Achtergronden bij de Richtlijn interpretatie en gebruik van toetsresultaten in het portfolio

1. Doelstelling van de Richtlijn en het gebruik door verschillende doelgroepen

Doel van deze Richtlijn is het ondersteunen van aios en opleider (-team) bij het bespreken van de voortgang c.q. bij het nemen van beslissingen over de voortgang in de opleiding, en bij het gebruik van toetsresultaten in dit proces. De Richtlijn beoogt dit te doen door middel van:

- Een beschrijving van de doelstellingen van toetsing in de medisch-specialistische vervolgopleiding en het gebruik van toetsresultaten in begeleidingsgesprekken (voortgangs- en beoordelingsgesprekken).
- Een beschrijving van kwaliteitseisen bij toetsing en bij gebruik van toetsresultaten (met nadruk op werkplekbeoordelingen).
- Een omschrijving van stappen in zorgvuldige besluitvorming.
- Een samenvatting van een aantal randvoorwaarden bij gebruik van toetsresultaten ten behoeve van besluitvorming.

Voor aios en opleider (-teams) worden in de Richtlijn de belangrijkste aandachtspunten bij het gebruik van toetsresultaten (werkplekbeoordelingen) in voortgangs- en beoordelingsgesprekken beschreven. De Richtlijn geeft een overzicht van de criteria aan de hand waarvan een inschatting gemaakt kan worden van de kwaliteit en bruikbaarheid van toetsresultaten, en waaraan de zorgvuldigheid in besluitvorming getoetst kan worden.

Voor opleiders en opleiderteams worden daarnaast een aantal richtlijnen beschre-

ven voor de inrichting van de toetsing op de werkplek en de procedures ten behoeve van besluitvorming.

Voor *opleidingsmanagers en kwaliteitszorgmedewerkers*, tenslotte, worden in de Richtlijn de belangrijkste randvoorwaarden voor een goed toetssysteem beschreven. De Richtlijn is daarmee te gebruiken als basis voor een interne visitatie naar de kwaliteit van het toetssysteem op een afdeling c.q. binnen een instelling.

Deze Richtlijn is gebaseerd op literatuuronderzoek en best practices, zowel binnen de context van het medisch opleidingscontinuüm als binnen andere contexten waar leren en beoordelen op de werkplek een belangrijke plaats innemen.

2. Definities en begripsbepalingen

CanMEDS: een 7-tal algemene competenties die de bekwaamheid van de Nederlandse medisch specialist beschrijven. Zij vormen de generieke eindtermen voor alle medisch-specialistische vervolgopleidingen in Nederland en zijn afgeleid van de door 'the Royal College of Physicians and Surgeons of Canada' geformuleerde 'CanMEDS-roles'. (<http://knmg.artsennet.nl/>).

Competentie: de bekwaamheid om op basis van kennis, vaardigheden, oordeelsvermogen, houdingen (opvattingen, visies) en persoonskenmerken, keuzes te maken uit een handelingsrepertoire die tot resultaat hebben dat de taken die zich in een specifieke beroepscontext voordoen, adequaat worden aangepakt, conform rol en verantwoordelijkheid.

Competentieprofiel: profielbeschrijving van de individuele aios die inzicht geeft in de mate van bekwaamheid voor de verschillende CanMEDS competenties.

Halo: bepaalde kenmerken van de beoordeelde of bepaalde aspecten in het functioneren worden overbelicht en hebben een onterechte (positieve of negatieve) invloed op de beoordeling van andere kenmerken of aspecten in het functioneren.

KBS: Kritische BeroepsSituatie / Kenmerkende BeroepsSituatie.

KPB: Korte PraktijkBeoordeling; vorm van werkplekbeoordeling. Kortdurende evaluatie van het functioneren van de aios op basis van directe observatie, bij het uitvoeren van een authentieke taak.

Leniency: neiging tot geven van milde oordelen; score-inflatie.

Opleidingsplan: een door de betreffende wetenschappelijke vereniging opgesteld en door het Centraal College Medische Specialismen (CCMS) vastgesteld plan dat de structuur en inhoud van de opleiding tot medisch specialist beschrijft en dat de basis vormt voor het individuele opleidingsplan.

OSATS: Objective Structured Assessment of Technical Skills; vorm van werkplekbeoordeling. Oorspronkelijk ontwikkeld ten behoeve van evaluatie van met name technische vaardigheden bij operatieve ingrepen.

POP: Persoonlijk OntwikkelPlan, ook wel Individueel OpleidingPlan (IOP). In het Persoonlijk Ontwikkelplan beschrijft de aios wat hij/zij wil bereiken gedurende een bepaalde periode (persoonlijk leerplan), aan de hand van SMART geformuleerde

leerde doelen. Een POP wordt opgesteld op basis van sterkte-zwakteanalyse, individuele leerwensen van de aios en van de mogelijkheden die in betreffende periode geboden worden. Een POP wordt gedurende de opleiding regelmatig geëvalueerd en bijgesteld.

Portfolio: persoonlijk dossier, verzamelinstrument al dan niet digitaal, dat informatie bevat op basis waarvan het leerproces van de aios inzichtelijk wordt gemaakt. Het portfolio in de medisch-specialistische vervolgopleiding vervult verschillende functies, met nadruk op ontwikkeling en reflectie. Daarnaast vormt de informatie in het portfolio 'bewijsmateriaal' bij het nemen van beslissingen ten aanzien van bekwaamheid.

SMART: acroniem ter aanduiding van de eisen die aan goed geformuleerde leerdoelen gesteld kunnen worden. Leerdoelen dienen liefst Specifiek, Meetbaar, Acceptabel, Realiseerbaar/Realistisch en Tijdgebonden te zijn.

Sterkte-zwakteanalyse: analyse van het eigen functioneren waarbij in kaart wordt gebracht waar de sterke punten liggen in het functioneren en waar verbetering noodzakelijk/wenselijk is.

Themaprofiel: profielbeschrijving van de individuele aios die inzicht geeft in de mate van bekwaamheid voor de verschillende in het opleidingsplan gedefinieerde thema's.

Toetsing: systematische en structurele evaluatie van en feedback op het functioneren van de aios, met formatieve en/of summatieve doeleinden.

Werkplekbeoordeling: beoordeling van het functioneren van de aios op de werkplek.

3. Doelstellingen van toetsing – algemene inleiding

A. Rol van toetsen in de opleiding

Toetsen kunnen binnen opleidingen een aantal functies vervullen¹⁻⁴:

1. Bevorderen van het leren en ondersteunen van effectieve en efficiënte ontwikkeling van (beroeps)competenties. Toetsing motiveert, bekrachtigt en stuurt de professionele ontwikkeling van de aios, door:
 - a. het expliciet maken van de eisen die op bepaalde momenten in de opleiding aan de aios gesteld worden,
 - b. feedback op functioneren, gerelateerd aan de doelstellingen van de opleiding: zichtbaar maken van sterke en minder sterke kanten in het functioneren.
2. Onderbouwen/verantwoording van selectie-/promotiebeslissingen: Toetsresultaten geven inzicht in het bereikte niveau van competentieontwikkeling en de kwaliteit van functioneren van de aios en vormen daarmee een belangrijke basis voor het toekennen van bekwaamheidsverklaringen, certificering en registratie.
3. Bewaken van opleidingskwaliteit:
 - a. toetsresultaten geven informatie over de effectiviteit van de opleiding,
 - b. bescherming van publiek tegen onbekwame professionals.

Er wordt onderscheid gemaakt tussen formatief en summatief gebruik van toetsresultaten.

Formatieve toetsing heeft met name ten doel de aios feedback te verschaffen over de voortgang in het eigen leerproces, op basis daarvan reflectie te stimuleren en leeractiviteiten te sturen. De formatieve functie van toetsen staat centraal in de regelmatig terugkerende, formele beoorde-

lingen van het functioneren op de werkplek in de vorm van KPB, OSATS of 360 graden feedback. Ook in voortgangsgesprekken ligt de nadruk op formatief gebruik van toetsresultaten.

Van *summatief* gebruik van toetsresultaten wordt gesproken als toetsresultaten de basis vormen voor het nemen van beslissingen over voortgang c.q. certificering. Summatief gebruik van toetsresultaten staat centraal in beoordelingsgesprekken. Beslissingen over bekwaamheid respectievelijk registratie/certificering worden altijd genomen op basis van informatie over het functioneren van de aios die gedurende *langere* tijd, min of meer *systematisch*, wordt verzameld, bijvoorbeeld in een portfolio. Een enkele werkplekbeoordeling (momentopname) vormt nooit de enige en unieke basis voor het nemen van summatieve beslissingen!

