

360-gradersfeedback i speciallægeuddannelsen

Ellen Holm¹, Kirsten Holm² & Jette Led Sørensen³

STATUSARTIKEL

1) Geriatrik Afdeling, Nykøbing Falster Sygehus
 2) Børne- og Ungeafdelingen, Nordsjællands Hospital, Hillerød
 3) Juliane Marie Centret for Børn, Kvinder og Reproduktion, Rigshospitalet

Ugeskr Læger
 2014;176:V05140295

Et voksende krav fra sundhedsmyndigheder og offentlighed om kontrol af kvaliteten i speciallægeuddannelsen førte i Danmark i 2003 til en reform. Et af de centrale punkter i reformen var indførelsen af de syv lægeroller (medicinsk ekspert, kommunikator, samarbejder, sundhedsfremmer, leder/administrator, akademiker og professionel). Et andet centralt punkt var, at en læges kompetenceniveau skulle være målbart, dvs. man skulle objektivt kunne bedømme, om speciallæger faktisk mestrede de opgaver, som de skulle varetage. Men hvordan skal man objektivt bedømme en læges evne til at samarbejde, kommunikere og forebygge? En måde at gøre dette på kunne være at indhente information fra lægens samarbejdspartnere og spørge om, hvordan de oplever lægens kompetence på disse områder. Denne metode kaldes 360-gradersevaluering/feedback (360F), på engelsk også *multisource feedback*. Idéen er, at man indsamler information om lægens kompetencer fra alle persongrupper omkring vedkommende, dvs. fra mange kilder. Metoden stammer oprindeligt fra den tyske værnemagt, som anvendte den til udvælgelse af officerer allerede omkring 1930. Op gennem 1980'erne blev metoden udbredt i private virksomheder i USA. De første rapporter om anvendelse til vurdering af læger kom sidst i 1980'erne [1]. Metoden er siden blevet meget udbredt som kompetenceevaluering i speciallægeuddannelsen og til recertificering af læger i bl.a. USA [2], Storbritannien [3] og Canada [4]. I Danmark anvendes metoden i langt de fleste lægelige specialer, og den er beskrevet i en nyligt publiceret rapport [5].

Vi vil i denne artikel gennemgå en del af litteraturen og de praktiske spørgsmål, der opstår i relation til implementering af 360F i speciallægeuddannelsen.

FREM GANGSMÅDE

360F gennemføres ved at fordele bedømmelseskemaer til de udvalgte bedømmere. Lægen, som skal bedømmes, udfylder også selv et skema (selvevaluering). Spørgeskemaerne uddeles i papirform eller elektronisk form. De udfyldte skemaer tilbagesendes til vejleder/uddannelsesansvarlig overlæge (UAO), som udarbejder en svarrapport til lægen. Alternativt genereres der en elektronisk rapport, som kan downloades af UAO/vejleder/anden feedbackgivende person [6]. Herefter afholdes der en feedbacksamtale, hvor lægen og feedbackgiveren drøfter resultatet og hvilke handlinger, dette skal medføre.

FORMÅL

Formålet med 360F er at indhente feedback fra lægens samarbejdspartnere. Denne feedback anvendes til vurdering af lægens kompetencer, og 360F er således både en feedback- og en kompetencevurderingsmetode. 360F er fundet at være en af de mest effektive kompetencevurderingsmetoder, når man vurderer, hvad der faktisk kan ændre lægers adfærd [7, 8]. Men der er en række mulige bias [9]. Metoden bør derfor ikke stå alene [10, 11] og bør ikke anvendes til en summativ (hvor man kan »dumpe« eller »bestå«) vurdering, men alene som en formativ (informerende) feedbackmetode [12]. Metodemæssigt adskiller gennemførelsen af 360F med et formativt sigte sig ikke fra 360F gennemført med summativt sigte. Men da en summativ evaluering har en absolut konsekvens, er der strengere krav til validitet. Derfor anbefaler Sundhedsstyrelsen, at 360F kun bruges formativt.

