnese of lichamelijk onderzoek). De Minicard bleek beter te discrimineren. Bij gebruik in de praktijk bleek dat de Minicard tot strengere oordelen leidde.²⁶

De ITER kent meerdere schalen, waaronder 14 items en een tienpuntsschaal, onderverdeeld in vijf omschreven beoordelingen: 'unacceptabel', 'insufficient', 'marginal', 'good' en 'excellent'.²⁷ Op een apart blad staat beschreven wat voor elk item als 'onacceptabel' en wat als 'excellent' wordt gezien. Veel ITER varianten kennen echter drie waarden: 'below expectation', 'meets expectation' en 'exceeds expectation'.⁶⁵

Een andere benadering is om voor items verschillende wegen toe te kennen.³⁷ Het correct plaatsen van een fundusklem bij laparoscopie is twee punten waard terwijl andere handelingen acht punten opleveren. Voor fouten worden verschillende aantallen punten afgetrokken: leverschade zonder bloeding is min één punt, met bloeding min vijf punten.

Cook et al. vonden dat bij de mini-CEX met een negenpuntsschaal een beter onderscheid te maken viel tussen onvoldoende, voldoende en goed dan met een vijfpuntsschaal, waarschijnlijk omdat beoordelaars extremen op de schaal graag vermijden.⁴⁵ Williams et al. raden in hun review om die reden een schaal aan van

minimaal vijf, zes of zeven punten.⁴² Een schaal waarbij naast een getal ook onderscheid gemaakt moet worden tussen een onvoldoende, voldoende of goed, of iets soortgelijks, raden zij af. Extra categorieën zouden de beoordeling lastiger maken. Wat wel gehanteerd wordt is drie categorieën waarbinnen steeds twee of drie scores vallen. Een schaal die loopt van 4 tot 10 kan gehanteerd worden in de Nederlandse context. Lager dan 4 zal niet worden gescoord. In een dergelijke schaal is 7 de middenwaarde die ook als een beginselstandaard van 'adequaat functioneren' gehanteerd kan worden. Scores 4 en 5 kunnen worden beschouwd als lager dan verwacht; 9 en 10 als hoger dan verwacht. Waarden 6, 7 en 8 vallen dan alle binnen het 'verwachte' niveau. Een dergelijke schaalverdeling kan aangeraden worden.⁶⁵

Normering

Er zijn in principe vier benaderingen van de vraag waarmee bij de invulling van het KPB formulier de vaardigheid van de aios moet worden vergeleken: 1) een score ten opzichte van een beoogd eindresultaat, 2) een score ten opzichte van het opleidingsstadium waarin de aios zich bevindt, 3) een score ten opzichte van een peer-groep van andere aios en 4) een score ten opzichte van de eigen ontwikkeling.

Indien geen afspraken zijn gemaakt over normering hanteren beoordelaars soms verschillende normen, die niet expliciet zijn. Op het eerste gezicht lijkt een normering ten opzichte van het gewenste eindresultaat, dat wil zeggen een 'absolute' normering, logisch. Men wil immers laten zien hoe de aios scoort in vergelijking met een standaard en men verwacht ook dat er een stijgende lijn te zien zal zijn. In de praktijk echter valt het beoordelaars moeilijk om consequent eerstejaars aios onvoldoende scores geven. Men wil, terecht, graag inspanning belonen en de aios waarderen ten

opzichte van het verwachte niveau. Om deze reden zijn de normen 'below', 'meets', en 'exceeds expectation', c.q. onder verwacht niveau, op verwacht niveau en boven verwacht niveau, met andere woorden een relatieve normering, vaak een beter hanteerbare normering.⁶⁵ Relatief wil dan zeggen ten opzichte van wat men in dit opleidingsstadium verwacht. Dat is een goed verdedigbare benadering. Tenslotte wil men soms ook aios belonen voor de individuele vooruitgang die gemaakt is, ongeacht het werkelijke niveau. Daar moet voorzigtiger mee worden omgegaan, omdat het vergelijken van aios op die manier lastig wordt. In zo'n geval is het verstandig om de waardering voor individuele vooruitgang vooral in de opmerkingen naar voren te laten komen en minder in de scores.