Zowel ten behoeve van voortgangsgesprekken als ten behoeve van beoordelingen moeten toetsresultaten een *accuraat* en *geloofwaardig* beeld geven van het *leerproces*, *competentieontwikkeling* en van de *bekwaamheden* van de aios. Dit kan alleen als de informatie op basis waarvan beslissingen worden genomen aan specifieke voorwaarden voldoet (zie hoofdstuk 4).

B. Toetsen in de medisch-specialistische vervolgopleiding: competentie- en themagericht toetsprogramma

Toetsresultaten geven inzicht in het functioneren van de aios. Zowel bij het formuleren van individuele ontwikkelingsplannen, als bij het nemen van gemotiveerde beslissingen over professionele bekwaamheid is het noodzakelijk om een gedetailleerd beeld te hebben van de ontwikkeling van de aios, gerelateerd aan de specifieke eindtermen van de opleiding. Een holistisch, algemeen beeld van het functioneren van de aios is daarbij meestal niet voldoende.

In het kader van de vernieuwing van de medisch-specialistische vervolgopleidingen zijn de eindtermen van de opleiding langs verschillende invalshoeken gedefinieerd. Enerzijds worden ten behoeve van de competentieontwikkeling van de aios een 7-tal algemene competentiedomeinen onderscheiden die vastgelegd zijn in de CanMEDS rollen:

1. Medisch handelen
2. Communicatie
3. Samenwerking
4. Organisatie
5. Professionaliteit
6. Maatschappelijk handelen
7. Kennis en wetenschap

Anderzijds zijn vanuit vakinhoudelijk perspectief specifieke (specialisme-gebonden) einddoelen beschreven in een aantal verschillende thema's en/of kenmerkende c.q. kritische beroepssituaties (taken, clusters van taken of verrichtingen).

De professionele ontwikkeling van de aios vindt enerzijds plaats binnen elk van de zeven CanMEDS competenties, anderzijds binnen de verschillende thema's van het vakgebied. Bij het bespreken c.q. beoordelen van de voortgang van de aios moet zo goed mogelijk in beeld gebracht worden hoe het staat met:

- de ontwikkeling van de aios in de verschillende CanMEDS competenties (*competentieprofiel*),
- de ervaringen en bekwaamheden van de aios binnen de verschillende thema's (*themaprofiel*).

Analoog hieraan richt toetsing van de aios zich dus enerzijds op de competentieontwikkeling (over de verschillende thema's heen), anderzijds op de themagebonden bekwaamheden (over de verschillende competentiedomeinen heen).

Uitspraken over competenties worden altijd afgeleid uit beoordelingen van het

functioneren op verschillende taken in de beroepsuitoefening. Een aios is immers alleen in staat om de relevante beroepstaken adequaat uit te oefenen indien hij beschikt over een aantal competenties.¹⁻⁵ Op deze manier ontstaat voor elk thema een matrix van relevante taken en competenties, zoals uitgewerkt in onderstaande voorbeelden (Thema A respectievelijk B).

Uit de gehanteerde definitie van professionele competentie en de opleidingsspecifieke thema-competentiematrices volgt dat in de medisch-specialistische vervolgopleiding een *complete* toetsprogramma minimaal omvat:

- A. Beoordelingen van performance op relevante taken uit de beroepspraktijk:
 - a. beoordeling van de kwaliteit van handelen (proces) en
 - b. beoordeling van de uitkomst van het handelen (= outcomes).
- B. Beoordeling van de verantwoording van gemaakte keuzes en van reflectie op handelen.
- C. Beoordeling van ervaring met specifieke taken uit het vak (registratie van ervaringen/verrichtingen).

A-C betreffen voornamelijk registratie en beoordeling van functioneren in de praktijk (werkplekbeoordelingen). Het leren van de aios in de medisch-specialistische vervolgopleiding vindt voor 90-95% plaats op de werkplek. Evaluaties c.q. beoordelingen van het functioneren op de werkplek vormen dan ook de kern van het toetsprogramma in de medisch-specialistische vervolgopleiding. Bij het doen van uitspraken over professionele competentie is vaak behoefte aan aanvullend 'bewijsmateriaal', omdat het onmogelijk is om alle facetten van het vak in de praktijk voldoende breed te toetsen.⁶⁻⁷ Aanvullend bewijsmateriaal kan verkregen worden op basis van:

- D. Beoordeling van gevolgde scholing/onderwijs/training.
- E. Beoordeling van kennis (kennistoetsing).
- F. Beoordeling van vaardigheden (met name relevant indien de praktijk te weinig mogelijkheden biedt om dit gedegen te doen; als beoordeling van vaardigheden in de praktijk lastig is in verband met weinig voorkomen óf als bewijs van beheersing noodzakelijk is in verband met patiëntveiligheid).

C. Portfolio als verzamelinstrument

Om te komen tot uitspraken over competenties c.q. functioneren van de aios is bewijsmateriaal nodig. Dit bewijs moet:

- over langere tijd verzameld zijn,
- afkomstig zijn van verschillende bronnen,
- gemeten zijn met behulp van een verscheidenheid aan instrumenten.

Er zijn binnen de medisch-specialistische vervolgopleiding vele toetsvormen en -instrumenten beschikbaar. Elk van deze instrumenten levert specifieke informatie over specifieke deelaspecten van het professionele handelen van de aios² (zie bijvoorbeeld ook <http://www.acgme.org/outcome/assess/toolbox.pdf>; <http://www.foundationprogramme.nhs.uk/pages/home/training-and-assessment>). De informatie over het functioneren van aios kan in een (elektronisch) portfolio verzameld en gepresenteerd worden zodat voor elke aios een competentieprofiel c.q. een taak- of themaprofiel opgesteld kan worden.

Door de informatie en toetsresultaten in het portfolio op een zorgvuldige manier te combineren en te wegen, is het mogelijk om enerzijds vast te stellen waar sterke c.q. zwakke kanten van de aios liggen voor wat betreft de zeven CanMEDS com-

Thema A							
Competentie	Taak 1	Taak 2	Taak 3	Taak 4	Taak 5	Taak ...	Taak n
Medisch handelen	X	X		X	X	X	
Communicatie	X	X	X	X		X	X
Samenwerken	X				X		X
Organisatie			X			X	
Kennis en wetenschap		X				X	
Maatschappelijk handelen	X						X
Professionaliteit	X	X	X	X	X	X	X

Thema B							
Competentie	Taak 1	Taak 2	Taak 3	Taak 4	Taak 5	Taak ...	Taak n
Medisch handelen	X	X		X		X	
Communicatie	X		X	X	X		X
Samenwerken	X			X	X		
Organisatie	X		X			X	
Kennis en wetenschap	X						X
Maatschappelijk handelen	X	X	X				X
Professionaliteit	X	X	X	X	X	X	X

petenties en anderzijds vast te stellen hoe de aios in de verschillende inhoudelijke thema's functioneert (met welke mate van zelfstandigheid, effectiviteit, efficiëntie, etc.). Op basis hiervan kunnen vervolgens onderbouwde beslissingen genomen worden ten aanzien van persoonlijke ontwikkelingsplannen, bekwaamheidsverklaringen, voortgang in de opleiding, etc.⁸

D. Gebruik van toetsresultaten in het voortgangsgesprek (formatief)

Een voortgangsgesprek is bedoeld voor de evaluatie van en reflectie op het functioneren van de aios in de afgelopen periode (drie-zes maanden), en voor het maken van specifieke afspraken voor de volgende periode.

Het Persoonlijk OntwikkelingsPlan (POP) van de aios vormt de basis voor de inhoud van het voortgangsgesprek. In het POP staan specifieke leerdoelen beschreven. Aan de hand van alle voorliggende informatie in het portfolio wordt zowel de voortgang in het leerproces als de mate waarin leerdoelen in de afgelopen periode zijn bereikt, geëvalueerd. De formatieve functie van toetsing staat in het voortgangsgesprek centraal. Toetsresultaten/beoordelingen helpen om sterke en minder sterke kanten in het functioneren te identificeren. Op basis hiervan kunnen concrete afspraken voor de volgende opleidingsperiode worden gemaakt. Hierbij is het belangrijk duidelijk af te spreken op welke manier getoetst kan en zal worden of de aios succesvol aan zijn leerdoelen heeft voldaan. Toetsresultaten zijn vooral bruikbaar voor het (bij)sturen van het leerproces als ze aan bepaalde voorwaarden voldoen (zie hoofdstuk 4).