VALIDITET

Begrebet validitet dækker over meget forskelligt. Helt overordnet udtrykker validitet den grad, hvormed man med et instrument/en metode faktisk måler det, som man har ønsket at måle og omvendt ikke måler det, som man ikke har ønsket at måle. Man skelner mellem forskellige slags validitet herunder *face validity* (oplever brugerne metoden som umiddelbart fornuftig?), begrebsvaliditet (kan man vha. testen skelne mellem novicer og eksperter?) og prædiktiv eller fremtidig kriterievaliditet (kan man med instrumentet forudsige

FAKTABOKS

- ▶ Ved 360-gradersfeedback indhentes der struktureret feedback fra en uddannelseslæges samarbejdspartnere (bedømmerne).
- ▶ Bedømmerne kan vælges af uddannelseslægen selv eller af andre, f.eks. af den uddannelsesansvarlige overlæge, eller en afdeling kan have et fast sammensat bedømmerpanel. Lægen bedømmer også sig selv (selvevaluering).
- ▶ Det nødvendige antal af bedømmere er afhængigt af det anvendte spørgeskema samt af sammensætning og uddannelse af bedømmerne, men er 8-15, når der anvendes en skalamodel. En narrativ vurdering kan potentielt være nyttig også med færre bedømmere; viden herom er dog begrænset.
- ▶ Efter indhentning af svar fra bedømmerne afholdes en feedbacksamtale, typisk med hovedvejleder eller uddannelsesansvarlig overlæge.
- ▶ Feedbackgiver skal være trænet og have mulighed for supervision ved behov.

noget centralt vedrørende det fænomen, man ønsker målt?). Validitetsbegrebet er kort gennemgået i Sundhedsstyrelsens rapport om kompetencevurderingsmetoder [5].

Nogle mulige bias, som kan true validiteten af 360F, er varigheden af bedømmernes kendskab til lægen (længere varighed giver mere positiv bedømmelse), haloeffekt (når en bedømmer afgiver sine svar ud fra en generel positiv eller negativ opfattelse af lægen) [13] og mere eller mindre »skrappe« bedømmere. 360F har i de fleste publicerede studier en god *face validity*. I enkelte studier har man påvist, at metoden er valid, bedømt på muligheden for at skelne mellem mindre og mere erfarne læger [14]. I et enkelt longitudinelt studie har man påvist, at læger, som har været gennem 360F flere gange, har progression i deres kompetence-niveau [15].

UDVÆLGELSE AF BEDØMMERE

Ideen med 360F er, at bedømmerne skal have et så grundigt kendskab til uddannelseslægen, at de har mulighed for at afgive en bedømmelse, som baseres på egne erfaringer med lægen. Derfor har det været almindelig praksis, at uddannelseslægen selv vælger sine bedømmere. I nogle studier har man dog påvist, at det har givet et andet resultat, hvis lægen selv vælger bedømmerne, end hvis bedømmerne vælges af en tredjepart (f.eks. af den UAO) [16]. Det har betydning, om bedømmerne er yngre læge-kolleger (som bedømmer mildere) eller overlæger (som bedømmer skrappe) [16]. Desuden bedømmer oversygeplejersker skrap [17], mens det er uvist, hvordan andre faggruppers bedømmelser er.

For at bedømmelserne bliver brugbare er det vigtigt, at bedømmerne forstår baggrund og formål med metoden [14]. I et kvalitativt studie lod man bedømmere tænke højt, mens de scorede lægerne. Det viste sig, at bedømmerne meget ofte refererede til mere generelle karaktertræk hos den uddannelsessøgende i stedet for at bedømme ud fra konkrete observationer [18]. Dette er i modstrid med idéen i metoden og taler for, at et uddannet bedømmerpanel, som kan instrueres grundigt i metoden, kunne være en fordel. Når lægerne selv vælger bedømmere, kræver det et større antal bedømmere, som man ikke på alle mindre afdelinger råder over. Bruger man i stedet et trænet fast bedømmerpanel kan man formentlig få en valid bedømmelse med færre bedømmere [10, 11].

Skal bedømmerne være anonyme?