10. Relatie tot portfolio, voortgangsgesprek en bekwaamverklaringen

De KPB wordt gedocumenteerd omdat er een overzicht moet ontstaan van het functioneren van de aios door de tijd. Het kader waarin dat moet plaatsvinden is het portfolio. Daarin vormt de set van KPB's, minimaal tien per jaar, een van de centrale onderdelen. De Richtlijn "Interpretatie en gebruik van toetsresultaten in het portfolio" gaat hier nader op in.⁶⁶ Bij de voortgangsgesprekken vormen de KPB's vervolgens essentiële achtergrondinformatie. Vooral trends over de KPB's heen zijn de moeite waard om te bepreken. Ook over voortgangsgesprekken zal een richtlijn verschijnen.

De KPB kan bijdragen aan de informatie waarop een bekwaamverklaring voor een specifieke professionele activiteit kan worden afgegeven. Er zijn voorbeelden in de literatuur waar dit met een variant van de KPB geschiedt. Larson et al. beschrijven hoe voor verschillende operaties een 'Operative Performance Rating System' (OPRS) kan worden gehanteerd waarvoor

specifieke beoordelingsformulieren bestaan.³⁸ De OPRS is niet alleen bedoeld om de aios chirurgie te evalueren, maar ook om deze te certificeren. Voor certificering dient elke procedure tenminste drie maal geobserveerd te worden door tenminste twee verschillende beoordelaars. De aios die drie keer achtereenvolgend de cijfers 4 en 5 (op een vijfpuntsschaal) heeft gekregen, krijgt een bekwaamverklaring. Achtergronden van de bekwaamverklaring zijn voor de Nederlandse context beschreven in de Richtlijn Bekwaamverklaringen in de specialistenopleiding.⁶⁷

11. Invoering van korte praktijkbeoordelingen als systeem en training van klinische staf

Norcini en Burch noemen drie vereisten bij de invoering van KPB's: a) een goede planning van observatiemomenten en feedbackgesprekken, b) een goede informatievoorziening aan stafleden en aios over doel en werkwijze en c) een voortdurende herinnering aan het systeem en de waarde van formatieve beoordeling om verwatering te voorkomen.¹⁶

Planning van KPB's is een kwetsbaar punt. Hoe minder sturing hierin optreedt, hoe lastiger het wordt om aan de eisen te voldoen. Tijd is vaak een probleem en soms een excuus voor het feit dat KPB's niet worden uitgevoerd.⁴² Er zijn poliklinische afdelingen waarin eens per week een staflid een uur wordt vrijgemaakt om achtereenvolgens drie of vier KPB's te doen bij verschillende aios op de poli. Ook aios moeten gewend raken aan regelmatige observaties. Er zijn indicaties dat het leereffect pas ontstaat als deze gewenning de angst wegneemt die ontstaat bij incidentele aanwezigheid van een senior staflid met een scoringsformulier.⁶⁸

Informatievoorziening, en in het bijzonder training van stafleden, is nodig om vertrouwd te raken met de KPB als me-

thode, met de normering die gehanteerd wordt en met het voeren van korte feedbackgesprekken.²⁴ In alle Nederlandse Onderwijs- en Opleidingsregio's (OOR's) worden dergelijke trainingen aangeboden. Holmboe et al. vonden dat getrainde stafleden strenger beoordeelden dan ongetrainde stafleden, ongeacht de accurate score. Getrainde beoordelaars gaven aios die onvoldoende vaardigheden bezaten, ook eerder daadwerkelijk een onvoldoende score. Daarnaast vond men dat stafleden zich meer op hun gemak voelden om te beoordelen indien zij getraind zijn.²⁴ Anderen rapporteren echter minder effect van training,^{42 69} of voornamelijk een effect op zelfvertrouwen in het uitvoeren van de beoordeling.⁶⁹ Een 'frame of reference' training wordt aangeraden voor het leren uitvoeren van KPB's.¹⁵ Deze training omvat zeven stappen: 1) bespreking van, en het bereiken van consensus over de aspecten van competentie die belangrijk zijn om te observeren, 2) definiëren wat als optimaal gedrag beschouwd kan worden, 3) definiëren wat als net voldoende gedrag beschouwd kan worden, 4) observeren van een klinisch contact, bij voorkeur via video, 5) beoordelen van het contact op een schaal met ankerwaarden, 6) vergelijken en bespreken van de gegeven oordelen, 7) diepgaande bespreking van discrepanties in oordelen.