E. Gebruik van toetsresultaten in het beoordelingsgesprek (summatief)

In het beoordelingsgesprek staat de vraag centraal of de betreffende aios voldoet

aan de (minimum)eisen die op dat moment gesteld worden. Hieraan gekoppeld wordt de vraag beantwoord of het verantwoord is om de aios te laten doorstromen naar een volgende fase van de opleiding c.q. loopbaan.

De basis voor het gesprek wordt gevormd door de aan de aios te stellen eisen enerzijds, en door de in het portfolio beschikbare informatie over het functioneren van de aios anderzijds. De minimeisen waaraan een aios op specifieke momenten in de opleiding moet voldoen staan deels omschreven in het opleidingsplan zoals opgesteld door de betreffende wetenschappelijke vereniging, en zijn uitgewerkt in het regionale opleidingsplan. Het portfolio, met de daarin opgenomen toetsresultaten, vormt het belangrijkste bewijsmateriaal ter onderbouwing van de te nemen beslissing. De beslissing is alleen geloofwaardig en verdedigbaar als bewijsmateriaal en besluitvormingsproces aan specifieke voorwaarden voldoen (zie hoofdstukken 4-5).

Elke bekwaamheidsverklaring die wordt afgegeven betreffende deelaspecten van thema's (bijvoorbeeld een specifieke verrichting) kan beschouwd worden als een *summatief* toetsmoment. Er wordt dan vastgelegd dat de aios aan de minimumvoorwaarden voldoet om de betreffende verrichting zonder directe supervisie uit te voeren (bijvoorbeeld het zelfstandig uitvoeren van een vacuümextractie). Werkplekbeoordelingen en expertoordelen van het opleiderteam vormen ook bij dergelijke beslissingen het belangrijkste bewijsmateriaal.

4. Interpretatie van toetsresultaten: kwaliteit en kwantiteit van werkplekbeoordelingen

Een toets *meet* (een deel van) het functioneren van de aios. Met behulp van verschillende toetsen (verscheidenheid van

Box 1. Een voorbeeld.

Als een kandidaat slaagt voor een kennistoets is onze conclusie vaak dat de betreffende kandidaat over voldoende kennis beschikt. Daarbij gaan we uit van een aantal aannames:

- dat de behaalde score een *nauwkeurige/precieze meting* is van de *kennis* van de kandidaat (en niet van andere, voor deze toets irrelevante zaken zoals IT-vaardigheden, leesvaardigheid of stressbestendigheid)
- dat we op basis van de toets uitspraken kunnen doen over de kennis van de kandidaat die verder gaan dan de specifieke inhoud van de items in de betreffende toets
- en dat de zak-slaaggrens/cesuur zodanig gekozen is dat er een betrouwbaar onderscheid wordt gemaakt tussen kandidaten die over voldoende kennis beschikken en kandidaten wier kennisniveau daadwerkelijk onvoldoende is.

Vaak gaat onze interpretatie van toetsscores nog een stap verder en gebruiken we toetsscores om daarmee een uitspraak te doen over de bekwaamheid van de kandidaat als beroepsbeoefenaar. We gaan er dan, op basis van de toetsresultaten, van uit dat de kandidaat ook in staat is om de kennis adequaat toe te passen in relevante beroepssituaties.

toetsinstrumenten en beoordelaars) worden toetsresultaten verkregen die worden gebruikt ter onderbouwing van *beslissingen*. Deze beslissingen kunnen betrekking hebben op (bij)sturing van het leerproces ofwel op voortgang in de opleiding (selectie). Beslissingen kunnen pas genomen worden na *interpretatie* van de verkregen meetresultaten. Bij het nemen van beslissingen op basis van toetsresultaten gaan we vaak uit van een aantal impliciete aannames. Zie voor een voorbeeld Box 1. De vraag is hoe gerechtvaardigd al deze aannames en conclusies zijn. Zijn de kandidaten met de hoogste scores ook het meest competent, en functioneren zij het best op de werkvloer?

De mate waarin de interpretatie van de toetsscores en de zak/slaagbeslissingen kloppen, hangt samen met de mate waarin de aannames over de toets terecht zijn. Bij meer bewijs dat de aannames terecht zijn, is de geldigheid van de conclusies die we aan de toetsresultaten verbinden groter. De basisgegevens, de meetresultaten, moeten dus aan een aantal kwaliteitseisen vol-

doen.⁹⁻¹⁰ Deze basiseisen worden hierna beschreven, met een specifieke nadruk op implicaties voor werkplekbeoordelingen.

A. Toetsresultaten zijn accuraat

De toetsresultaten moeten een *precieze* (betrouwbare) indicatie vormen voor het prestatieniveau op de toets c.q. te beoordelen taak. Zij moeten precieze informatie geven over sterke en zwakke kanten in het functioneren van de aios. Verschillen in toetsresultaten bij vergelijking van verschillende aios horen dus uitsluitend het gevolg te zijn van daadwerkelijke (verschillen in) bekwaamheid van aios en niet beïnvloed te worden door andere, min of meer toevallige factoren. Mogelijke en veel voorkomende foutenbronnen zijn bijvoorbeeld:

- extreme vermoeidheid of ziekte bij de kandidaat,
- onduidelijke scoringssystematiek,
- verkeerd gebruik van scoringsmodellen,
- een toets bestaande uit extreem moeilijke of gemakkelijke taken.

Accuraatesse van werkplekbeoordelingen in de praktijk

Uit onderzoek blijkt dat werkplekbeoordelingen in het algemeen niet erg accuraat zijn, vergeleken met scores die verkregen worden uit meer 'objectieve' of gestandaardiseerde toetsvormen.¹¹⁻¹³ Typische kenmerken van werkplekbeoordelingen in de klinische setting zijn onder andere:

- leniency (score-inflatie): beoordelaars zijn geneigd om mild en optimistisch te oordelen,¹⁴
- halo-effecten: de algemene indruk over de aios is bepalend voor het oordeel op deelcompetenties, of één aspect in het functioneren bepaalt oordelen over andere onderdelen van het functioneren,¹⁵
- verschillen in de manier waarop beoordelaars de kwaliteit van waargenomen functioneren interpreteren en omzetten in een score: sommige beoordelaars kunnen het eens zijn over kwaliteit van functioneren en dit toch anders scoren, terwijl andere beoordelaars van mening verschillen over de waargenomen kwaliteit en toch hetzelfde scoren.¹⁶⁻¹⁷

Verschillende oorzaken kunnen hieraan ten grondslag liggen:

- *Dubbelrol van de beoordelaar:* Beoordelaars (staf) hebben vaak een langdurige, min of meer intensieve werkrelatie met de aios en vervullen naast de rol van beoordelaar ook de rol van coach-supervisor. Onderzoek toont tamelijk consistent aan dat het vervullen van deze dubbelrol het geven van eerlijke en volledige feedback belemmert. Met name het toekennen van 'negatieve' beoordelingen en feedback wordt ervaren als lastig, en weinig motiveerend voor betrokkenen.
- *Gebrek aan directe observatie:* Oordelen met betrekking tot specifieke vaar-

digheden als communicatie in de arts-patiëntrelatie, of systematiek in anamnese en fysisch-diagnostisch onderzoek komen vaak tot stand op basis van indirecte informatie zoals verslaglegging in de status, of worden afgeleid uit de manier waarop de aios de patiëntcasus presenteert. Concrete feedback op en beoordeling van bepaalde klinische vaardigheden is zonder directe observatie niet mogelijk.¹⁶

- *Niet (tijdig) documenteren van observaties:* Door niet te scoren/schrijven tijdens de observaties, maar vaak (veel) later achteraf te documenteren treedt informatieverlies en -vervalsing op.¹⁸⁻²⁰
- *Eenzijdige focus op kwaliteit van medisch (technisch) handelen:* De beoordeling van de kwaliteit van medisch-technisch handelen bepaalt in sterke mate de scores op andere competenties.
- *Gebrek aan training en onderlinge afstemming tussen beoordelaars, leidend tot:*
 - onzekerheid bij beoordelaars over eigen competentie als beoordelaar,
 - gebrek aan consensus over kwaliteit van performance en te hanteren standaarden,
 - gebrek aan consensus over hoe het scoresysteem te gebruiken.

Geschreven, aanvullende opmerkingen over het functioneren van de aios, waarin een aantal sterke en minder sterke punten concreet gemaakt worden, zijn daarom veel belangrijker en informatiever dan numerieke scores.