I de fleste modeller anvender man anonyme bedømmere ud fra den betragtning, at man på denne måde får mere ærlige svar. I enkelte modeller har man beskrevet anvendelse af ikkeanonyme bedømmere forstået på den måde, at bedømmerne underskriver deres vurde-

ring, og således er kendt af eller ved behov kan afsløres for den, der modtager svarene, f.eks. den UAO eller feedbackgivende vejleder [10, 11, 19, 20]. I den svarrapport, som udarbejdes til uddannelseslægen, fremgår bedømmernes navne ikke. Et argument for at bruge ikkeanonyme bedømmere er muligheden for, at vejleder/UAO kan opsøge en bedømmer og få uddybende information. Et andet argument er lægernes retssikkerhed. Dette skisma er beskrevet i en evalueringsrapport, som omhandler Region Nord-modellen [21] »Det er ikke indlysende, hvad der er etisk korrekt – at give mulighed for kritik uden at frygte negative reaktioner, eller at beskytte den evaluerede mod risikoen for ondsindet anonym kritik«. Det svarer helt til argumentationen i et andet studie »... a rater might score a trainee's behaviour as of major concern for personal, rather than professional reasons. Such instances should be discouraged by the lack of anonymity of the process and identifiable by the educational supervisor's investigations in such cases« [20]. Valget af anonyme versus ikkeanonyme bedømmere er således i høj grad en holdningssag.

Hvor mange bedømmere skal der anvendes?

I oversigtsartikler angives det ofte, at der skal være 8-15 bedømmere, antallet afhænger dog af fordeling på personalegrupper.

I nogle modeller anvender man spørgeskemaer, som er specifikt rettede mod bestemte grupper f.eks. sygeplejersker eller lægekolleger. Ved testning af en af disse modeller fandt man, at der skulle være otte lægekolleger og seks ikkelægelige kolleger, for at bedømmelsen kunne være valid [22]. I en del senere studier har man peget på, at sekretærer kun kan besvare et ret begrænset antal spørgsmål og derfor ikke skal indgå. I enkelte studier har man anvendt helt ned til tre-fire bedømmere [10, 11]. I den danske Region Nord-model anbefales fem personer fra hver af grupperne: overlæger/speciallæger, yngre læge-kolleger, plejepersonale og »andre«. Det anføres dog, at der fra gruppen af speciallæger/overlæger måske er behov for seks, fordi der i denne gruppe er det største frafald [21].

360-gradersfeedbackskemaet

Spørgsmålene i 360F-skemaet skal afspejle indholdet i specialets målbeskrivelse. I Danmark er 360F primært blevet brugt til bedømmelse af kompetencer inden for rollerne kommunikator, samarbejder, leder/administrator og professionel. Kompetencerne inden for disse roller må formodes at være nogenlunde ens for alle læger uafhængigt af speciale. Man kan dog også forestille sig, at man i et speciale ønsker at vurdere mere specifikke kompetencer inden for disse roller f.eks. evnen til at kommunikere med særlige patientgrupper såsom børn eller ældre. I nogle udenlandske modeller og i et enkelt dansk studie er der inkluderet spørgsmål, hvor

medicinske ekspertkompetencer skal evalueres, hvilket har ført til udarbejdelse af forskellige skemaer til forskellige specialer [23-25].

Et mere kontroversiel spørgsmål i Danmark er, om man nødvendigvis skal anvende en skalamodel. I en skalamodel beskrives f.eks. en positiv adfærd, og bedømmerne skal svare på, hvor ofte den uddannelsessøgende viser denne adfærd på en skala fra f.eks. »næsten aldrig« til »altid« [21]. I et enkelt studie har man som alternativ til skalamodellen testet en rent narrativ model, hvor bedømmerne vurderer adfærd alene med et beskrivende åbent udsagn [26]. De fleste skemaer bruger en blanding, hvor der er en skalabedømmelse og mulighed for åbne kommentarer. Men da bedømmere typisk foretrækker at sætte et kryds, er der ved en blandet model en risiko for, at det giver få brugbare kommentarer. Et kvalitativt studie af de åbne kommentarer i et skalabaseret 360F-skema viste, at de fleste kommentarer refererede til generelle karaktertræk og ikke var konkrete og/eller handlingsanvisende [27]. De læger, som bedømmes, foretrækker en narrativ model [26] eller lægger meget vægt på brugbare åbne kommentarer [28, 29]. Valget af skematype hænger sammen med evalueringens formål. Kvaliteten af de åbne kommentarer er væsentlig, hvis evalueringen skal bidrage til lægens videre udvikling. Hvis formålet derimod er at udpege læger, som ikke lever op til kravene, er en ren skalamodel tilstrækkelig [30]. I de ni internmedicinske specialer har man ved den seneste revision af målbeskrivelse valgt en rent narrativ model. Det er ukendt, hvor mange bedømmere man skal bruge, når man anvender et rent narrativt skema, idet dette ikke

kan beregnes med de psykometriske metoder, som er anvendt til vurdering af, hvor mange bedømmere der skal bruges ved en skalamodel.