12. Evaluatie en herziening van deze Richtlijn

Deze Richtlijn weerspiegelt een dynamische praktijk van beoordelingen en feedback in de klinische opleidingen. Volgens het Kaderbesluit CCMS 2009 moet in 2011 de KPB in alle opleidingen standaard zijn ingevoerd. Dat impliceert dat er sprake is van een zeer grootschalige ontwikkeling die ongetwijfeld in de komende jaren tot nieuwe inzichten zal leiden. Daarnaast staan ook internationaal

het onderzoek en de ontwikkeling niet stil. Het is een van de zich het sterkst ontwikkelende terreinen van de medische onderwijskunde. Daarom wordt aanbevolen deze opleidingsrichtlijn over circa vijf jaar opnieuw tegen het licht te houden.

Literatuur

- Norcini JJ. Workplace-based assessment in clinical training. ASME series Understanding Medical Education, Association of the Study of Medical Education, Edinburgh, 2007.
- Hiemstra E, Jansen FW. Opleidingsrichtlijn OSATS. Tijdschrift voor Medisch Onderwijs 2010;29(5 Suppl 4):S135-143.
- Horsman MA & Ten Cate ThJ. Richtlijn Multi-source feedback. Tijdschrift voor Medisch Onderwijs 2010;29(2 Suppl 1):S3-51.
- Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (Clinical Evaluation Exercise): A preliminary Investigation. Ann Intern Med 1995;123(10):795-799.
- Guide to Evaluation of Residents in Internal Medicine – a System approach. American Board of Internal Medicine 1994.
- Elstein AS, Shulman LS, Sprafka SA. Medical problem-solving. J Med Educ 1981;56(1):75-76.
- Noel GL, Herbers JE, Jr., Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents? Ann Intern Med 1992;117(9):757-765.
- Kroboth FJ, Hanusa BH, Parker S, Coulehan JL, Kapoor WN, Brown FH et al. The inter-rater reliability and internal consistency of a clinical evaluation exercise. J Gen Intern Med 1992;7(2):174-179.
- Woolliscroft JO, Stross JK, Silva J Jr. Clinical competence certification: a critical appraisal. J Med Educ 1984; 59(10):799-805.
- Miller GE. The assessment of clinical skills/competence/performance. Acad Med. 1990; 65(9 Suppl):S63-7.
- Ten Cate ThJ, Bleker OP, Büller HA, Scherpbier AJJA. Opleiden van medisch specialisten – achtergronden en praktijk. ISBN 9031342851 Houten: Bohn Stafleu Van Loghum; 2003, p206-209.
- Centraal College Medische Specialismen. Kaderbesluit 2004. Van kracht per 1 januari 2005. KNMG, 2004.
- CMS protocol toetsing en beoordeling in de medisch specialistische opleidingen.
- Centraal College Medische Specialismen. Kaderbesluit 2009. Van kracht per 1 januari 2011. KNMG, 2009.
- Holmboe ES, Hawkins RE (eds). Practical guide to the evaluation of clinical competence. Mosby: Philadelphia; 2008.
- Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach 2007; 29(9):855-871.
- Schuwirth LW, Vleuten CP van der. Competentie-gerichte toetsing in het klinisch onderwijs. Ned Tijdschr Geneesk 2005;149(49):2752-2755.
- Alves de Lima A, Henquin R, Thierer J, Paulin J, Lamari S, Belcastro F, Vleuten CP van der. A qualitative study of the impact on learning of the mini clinical evaluation exercise in postgraduate training. Med Teach 2005; 27(1):46-52.
- Alves de Lima A, Barrero C, Baratta S, Castillo Costa Y, Bortman G, Carabajales J, Conde D, Galli A, Degrange G, Vleuten C van der. Validity, reliability, feasibility and satisfaction of the Mini-Clinical Evaluation Exercise (Mini-CEX) for cardiology residency training. Med Teach 2007; 29(8):785-790.
- Schuwirth LW, Vleuten CP van der. A plea for new psychometric models in educational assessment. Med Educ 2006;40(4):296-300.
- Day SC, Grosso LJ, Norcini JJ, Jr., Blank LL, Swanson DB, Horne MH. Residents' perception of evaluation procedures used by their training program. J Gen Intern Med 1990;5(5):421-426.
- Ten Cate O. Entrustability of professional activities and competency-based training. Med Educ 2005;39:1176-1177.
- Durning SJ, Cation LJ, Markert RJ, Pangaro LN. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med 2002;77(9):900-904.
- Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. Ann Intern Med 2004;140(11):874-881.
- Burr WA. Competence-based assessment for higher specialist training in endocrinology and diabetes mellitus. Diabet Med 2004; 21 Suppl 3:13-15.
- Donato AA, Pangaro L, Smith C, Rencic J, Diaz Y, Mensinger J, Holmboe E. Evaluation of a novel assessment form for observing medical residents: a randomised, controlled trial. Med Educ 2008;42(12):1234-1242.
- Skakun EN, Wilson DR, Taylor WC, Langley GR. A preliminary examination of the In-Training Evaluation Report. J Med Educ 1975;50(8):817-819.
- Cruess R, Herold McIlroy J, Cruess S, Ginsburg S, Steinert Y. The professionalism Mini-evaluation exercise: a preliminary investigation. Acad Med 2006;81(10):S74-S78.
- Golnik KC, Goldenhar LM, Gittinger JW Jr, Lustbader JM. The Ophthalmic Clinical Evaluation Exercise (OCEX). Ophthalmology 2004; 111(7):1271-1274.