B. Toetsresultaten zijn generaliseerbaar

Bij de interpretatie van toetsscores wil men meestal niet alleen uitspraken doen over het presteren van de aios op die ene geobserveerde taak of toets, maar conclusies trekken over verwachte prestaties in een veel groter domein van vergelijkbare taken (criteriumdomein). Meer concreet

betekent dit dat in de medisch-specialistische vervolgopleiding de vraag gesteld moet worden in hoeverre de toetsresultaten een betrouwbare indicatie vormen voor:

- beheersing van de betreffende CanMEDS competentie (bij de uitoefening van uiteenlopende beroepstaken) en/of
- de beheersing van alle vergelijkbare taken/toetsen in het betreffende thema of vakgebied.

Anders gezegd: in hoeverre zijn behaalde toetsresultaten reproduceerbaar bij herhaalde meting met andere (maar vergelijkbare) taken en beoordelaars? Uit onderzoek blijkt dat van alle factoren die van invloed zijn op de generaliseerbaarheid van toetsresultaten, de zogenoemde *'inhoudsspecificiteit'* de belangrijkste is.²¹⁻²² Hiermee wordt bedoeld dat de kwaliteit van functioneren nogal variabel is, en sterk afhankelijk van de specifieke inhoud van de taak of van het probleem waarmee iemand wordt geconfronteerd. Dat wil zeggen dat de prestatie op taak 1 niet of nauwelijks voorspellend is voor de prestatie op taak 2 uit hetzelfde inhoudelijke taakgebied. Zo is de kwaliteit van klinisch redeneren bij het oplossen van een patiëntprobleem sterk afhankelijk van de inhoud van dat specifieke probleem.

Het probleem van de inhoudsspecificiteit doet zich voor bij alle vormen van toetsing, zowel bij meer objectieve, gestandaardiseerde toetsen (zoals multiple choice examens) als bij beoordelingen op de werkplek (zoals KPB of OSATS). Behaalde toetsresultaten worden daarmee sterk afhankelijk van de toevallige keuze van taken of toetsitems. Dit impliceert dat, om betrouwbare uitspraken te kunnen doen over functioneren in een bepaald domein, een grote steekproef getrokken moet worden uit het betreffende (taak)domein. Hoe breder het domein

waarnaar we uitspraken willen generaliseren, hoe groter de steekproef van te beoordelen taken dient te zijn.²³⁻²⁴

Een andere factor die generaliseerbaarheid kan beïnvloeden wordt gevormd door beoordelaars. Zo kunnen beoordelaars soms sterk van elkaar verschillen in de manier waarop ze geobserveerd gedrag interpreteren en waarderen. Er zijn strenge en milde beoordelaars, maar er bestaat vaak ook een verschil in opvatting over wat in welke fase van de opleiding beheerst moet worden of wat belangrijk is voor geleverde kwaliteit van zorg. De opzet van het scoresysteem en het beoordelingssysteem vormt een andere factor die van invloed is op betrouwbaarheid van toetsresultaten. Indien er ruimte voor subjectiviteit in de interpretatie van criteria bestaat zal de betrouwbaarheid strikt genomen afnemen. Ook hierbij geldt dat betrouwbaarheid positief beïnvloed wordt door een grotere steekproef: oordelen over aios zijn meer betrouwbaar naarmate meer verschillende examinatoren betrokken zijn bij de totstandkoming ervan. Onderzoek toont echter consistent aan dat laatstgenoemde factoren bij zorgvuldig ontworpen beoordelingsinstrumenten en goed getrainde beoordelaars van (veel) minder belang zijn dan de inhoudsspecificiteit.²⁵

Generaliseerbaarheid van werkplekbeoordelingen in de praktijk

Werkplekbeoordelingen zijn arbeidsintensief en niet eenvoudig te realiseren. Uitspraken over het functioneren van de aios zijn daarom vaak gebaseerd op een zeer beperkt aantal beoordelingen, bij een beperkt aantal taken en door een beperkt aantal verschillende beoordelaars. Dit maakt het doen van gegeneraliseerde uitspraken vaak lastig. Onderzoek toont aan dat voor werkplekbeoordelingen, gebaseerd op individuele patiëntcasus, in grote lijn

geldt dat 6-11 beoordelingen noodzakelijk zijn om te komen tot betrouwbare oordelen over professionele competentie.^{11 26-27} Onderzoek toont daarbij overigens aan dat het aantal benodigde beoordelingen afhankelijk is van het te beoordelen domein en van de te beoordelen competenties.¹¹

Wat betreft het aantal beoordelaars dat nodig is om te komen tot betrouwbare, generaliseerbare oordelen, geldt ten aanzien van werkplekbeoordelingen dat 6-10 beoordelaars nodig zijn om van betrouwbare oordelen te kunnen spreken. Indien gebruik wordt gemaakt van patiëntoordelen, zijn grotere aantallen noodzakelijk.¹¹⁻¹²

360 graden feedbackprocedure en betrouwbaarheid van oordelen

In de context van *360 graden feedback* verdienen verschillen tussen beoordelaars specifieke aandacht. In het algemeen geldt dat medisch specialisten/stafleden de belangrijkste groep beoordelaars vormen bij het beoordelen van aios op de werkplek. Kenmerkend voor 360 graden feedback is dat verschillende beoordelaargroepen in de procedure betrokken worden. Deze beoordelaargroepen hebben op verschillende manieren en in verschillende situaties met aios te maken, en verschillen ook meestal van elkaar wat betreft de manier waarop ze het functioneren van een aios waarderen en beoordelen. Verpleegkundigen bijvoorbeeld, observeren niet alleen ander gedrag van een aios dan patiënten of stafleden, zij zullen mogelijk ook ander gewicht toekennen aan bepaalde gedragingen en dus anders oordelen. Dit betekent echter niet dat deze – verschillende – beoordelingen daarom niet betrouwbaar zijn. De verschillende invalshoeken waarmee naar het functioneren van de aios wordt gekeken vormen de belangrijkste basis voor het gebruik van 360 graden feedback. Het betrekken van verschillende groepen beoordelaars verrijkt de informa-

tie op basis waarvan conclusies getrokken kunnen worden.²⁸

C. Toetsresultaten zijn transferabel/ extrapoleerbaar

In het algemeen zijn we minder geïnteresseerd in de bekwaamheid van de aios met betrekking tot het beantwoorden van meerkeuzevragen, of met betrekking tot het uitvoeren van een technische handeling op een dummy in een simulatietoets. Belangrijk is in hoeverre de toetsresultaten een relevante indicatie vormen voor feitelijk functioneren in de praktijk. De toetsscore moet vertaalbaar zijn naar de situatie waarin datgene wat we meten praktisch relevant is. Naarmate toetsing meer plaatsvindt op basis van levens-echte, authentieke taken, uitgevoerd in de context van de alledaagse beroepspraktijk, is extrapoleren van resultaten gemakkelijker, en vormen toetsresultaten een meer ‘directe’ aanwijzing en bewijs voor gedrag waarin we uiteindelijk geïnteresseerd zijn.

Transferabiliteit/extrapoleerbaarheid van werkplekbeoordelingen in de praktijk

Extrapoleerbaarheid van werkplekbeoordelingen lijkt in eerste instantie geen problemen op te leveren: immers het geobserveerde en beoordeelde gedrag is identiek aan gedrag dat in de praktijk van alledag getoond wordt. Extrapoleerbaarheid van werkplekbeoordelingen kan echter worden beïnvloed door een aantal factoren:

- Directe observatie (zeker indien gevolgd door een beoordeling) kan van invloed zijn op het functioneren van de aios. Een aios kan in een dergelijke situatie ófwel beter ófwel slechter presteren dan hij/zij normaal gesproken gewoon is te doen.
- Werkplekbeoordelingen geven informatie over wat een aios in de dagelijkse praktijk doet. Het beoordelen van func-

tioneren op de werkplek aan de hand van een (gestructureerd/gestandaardiseerd) beoordelingsformulier betekent niet automatisch dat de uiteindelijke scores een goed beeld geven van het feitelijke functioneren van de aios. Zo kunnen bijvoorbeeld de items op het beoordelingsformulier meer of minder relevant en representatief zijn voor wat we feitelijk willen meten. In andere woorden: de manier waarop geobserveerd gedrag wordt vertaald in scores en wordt vastgelegd op een formulier, bepaalt de informatiewaarde van de werkplekbeoordeling.