SVARRAPPORT OG FEEDBACKSAMTALE

Når der er indhentet resultater fra alle bedømmere, samles resultaterne i en svarrapport. En sådan svarrapport kan bestå af et enkelt ark eller være en mere omfattende rapport. Et eksempel på en enkelt rapport er et resultatark, hvor der for hvert spørgsmål vises gennemsnitsresultat fra bedømmere, resultat fra selvurdering og evt. et gennemsnitsresultat for gruppen af uddannelsessøgende, vist som søjler i tre farver, se eksempel i **Figur 1**. Et eksempel på en mere omfattende rapport ses i den danske Region Nord-model [21]. Feedbacksamtalen afholdes typisk af hovedvejleder eller UAO, men kan også afholdes af andre, når blot disse er uddannede til opgaven. Feedbacksamtalen har en afgørende betydning for, om evalueringen får noget resultat, især hvis der er stor diskrepans mellem selvurdering og bedømmernes vurdering. Da feedbacksamtalen er af afgørende betydning, bør læger, der varetager denne, være trænet i feedbackgivning, og de bør have mulighed for supervision og støtte ved behov. Mange hospitaler råder over organisationspsykologer, som vil kunne give en sådan støtte.

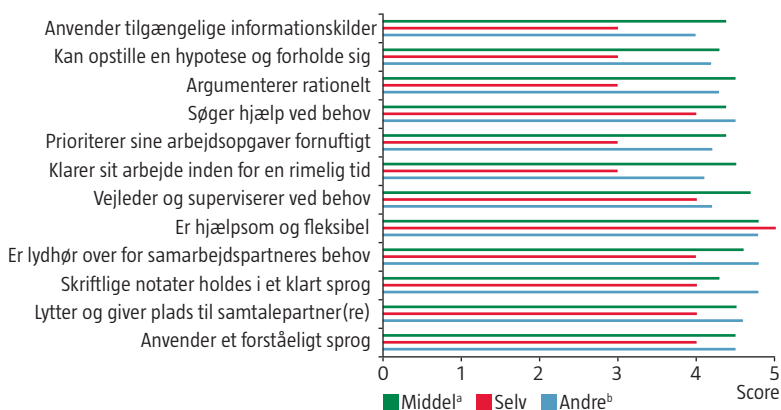
SAMMENFATNING OG PERSPEKTIVERING

360F er en valid og effektiv feedback-/kompetencevurderingsmetode. Der er dog en række kendte bias som kompromitterer validiteten, og der er stadig mange åbne spørgsmål, som skal overvejes, f.eks. valg af skematype, metode til udvælgelse af bedømmere og træning af bedømmere. Der er således fortsat brug for yderligere videnskabelige undersøgelser af forskellige modeller. Den narrative model kan potentielt anvendes i små miljøer, idet verbal feedback fra få personer også kan være brugbar.

For de modeller, som anvendes i Danmark, vil bl.a. følgende forskningsspørgsmål være relevante: 1) Hvor mange bedømmere svarer i gennemsnit fra hver enkelt personalegruppe? 2) Kan man se et mønster i relation til hvor skrappe bedømmerne er? – Er overlæger skrappe end yngre læger, og hvordan med sygeplejersker og »andre«? 3) Hvor hyppigt er der åbne kommentarer, og hvordan er kvaliteten af disse? 4) Kan man konstatere progression hos læger, som er bedømt flere gange? 5) Hvor mange bedømmere skal der bruges, når man anvender et rent narrativt skema? 6) Opleves evalueringen som mere/mindre brugbar af de evaluerede læger, når der anvendes et rent narrativt skema? 7) Opleves evalueringen som mere/mindre brugbar af de evaluerede læger, hvis man anvender et uddannet fast bedømmerpanel frem for bedømmere valgt af lægen som evalueres?