30. Golnik KC, Goldenhar L. The ophthalmic clinical evaluation exercise: reliability determination. *Ophthalmology* 2005; 112(10):1649-1654.
31. Scheele F en Schutte MF red. Curriculum opleiding tot gynaecoloog NVOG-HOOG. Utrecht: NVOG; 2005.
32. Reznick, R., et al., Testing technical skill via an innovative "Bench Examination". *Am J Surg*, 1997;173: p. 226-230.
33. Strating WJ, Van Acker GJD, Brand PLP, Pierik EGJM. Beoordeling van klinische en chirurgische vaardigheden: meer overeenstemming tussen beoordelaars bij gebruik van de 'Korte klinische beoordeling' (KKB) dan bij de 'Objective structured assessment of technical skills' (OSATS). *Ned Tijdschr Geneesk* 2008;152:B9.
34. Cremers SL, Lora AN, Ferrufino-Ponce ZK. Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). *Ophthalmology* 2005; 112(10):1655-1660.
35. Ringsted C, Østergaard D, Ravn L, Pedersen JA, Berlac PA, Vleuten CP van der. A feasibility study comparing checklist and global rating forms to assess resident performance in clinical skills. *Med Teach* 2003;25(6):654-658.
36. Griffiths CE. Competency assessment of dermatology trainees in the UK. *Clin Exp Dermatol* 2004; 29(5):571-575.
37. Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S, Pellegrini C. An objective scoring system for laparoscopic cholecystectomy. *J Am Coll Surg* 1999;189(6):566-574.
38. Larsen CR, Grantcharov T, Schouenborg L, Ottosen C, Soerensen JL, Ottesen B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG*. 2008;115(7):908-916.
39. Mulder H, Ten Cate O, Daalder R, Berkvens J. Building a competency-based workplace curriculum around entrustable professional activities: the case of physician assistant training (submitted).
40. Epstein RM. Assessment in Medical Education. *N Engl J Med* 2007;356:387-96.
41. Holmboe ES, Huot S, Chung J, Norcini J, Hawkins RE. Construct validity of the miniclinical evaluation exercise (miniCEX). *Acad Med* 2003; 78(8):826-830.
42. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med* 2003; 15(4):270-292.
43. Margolis MJ, Clauser BE, Cuddy MM, Ciccone A, Mee J, Harik P, Hawkins RE. Use of the mini-clinical evaluation exercise to rate examinee performance on a multiple-station clinical skills examination: a validity study. *Acad Med* 2006; 81(10 Suppl):S56-60.
44. Norcini JJ, Blank LL, Arnold GK, Kimball HR. Examiner Differences in the Mini-CEX. *Adv Health Sci Educ Theory Pract* 1997; 2(1):27-33.
45. Norcini JJ, Blank LL, Duffy F, Fortna G.S. The Mini-CEX: A method for Assessing Clinical Skills. *Ann Intern Med* 2003; 138:476-481.
46. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309-3170.
47. Govaerts MJB, Van der Vleuten CPM, Schuwirth LWT, Muijtens AMM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ* 2007;12:239-260.
48. Ten Cate O, Snell L, Carraccio C. Medical competence: The interplay between individual ability and the health care environment. *Med Teach* 2010;32(8):669-675.
49. Albanese M. Rating educational quality: factors in the erosion of professional standards. *Acad Med* 1999;74(6):652-658.
50. Albanese MA. Challenges in using rater judgments in medical education. *J Eval Clin Pract* 2000; 6(3):305-319.
51. Hatala R, Ainslie M, Kassen BO, Mackie I, Roberts JM. Assessing the mini-Clinical Evaluation Exercise in comparison to a national specialty examination. *Med Educ* 2006;40(10):950-956.
52. Begeleidingcommissie Beschrijving Opleidingsplannen Vervolgopleidingen (BBOV), december 2008.
53. Han PK, Keranen LB, Lescisin DA, Arnold RM. The Palliative Care Clinical Evaluation Exercise (CEX): An Experience-Based Intervention for Teaching End-of-Life Communication Skills. *Acad Med* 2005; 80(7):669-676.
54. Cook DA, Beckman TJ. Does scale length matter? A comparison of nine- versus fivepoint rating scales for the mini-CEX. *Adv Health Sci Educ Theory Pract* 2008 Nov 26. [Epub ahead of print].
55. Irby DM, Wilkerson LA. Teaching when time is limited. *BMJ* 2008;336:384-387.
56. Burr WA. Competence-based assessment for higher specialist training in endocrinology and diabetes mellitus. *Diabet Med* 2004; 21 Suppl 3:13-15.
57. Carr S. The Foundation Programme assessment tools: an opportunity to enhance feedback to trainees? *Postgrad Med J* 2006; 82(971):576-579.
58. Van de Ridder JJM, Stokking KM, McGaghie WC, Ten Cate OThJ. What is Feedback in Clinical Education? *Med Educ* 2008;2(42):189-197.
59. Pendleton D, Schofield T, Tate P. A method for giving feedback. In: The consultation: an approach to learning and teaching. Oxford: Oxford University Press, 1984:68-71.