In het algemeen geldt het volgende. Op basis van een enkele werkplekbeoordeling wordt *specifieke* informatie verkregen: specifieke informatie over functioneren met betrekking tot betreffende taak, in betreffende specifieke situatie. Gegeneraliseerde uitspraken (over competentie, bekwaamheid van de aios) zijn slechts mogelijk indien voldoende gevarieerde meetgegevens uit een domein beschikbaar zijn. Dit neemt echter niet weg dat separate (taak)beoordelingen een indicatie kunnen vormen voor specifieke sterke c.q. minder sterke kanten in het functioneren van de aios.

In het algemeen geldt bij het interpreteren van toetsscores dat het trekken van een juiste en verdedigbare conclusie over de competentie c.q. bekwaamheid van een aios alleen mogelijk is als de toetsscores zowel accuraat, generaliseerbaar als transferabel/extrapoleerbaar zijn. Helaas voldoet geen enkel toetsinstrument volledig aan deze eisen. Het gebruik van verschillende instrumenten die door ieder meer dan eens worden ingezet is daarom een absolute noodzaak. Dit palet van toetsen en beoordelingen vormt een toetsprogramma. Het portfolio dient hierbij als verzamelinstrument.

5. Gebruik van toetsresultaten: verdedigbaarheid van beslissingen

Zorgvuldig gebruik van toetsresultaten vereist een hoge kwaliteit van het besluitvormingsproces. Dit proces moet transparant, traceerbaar, geloofwaardig en rechtvaardig zijn. Beslissingen moeten gefundeerd zijn en gemotiveerd kunnen worden. Interpretatie en gebruik van toetsresultaten zou plaats moeten vinden aan de hand van overeengekomen *standaarden* (dit kunnen leerdoelen, opleidingsdoelstellingen of minimaal vereiste bekwaamheidsniveaus zijn) en van tevoren vastgestelde *procedures*.

A. Werkplekbeoordelingen en kwaliteit van standaarden

De gehanteerde standaarden en cesuur moeten verdedigbaar zijn. Een groot probleem op dit moment is het ontbreken van duidelijke en evidence-based standaarden waaraan een aios op specifieke momenten in de opleiding moet voldoen. Hiervoor zijn een aantal oorzaken aan te wijzen:

1. Voor een aantal algemene competentiedomeinen, zoals communicatie, professionaliteit of samenwerken, zijn nog geen absolute standaarden vastgesteld. Hoewel dit niet eenvoudig lijkt, zullen ook voor deze competenties standaarden ontwikkeld moeten worden.
2. Bij werkplekbeoordelingen is sprake van een hoge mate van leniency (score-inflatie). Uit onderzoek in de V.S. blijkt dat het merendeel (> 80%) van de arts-assistenten beoordeeld wordt als 'excellent' of 'very good'.¹⁴ Een score 'good' of 'satisfactory' zou dan wel eens geïnterpreteerd moeten worden als 'twijfelachtig' respectievelijk 'onvoldoende'.
3. Er bestaan soms grote verschillen in beoordelingssystematiek bij vergelijking van verschillende opleidingszie-

kenhuizen, als gevolg van verschillen in gehanteerde scoresystemen (instrumenten), verschillen in opleidingsklimaat, verschillen in opvattingen en afspraken over toetsing, verschillen in patiëntaanbod, etc.

Op grond van het bovenstaande kan geconcludeerd worden dat het ontwikkelen van goed onderbouwde standaarden alleen mogelijk is indien aandacht wordt besteed aan training van beoordelaars, er een onderlinge afstemming van beoordelingssystemen is en indien systematisch, en over een langere tijdsperiode, gegevens over functioneren van aios op de werkplek, in verschillende fasen van de opleiding en in verschillende settings, worden verzameld en toegankelijk worden gemaakt. Systematische documentatie van werkplekbeoordelingen is daarbij een noodzakelijke randvoorwaarde.²⁰

B. Zorgvuldige besluitvorming

Verdedigbaarheid van besluiten vraagt om zorgvuldigheid in het besluitvormingsproces. Van zorgvuldige besluitvorming kan worden gesproken indien minimaal is voldaan aan een aantal randvoorwaarden.²⁹⁻³¹

1. *Adequate notice (transparantie)*: Voor de medisch-specialistische vervolgopleiding betekent dit dat aios tijdig geïnformeerd zijn over de opzet en inhoud van het beoordelingsproces (instrumenten en procedures, beoordelingscriteria, standaarden c.q. te bereiken doelstellingen) en dat er regelmatig en tijdig tussentijdse feedback gegeven is over het functioneren (met eventueel mogelijkheid tot remediatie).
2. *Fair hearing*: Zorgvuldige besluitvorming berust niet alleen op input van voldoende geloofwaardige beoordelaars, maar ook van beoordeelde zelf. Dit betekent dat de aios in staat gesteld moet worden om input te leveren in het beoordelingsproces, bijvoorbeeld in de vorm van zelfbeoordeling, reflectie. In elk geval dient hij/zij tijdig in de gelegenheid gesteld te worden om verkregen (voorgenomen) beoordelingen te becommentariëren. Regelmatig ingeplande (formele) voortgangsgesprekken en andere feedbacksessies maken deel uit van een 'fair' systeem van beoordelen. Daarnaast horen aios voldoende geschoold/getraind te zijn in het voeren van feedbackgesprekken, voortgangsgesprekken en beoordelingsgesprekken.
3. *Beslissingen worden volgens vastgestelde procedures genomen*: Er wordt een duidelijk onderscheid gemaakt tussen formatief en summatief gebruik van beoordelingsresultaten. Beslissingen (summatief) zijn altijd gebaseerd op geaggregeerde, dat wil zeggen over langere tijd en op basis van verschillende instrumenten verzamelde informatie. Beslissingen berusten daarbij op input van voldoende beoordelaars, en worden genomen door een daartoe verantwoordelijk gestelde groep (bijvoorbeeld de opleidergroep). De kwaliteit van de besluitvorming neemt hierdoor toe.³²⁻³³
4. *Beslissingen worden gemotiveerd en onderbouwd*: Besluitvorming moet zoveel als mogelijk 'evidence-based' zijn. Zorgvuldige en volledige documentatie van feedback gedurende de opleiding is belangrijk, zowel voor het leerproces van de aios, als voor de besluitvorming. Documentatie van positieve feedback heeft een sterk motiverend en bekrachtigend effect op het leren door de aios.³⁴⁻³⁶ Uit onderzoek blijkt dat het ontbreken van documentatie een belangrijke reden kan vormen voor het uitstellen van negatieve beslissingen.³⁷
5. *Training, coaching van beoordelaars*: Beoordelaars moeten voldoende ver-

trouwd zijn met de achtergronden en doelstellingen van het beoordelings-systeem, en getraind zijn in het gebruik van instrumenten.³⁸⁻³⁹

Effectief gebruik van werkplekbeoordelingen (toetsresultaten) vereist een beoordelingssysteem dat door alle betrokkenen ervaren wordt als rechtvaardig en zinvol. Doorslaggevend daarbij zijn vertrouwen in opzet en procedures, veiligheid, training, coaching en in gezamenlijk gevoelde verantwoordelijkheid.^{13 31 40-41}

6. Randvoorwaarden en algemene aanbevelingen

Opleidergroep

- Zorg voor scholing van opleiderteam en bij opleiden betrokken beoordelaars. Faciliteer coaching 'on-the-job', bijvoorbeeld door een senior-stafid en/of onderwijskundige.⁴²
- Formaliseer procedures, waar nodig. Maak duidelijke afspraken over taken en verantwoordelijkheden wat betreft begeleiding en toetsing van aios.
- Draag als opleidergroep gezamenlijk verantwoordelijkheid voor de kwaliteit van toetsing.
- Zorg voor duidelijke standaarden (consensus) en communiceer deze. Maak duidelijk wat op welk moment van de aios wordt verwacht.
- Evalueer regelmatig als groep het functioneren van *alle* aios aan de hand van portfolio's.
- Neem als groep besluiten over aios.
- Bespreek 'probleem-aios' en stel zonedig een remediërsprogramma vast.

Instrumenten afgestemd op doelstellingen van toetsing

- Gebruik instrumenten die inzicht geven in ontwikkeling/groei met betrekking tot CanMEDS competenties.

- Gebruik instrumenten die inzicht geven in ontwikkeling/groei met betrekking tot opleidingsspecifieke thema's.
- Zorg voor verzamelen van voldoende kwalitatieve informatie (narratieve feedback), naast kwantitatieve informatie.

Faciliteer en bewaak zorgvuldige en volledige dossiervoering (portfolio)

- Zorg voor een gebruikersvriendelijk systeem (web-based).
- Zorg voor systematische en structurele evaluatie van aios op alle relevante (deel)competenties en thema's.
- Organiseer, waar nodig, de toetsing op de werkplek. Zorg voor inbedding van werkplekbeoordelingen in de dagelijkse werkroutine.
- Faciliteer directe observatie en documentatie van performance. Plan werkplekbeoordelingen, waar nodig, in.
- Zorg voor brede sampling.
- Betrek verschillende beoordelaargroepen in het proces.

Systematische en structurele monitoring van performance van grote groepen aios (referentiegroepen)

- Zorg voor documentatie van performance.
- Verzamel gegevens van aios over een langere periode en in verschillende opleidingsinstellingen.

Training/professionalisering van aios

- Zorg voor scholing van aios met betrekking tot voeren van feedbackgesprekken (geven en ontvangen van feedback; omzetten van feedback in concrete acties).
- Zorg voor scholing van aios in het voeren van voortgangs- en beoordelingsgesprekken.

7. Samenvattende overzichten

A. Toetsresultaten in begeleidingsgesprekken

Voortgangsgesprek

Doel: Evaluatie van en reflectie op het functioneren van de aios.
Bespreken van de ontwikkeling van de aios in relatie tot Persoonlijk Ontwikkelingsplan (POP) en het opleidingsplan.
Opstellen c.q. bijstellen van Persoonlijk Ontwikkelings Plan.

Materiaal (Portfolio): Afspraken voorgaand gesprek.
Persoonlijk Ontwikkelingsplan.
Sterkte-zwakke-analyse.
Bewijsmateriaal:

- Competentieprofiel (CanMEDS)
- Ontwikkeling vakinhoudelijke thema's (themaprofiel: KBS, verrichtingen; bekwaamheidsverklaringen)
- Gevolgd (cursorisch) onderwijs
- Onderzoek, publicaties, presentaties/referaten
- Onderwijstaken
- Toetsresultaten (werkplekbeoordelingen, overige toetsen)
- Overig

Gebruik van toetsresultaten: *formatief*. Feedback op functioneren; bekrachtiging van sterke punten, signaleren van punten ter verbetering.

Aandachtspunten: Wat wilde de aios bereiken?
Wat heeft aios afgelopen periode geleerd/bereikt?

- Waaruit blijkt dat (bewijsmateriaal!)?

Sterke punten in het functioneren van de aios?
Wat heeft aios afgelopen periode niet gerealiseerd of kunnen bereiken? Waar blijft groei achter bij verwachtingen?

- Wat ligt daaraan ten grondslag? Belangrijkste belemmerende factoren?

Belangrijkste aandachtspunten in het functioneren van de aios?
Wat betekent bovenstaande voor de komende periode?

Resultaat: (Bijgesteld) Persoonlijk Ontwikkelingsplan met SMART geformuleerde doelen. Hou hierbij rekening met:

- Eisen zoals in het opleidingsplan omschreven voor betreffende opleidingsperiode.
- Persoonlijke leerwensen van de aios.
- Takenpakket en mogelijkheden op de (nieuwe) werkplek.
- Let op hoge en lage scores: zij kunnen een indicatie vormen voor specifieke sterke en zwakke kanten in het functioneren van de aios. Check bijbehorende narratieve feedback.
- Let op alarmsignalen: extreem lage scores of een indicatie dat aios onder verwacht niveau functioneert.
- Let op critical incident reports. Zij kunnen een aanwijzing vormen voor specifieke problemen in het functioneren van de aios.

Beoordelingsgesprek

Doel: Beslissen over geschiktheid van aios om zelfstandig (een deel van) het vak uit te oefenen en/of de opleiding voort te zetten dan wel af te ronden.

Materiaal (Portfolio): Opleidingsplan.
Persoonlijk Ontwikkelingsplan.
Bewijsmateriaal:

- Competentieprofiel (CanMEDS)
- Ontwikkeling vakinhoudelijke thema's (themaprofiel: KBS, verrichtingen; bekwaamheidsverklaringen)
- Gevolgd (cursorisch) onderwijs
- Onderzoek, publicaties, presentaties/referaten
- Onderwijstaken
- Toetsresultaten (werkplekbeoordelingen, overige toetsen)
- Overig

Gebruik van toetsresultaten: *summatief*. Deel van de informatie die gebruikt wordt bij onderbouwing van beslissingen (bewijsmateriaal).

Aandachtspunten: Voldoet de aios aan alle (minimum)eisen die op dat moment van de opleiding aan hem/haar gesteld worden?
Waar blijkt dat uit?
Is het bewijsmateriaal c.q. de informatie in het portfolio van voldoende kwaliteit en kwantiteit?
Zijn besluitvormingsprocedures zorgvuldig gevolgd? Is het opleiderteam in voldoende mate betrokken/gehoord? Is de aios gehoord?
Is aanvullende informatie noodzakelijk (zo ja, op welke punten)?

Resultaat: Gemotiveerde, onderbouwde:

- Go/no-go beslissing en/of
- Bekwaamheidsverklaring(en)

Hierbij wordt rekening gehouden met: opleidingseisen, functioneren van de aios ten opzichte van relevante referentiegroep (peers) en verwachtingen omtrent verdere leerproces van de aios.

Bij twijfel:

- Heldere, schriftelijk vastgelegde afspraken betreffende benodigde aanvullende informatie
 - wat, door wie, hoe, op welke termijn

Remediatie (geïntensiveerd begeleidingstraject)

Hierbij wordt rekening gehouden met inschatting van mogelijkheden tot verdere groei / ontwikkeling en veiligheid voor aios / patiënten / studenten en collega's.

B. Kwaliteit en kwantiteit van werkplekbeoordelingen

Het gebruik van het portfolio, hetzij bij het coachen of sturen van het leerproces, hetzij bij het nemen van beslissingen, veronderstelt dat de in het portfolio verzamelde informatie een representatief beeld geeft van het feitelijke functioneren van de aios in de opleiding. De informatie in het portfolio dient daartoe aan een aantal kwaliteitseisen te voldoen.

Algemeen geldt bij het interpreteren van toetsscores dat het trekken van een juiste en verdedigbare conclusie over de compe-

tentie c.q. bekwaamheid van een aios alleen mogelijk is als de scores zowel *accuraat, generaliseerbaar en transferabel/extrapoleerbaar* zijn. Helaas voldoet geen enkel toetsinstrument volledig aan deze eisen. Het gebruik van verschillende instrumenten die door ieder meer dan eens worden ingezet is daarom een absolute noodzaak.

Meer specifiek voor werkplekbeoordelingen gelden aandachtspunten van accuratesse, generaliseerbaarheid en extrapoleerbaarheid. Deze zijn in de hierna volgende boxen toegelicht.

Accuratesse

Toetsresultaten moeten een accuraat beeld geven van het functioneren van de aios met betrekking tot de CanMEDS en themagebonden taken. Sterke kanten en verbeterpunten in het functioneren moeten zorgvuldig worden omschreven.

Geschreven, aanvullende opmerkingen over het functioneren van de aios, waarin een aantal sterke en minder sterke punten concreet gemaakt worden, zijn veel belangrijker en informatiever dan numerieke scores.

Aandachtspunten bij interpretatie van werkplekbeoordelingen in het portfolio:

1. Is de beoordeling tot stand gekomen op basis van *directe observatie* van het functioneren?
2. Is informatie voldoende specifiek en concreet? Worden numerieke scores onderbouwd met kwalitatieve gegevens (geschreven, beschrijvende feedback)?
3. Is variatie in scores aanwezig en verklaarbaar (zijn scoreverschillen zichtbaar)?
 - a. Op beoordelingsformulieren: is een te verwachten variatie in scores op verschillende deelcompetenties zichtbaar, of is veeleer sprake van globale, algemene indrukken (dat wil zeggen identieke scores op alle deelaspecten van functioneren)?
 - b. Over verschillende beoordelingsmomenten en -instrumenten heen: is groei in een bepaalde tijdsperiode zichtbaar (toenemende competentie en bekwaamheid)?
 - c. Zijn scoreverschillen tussen aios aanwezig en herkenbaar?

Veel voorkomende problemen bij werkplekbeoordelingen (zie hoofdstuk 4):

- Halo
- Leniency
- Variatie in beoordelagedrag o.a. als gevolg van tijdsdruk, ongetrainde en/of ongemotiveerde beoordelaars.