60. Hill F, Kendall K. Adopting and adapting the mini-CEX as an undergraduate assessment and learning tool. *The Clinical Teacher* 2007;4:244-248.
61. Caldwell G. Real-time assessment and feedback of junior doctors improves clinical performance. *The Clinical Teacher* 2006;3:185-188.
62. Govaerts M. In-training assessment: learning from practice. *The Clinical Teacher* 2006; 3: 242-247.
63. Borel Rinkes IH, Gouma DJ, Hamming JF. Surgical training in The Netherlands. *World Journal of Surgery* 2008;32 (10):2172-2177.
64. Van Schoor AN, Navsa N, Meiring JH, Treadwell I, Bosman MC, Greyling LM. Perspectives on the use of PDAs as assessment tools. *The Clinical Teacher* 2006;3:170-174.
65. Ten Cate OTHJ, Ter Braak EWTM, Frenkel J, Van de Pol A. De 4-tot-10-verwacht niveau schaal bij persoonlijke beoordelingen. *Tijdschrift voor Medisch Onderwijs* 2006;25(4):157-163.
66. Govaerts MJB, Driessen E, Verhoeven B, Van der Vleuten CPM, Brackel H, Van Hoorn J. Van de Laar R, Maas J, Oewi SG. Richtlijn Interpretatie en gebruik van toetsresultaten in het portfolio. *Tijdschrift voor Medisch Onderwijs* 2010;29(3 Suppl 3):S73-102.
67. Lips J, Scheele, F, Dijksterhuis M, Ten Cate O. Richtlijn bekwaamverklaringen in de specialistenopleiding. *Tijdschrift voor Medisch Onderwijs* 2010;29(2 Suppl 2):S53-72.
68. Malhorta S, Hatala R, Courneya CA. Internal medicine residents' perceptions of the Mini-Clinical Evaluation Exercise. *Med Teach* 2008;30(4):414-419.
69. Cook DA, Dupras DM, Beckman TJ, Thoman KG, Pankras VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med* 2009;24(1):74-79.

De auteurs:

Prof. dr. Th.J. ten Cate is hoogleraar medische onderwijskunde en directeur Expertisecentrum voor Onderwijs en Opleiding, UMC Utrecht.

Drs. C.R.M.G. Fluit is arts/onderwijskundige en hoofd afdeling Evaluatie, Kwaliteit en Ontwikkeling, UMC st Radboud Nijmegen.

Correspondentieadres:

Prof. dr. Th.J. ten Cate, Expertisecentrum voor Onderwijs en Opleiding, UMC Utrecht, postbus 85500, 3508 GA Utrecht. E-mail: t.j.tencate@umcutrecht.nl

Belangenconflict: geen gemeld

Financiële ondersteuning: geen gemeld