Generaliseerbaarheid

Toetsresultaten moeten een betrouwbare indicatie vormen voor daadwerkelijke beheersing van het deel van het vakgebied (het domein) waarover we uitspraken willen doen.

Op basis van één enkele werkplekbeoordeling wordt *specifieke* informatie verkregen: specifieke informatie over functioneren m.b.t. betreffende taak, in betreffende specifieke situatie. Gegeneraliseerde uitspraken (over competentie of bekwaamheid van de aios) zijn slechts mogelijk indien voldoende gevarieerde meetgegevens beschikbaar zijn.

Aandachtspunten:

1. Is er sprake van voldoende en voldoende variatie in beoordelingsmomenten?
 - a. Voldoende variatie in patiënten/patiëntproblematiek?
 - b. Voldoende variatie in beoordelingssituaties en - instrumenten?
 - c. Voldoende variatie in beoordelaars?
2. In hoeverre en op welke punten wijzen toetsresultaten in eenzelfde richting? Op welke punten lijken toetsresultaten elkaar tegen te spreken?
 - a. Indien informatie uit verschillende bronnen weinig consistent is, is verzamelen van aanvullende informatie noodzakelijk.
 - b. Toetsresultaten van momentopnames (KPB, OSATS) zijn doorgaans sterk afhankelijk van de specifieke kenmerken van de patiëntcasus. Als gevolg is hierbij meer variatie in scores te verwachten dan bij het beoordelen van het functioneren over langere tijd (zoals bij 360 graden feedback). Naarmate de aios meer ervaring opdoet, zal kwaliteit van functioneren minder afhankelijk zijn van het toevallige patiëntprobleem.

Let op het aantal beoordelingsmomenten c.q. beoordelaars (KPB, OSATS). Betrouwbare resultaten vereisen minimaal 6-11 beoordelingen c.q. beoordelaars. Variatie in patiëntpresentaties/casuïstiek is belangrijker dan variatie in beoordelaars!

Let bij 360 graden feedback op voldoende variatie in beoordelaargroepen (peers, staf, verpleegkundigen, paramedici, overige medewerkers, patiënten) en voldoende beoordelaars per groep (6-10; grotere aantallen bij patiënten!).

Veel voorkomende problemen bij werkplekbeoordelingen (zie hoofdstuk 4):

- beperkte sampling.

Extrapolerbaarheid

Toetsresultaten moeten een waardevolle en betekenisvolle indicatie geven over het feitelijke functioneren van de aios in de dagelijkse praktijk.

Aandachtspunten:

1. Zijn de beoordelingssituaties authentiek en voor de praktijk relevant?
 - a. Is de aios in voldoende mate beoordeeld tijdens de uitoefening van het werk, aan de hand van relevante en authentieke taken?
 - b. Zijn alle relevante beoordelingsaspecten meegenomen in het beoordelingsproces? Denk aan *alle* CanMEDS gebieden!

Veel voorkomende problemen bij werkplekbeoordelingen:

- Exclusieve nadruk op (medische) kennis.
- Exclusieve nadruk op medisch-technisch handelen.
- Bewijsvoering op basis van indirecte informatie (bijvoorbeeld uitspraken over communicatie op basis van schriftelijke verslaglegging in de status).

C. Verdedigbaarheid van beslissingen: standaarden en besluitvormingsproces

Standaarden

Beslissingen komen tot stand op basis van vergelijking van informatie met een bepaalde standaard. Met betrekking tot toetsing, en zeker in het geval van werkplekbeoordelingen, is het belangrijk te realiseren dat een 'gouden standaard' ontbreekt.

Aandachtspunten:

1. Zijn er duidelijke afspraken over de standaarden die gehanteerd worden bij interpretatie van resultaten ten behoeve van voortgangsbespreking c.q. beoordeling?
 - a. In het *voortgangsgesprek* worden de 'standaarden' merendeels gevormd door specifieke leerdoelen in het leerplan van de individuele aios. Uiteraard horen deze aan te sluiten bij, en in elk geval niet te conflicteren met het opleidingsplan.
 - b. In het *beoordelingsgesprek* worden bij voorkeur standaarden gehanteerd die een beschrijving geven van de minimumeisen die aan de aios op het moment van beoordelen gesteld moeten worden. Dergelijke *absolute* standaarden zijn deels beschreven in het opleidingsplan zoals opgesteld door de wetenschappelijke vereniging en uitgewerkt in het regionale opleidingsplan.
2. Waar geen absolute standaarden voorhanden zijn, kan de professionele ontwikkeling van de aios vergeleken worden met die van een adequaat gekozen referentiegroep. In het merendeel van de gevallen zal de referentiegroep bestaan uit aios in een zelfde fase van de opleiding. *Vergelijking met een referentiegroep is alleen mogelijk indien systematisch en over langere tijd verzamelde gegevens van grote groepen aios beschikbaar zijn.*

Veel voorkomende problemen bij werkplekbeoordelingen:

- Ontbreken van SMART-geformuleerde leerdoelen (bij voortgangsbesprekingen).
- Standaarden zijn weinig concreet en onderliggende criteria zijn op meer manieren uit te leggen.
- Instrumenten zijn onduidelijk wat betreft het te hanteren referentiekader bij beoordelen (bijvoorbeeld: is 'uitstekend' de best haalbare score voor deze aios in dit jaar van de opleiding, of weerspiegelt 'uitstekend' het best haalbare niveau aan het eind van de opleiding? Ofwel: beoordeling naar eindniveau opleiding, of naar opleidingsfase waarin aios zich op dat moment bevindt?). Als gevolg daarvan hanteren beoordelaars verschillende referentiekaders in hun werkplekbeoordeling.
- Gegevens van referentiegroepen zijn niet beschikbaar.

Het ontwikkelen van goed onderbouwde standaarden is alleen mogelijk indien aandacht wordt besteed aan training van beoordelaars, onderlinge afstemming van beoordelingssystemen en indien systematisch en over een langere tijdperiode gegevens over functioneren van aios op de werkplek, in verschillende fasen van de opleiding en in verschillende settings, worden verzameld en toegankelijk gemaakt. Systematische documentatie van werkplekbeoordelingen is daarbij een noodzakelijke randvoorwaarde.

Besluitvormingsproces

Beslissingen zijn alleen verdedigbaar als besluitvormingsprocedures zorgvuldig zijn gevolgd. Daarbij is het belangrijk een duidelijk onderscheid te maken tussen *formatief* en *summatief* gebruik van toetsresultaten/werkplekbeoordelingen (hoofdstuk 4).

Beslissingen (summatief) worden genomen op basis van *geaggregeerde* informatie in het portfolio. Een enkelvoudige werkplekbeoordeling (momentopname) vervult uitsluitend een formatieve functie (feedback).

Formaliseer summatieve besluitvorming door:

- het instellen van een opleidergroep die gezamenlijke verantwoordelijkheid draagt voor besluitvorming,
- beslissingen zorgvuldig te motiveren en te documenteren.

Aandachtspunten:

1. Is de informatie in het portfolio van voldoende kwaliteit en kwantiteit, voor alle te beoordelen competentiedomeinen?
2. Zijn te hanteren criteria en standaarden helder?
3. Wordt *alle* beschikbare informatie in het portfolio meegenomen in de beslissing?
4. Bij uitspraken over algemene competenties of professioneel handelen in brede zin: is er sprake van een mix van 'momentopnames' (KPBs, OSATS) en beoordelingen van functioneren over langere tijd (360 graden feedback, algemene oordelen van supervisors)?
5. Zijn voldoende stafleden uit de opleidergroep in voldoende mate gehoord; input van voldoende geloofwaardige beoordelaars?
6. Is de aios in voldoende mate gehoord: heeft de aios gelegenheid gehad om op de resultaten en voorgenomen beslissing te reageren?
 - a. Voortgangsgesprekken gevoerd en gedocumenteerd?
 - b. Regelmatig tussentijds feedback; gedocumenteerd?
 - c. Zelfbeoordeling(en) aanwezig en besproken?
 - d. Reflectieverslagen aanwezig en besproken?
7. Zijn procedures toetsbaar?
 - a. Waren er heldere afspraken m.b.t. beoordelingsprocedures, -instrumenten, en -criteria?
 - b. Is documentatie zorgvuldig en volledig (motivering, onderbouwing aanwezig)?

Bij twijfel: aanvullende informatie verzamelen!

Literatuur

1. Epstein RM, Hundert E.M. Defining and Assessing Professional Competence. *JAMA* 2002; 287(2):226-235.
2. Epstein RM. Assessment in Medical Education. *NEJM* 2007;356: 387-396.
3. Turnbull J, Gray J, MacFadyen J. Improving In-Training Evaluation Programs. *JGIM* 1998;13:317-323.
4. Tamblyn R. Outcomes in medical education: what is the standard and outcome of care delivered by our graduates. *Advances in Health Sciences Education Theory Pract* 1999;4:9-25.
5. Ten Cate O, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med* 2007; 82(6):542-547.
6. Van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309-317.
7. Hager P, Gonczi A, Athanasou J. General Issues about Assessment of Competence. *Assessment & Evaluation in Higher Education* 1994;19(1),3-16.
8. Van Tartwijk J, Driessen EW. Portfolios for assessment and learning AMEE Guide No 45. *Medical Teacher* 2009;31(9):790-80.
9. Kane M. Validation. In: R.L. Brennan (ed.) *Educational Measurement*. Westport CT: American Council on Education/Praeger; 2006, p.621-694.
10. Kane M, Crooks T, Cohen A. Validating Measures of Performance. *Educational Measurement: Issues and Practice*, Summer 1999, 5-17.
11. Williams RG, Klamen DA, MacGaghie WC. Cognitive, Social and Environmental Sources of Bias in Clinical Performance Ratings. *Teach Learn Med* 2003;15(4):270-292.
12. Turnbull J, Barneveld van C. Assessment of clinical performance: in-training evaluation. In: G.R. Norman, C.P.M. Van der Vleuten & D.I. Newble (eds.), *International Handbook of Research in Medical Education* Dordrecht: Kluwer Academic Publishers; 2002; p.793-810.
13. Murphy K R, Cleveland JN. *Understanding Performance Appraisal. Social, organizational and Goal-Based perspectives*. Thousand Oaks, 1995; CA: Sage Publications.
14. Williams RG, Dunnington G. Prognostic value of resident clinical performance ratings, *Journal of the American College of Surgeons* 2004;199(4): 620-627.
15. Verhulst SJ, Colliver JA, Paiva, RE, Williams RG. A factor analysis study of performance of first-year residents. *J. Med Educ* 1986;61:132-134.
16. Holmboe ES. Faculty and the Observation of Trainees' Clinical Skills: Problems and Opportunities. *Acad Med* 2004;79:16-22.
17. Littlefield JH, DaRosa DA, Anderson KD, Bell RM, Nicholas GG, Wolfson PJ. Assessing performance in clerkships: accuracy of surgery clerkship performance raters. *Acad Med* 1991;66(9):S16-S18.
18. DeNisi AS, Peters LH. Organization of information in memory and the performance appraisal process: evidence from the field. *Journal of Applied Psychology* 1996;81(6):717-737.
19. DeNisi AS, Robbins T, Cafferty TP. Organization of information used for performance appraisals: role of diary-keeping. *Journal of Applied Psychology* 1989;74(1):124-129.
20. Williams RG, Dunnington GL, Klamen DL. Forecasting Residents' Performance – Partly Cloudy. *Acad Medicine* 2005;80 (5):415-422.
21. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teach Learn Med* 1990; 2:58-76.
22. Elstein AS, Shulman LS, Sprafka SA. *Medical Problem-solving: An Analysis of Clinical Reasoning*. Cambridge; Massachusetts: Harvard University Press; 1978.
23. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;1:41-67.
24. Eva KW. On the generality of specificity. *Med Educ* 2003;37(7):587-588.
25. Brennan RL. Performance Assessments from the Perspective of Generalizability Theory. *Applied Psychology Measurement* 2000;24:339-353.
26. Williams RG, Verhulst S, Colliver JA, Dunnington GL. Assuring the reliability of resident performance appraisals: More items or more observations? *Surgery* 2005;137:141-147.
27. Wilkinson JR, Crossley JGM, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ* 2008;42:364-373.
28. Mount MK, Scullen SE. Multisource Feedback Ratings: What Do They Really Measure? In: M. London (ed). *How People Evaluate Others in Organizations*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001, p.155-176.
29. Taylor MS, Tracy KB, Renard MK, Harrison J.K, Carroll SJ. Due Process in Performance Appraisal: A Quasi-experiment in Procedural Justice. *Administrative Science Quarterly* 1995;40:495-523.
30. Erdogan B, Kraimer ML, Liden RC. Procedural justice as a two-dimensional construct. An examination in the performance appraisal context. *Journal of Applied Behavioural Science* 2001; 37(2):205-222.
31. Levy PE, Williams JR. The Social Context of Performance Appraisal: A Review and Framework for the Future. *Journal of Management* 2004; 30:881-905.
32. Schwind CJ, Williams RG, Boehler ML, Dunnington GL. Do individual attending post-rotation

- performance ratings detect resident clinical performance? Acad Med 2004;79:453-457.
33. Williams RG, Schwind CJ, Dunnington GL, Fortune J, Rogers D, Boehler M. The Effects of Group Dynamics on Resident Progress Committee Deliberation. Teach Learn Med 2005;17(3), 96-100.
 34. Norcini J, Burch V. Workplace-based assessment as an educational tool. AMEE Guide 31. 2007, Dundee: AMEE.
 35. Hatty J, Timperley H. The Power of Feedback. Review of Educational Research 2007;77(1):81-112.
 36. Shute VJ. Focus on Formative Feedback. Review of Educational Research 2008;78(1),153-189.
 37. Dudek NL, Marks MB, Regehr G. Failure to Fail: The Perspectives of Clinical Supervisors. Acad Med 2005; 80(10 suppl): S84-S87.
 38. Holmboe ES., Hawkins RE, Huot SJ. Effects of Training in Direct Observation of Medical Residents' Clinical Competence. A randomized trial. Ann Intern Med 2004;140:874-881.
 39. Woehr DJ, Huffcutt AI. Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organizational Psychology 1994; 67(3):189-205.
 40. Bernardin HJ, Orban JA, Carlyle JJ. Performance ratings as a function of trust in appraisal and rater individual differences. Academy of Management Proceedings 19981;311-315.
 41. Longenecker CO, Gioia DA. Confronting the "politics" in performance appraisal. Business Forum 2000;25(3,4):17-23.
 42. Littlefield JH, Darosa DA, Paukert J, Williams RG, Klamen DL, Schoolfield JD. Improving resident performance data: numeric precision and narrative specificity. Acad Med 2005;80(5):489-495.

De auteurs:

*Drs. M.J.B. Govaerts is universitair docent, werkzaam bij de vakgroep Onderwijsontwikkeling en Onderwijsresearch, Faculty of Health, Medicine and Life Sciences, Universiteit Maastricht.**

*Dr. E. Driessen is universitair docent, werkzaam bij de vakgroep Onderwijsontwikkeling en Onderwijsresearch, Faculty of Health, Medicine and Life Sciences, Universiteit Maastricht.**

Dr. B. Verhoeven is kinderchirurg in het MUMC+ te Maastricht en maakt deel uit van de taakgroep vervolgopleiding van de vakgroep Onderwijsontwikkeling en Onderwijsresearch, Faculty of Health, Medicine and Life Sciences, Universiteit Maastricht.

Prof. dr. C.P.M. van der Vleuten is hoogleraar Onderwijskunde, voorzitter van de vakgroep Onderwijsontwikkeling en Onderwijsresearch, Faculty of Health, Medicine and Life Sciences, Universiteit Maastricht.

*Dr. H. Brackel is kinderarts in het Catharina Ziekenhuis te Eindhoven.***

*Drs. J. van Hoorn is kinderarts-fellow neonatologie in het MUMC te Maastricht.***

*Drs. R. van de Laar is arts in opleiding tot gynaecoloog in het MUMC+ te Maastricht.***

*Dr. J. Maas is gynaecoloog in het MUMC+ te Maastricht.***

*Prof. dr. S.G. Oei is gynaecoloog in het Maxima Medisch Centrum te Veldhoven.***

** en als onderwijskundig adviseur verbonden aan het kernteam In VIVO, OOR Zuid-Oost Nederland.*

*** allen maken deel uit van het kernteam In VIVO, OOR Zuid-Oost Nederland.*

Correspondentieadres:

M.J.B. Govaerts, Vakgroep Onderwijsontwikkeling en -research, FHML, Universiteit Maastricht, postbus 616, 6200 MD Maastricht.

E-mail: marjan.govaerts@educ.unimaas.nl

Belangenconflict: geen gemeld

Financiële ondersteuning: geen gemeld