FIGUR 1

Eksempel på svarrapport for en introduktionslæge i intern medicin (udvalgte spørgsmål fra 2007-målbeskrivelsen). Ud over dette grafiske billede vedhæftes i svarrapporten i anonymiseret form eventuelle åbne kommentarer fra bedømmerne.



a) Angiver middelværdi for alle introduktionslæger der har modtaget 360-gradersfeedback i afdelingen inden for en nærmere angivet periode.

b) Angiver middelscore fra bedømmerne.

KORRESPONDANCE: Ellen Holm. E-mail: ellenholm@dadlnet.dk

ANTAGET: 25. september 2014

PUBLICERET PÅ UGESKRIFTET.DK: 1. december 2014

INTERESSEKONFLIKTER: Forfatterens ICMJE-formularer er tilgængelige sammen med artiklen på Ugeskriftet.dk

Artiklen bygger på en større litteraturgennemgang, den fulde litteraturliste kan rekvireres fra forfatterne.

I artikel Ugeskr Læger 2014;176:V05140295 er der den 28.9.2015 foretaget følgende rettelse: Litteraturlisten er korrigeret vedr. webadresser og nummerering af samtlige referencer.

LITTERATUR

1. Carline JD, Wenrich M, Ramsey PG. Characteristics of ratings of physician competence by professional associates. *Eval Health Prof* 1989; 12:409-23.
2. American Board of Medical Specialties. Maintenance of Certification (MOC). www.abms.org/board-certification/steps-toward-initial-certification-and-moc/ (20. nov 2013).
3. General Medical Council. Supporting information for appraisal and revalidation. www.gmc-uk.org/static/documents/content/RT_-_Supporting_information_for_appraisal_and_revalidation_-_DC5485.pdf (20. nov 2013).
4. College of Physicians & Surgeons of Alberta. Physician Achievement Review (PAR) Program. www.cpsa.ca/Services/PARprogram.aspx (20. nov 2013).
5. Sundhedsstyrelsen. Kompetencevurderingsmetoder – en oversigt. <http://sundhedsstyrelsen.dk/publ/Publ2013/08aug/KompetenceVurdMetoder.pdf> (20. nov 2013).
6. Malling B. Implementering af 360 graders feedback i den lægelige videreuddannelse på Århus Universitetshospital, Skejby. www.auh.dk/siteassets/om-auh/administrationen/hr/hr-aarhus-filer/hr---laegelig-videreuddannelse/pdf/360/rapporter/kul-afrap360impl_280710_fin_indsendt.pdf (7. okt 2014).
7. Brinkman WB, Geraghty SR, Lanphear BP et al. Effect of multisource feedback on resident communication skills and professionalism: a randomized controlled trial. *Arch Pediatr Adolesc Med* 2007;161:44-9.
8. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ* 2010;341:c5064.
9. Campbell JL, Roberts M, Wright C et al. Factors associated with variability in the assessment of UK doctors' professionalism: analysis of survey results. *BMJ* 2011;343:d6212.
10. Ben-David MF, Snadden D, Hesketh A. Linking appraisal of PRHO professional competence of junior doctors to their education. *Medical Teacher* 2004;26:63-70.
11. Hesketh EA, Anderson F, Bagnall GM et al. Using a 360 degree diagnostic screening tool to provide an evidence trail of junior doctor performance throughout their first postgraduate year. *Medical Teacher* 2005;27:219-33.
12. Campbell J, Hill J, Hobart J et al. GMC Multi-Source Feedback Study. www.gmc-uk.org/executive_summary_of_research.pdf_48212169.pdf (20. nov 2013).
13. Palmer JK, Loveland JM. The influence of group discussion on performance judgments: rating accuracy, contrast effects, and halo. *J Psychol* 2008;142:117-30.
14. Archer J, Norcini J, Southgate L et al. Mini-PAT (peer assessment tool): a valid component of a national assessment programme in the UK? *Adv Health Sci Educ* 2006;13:181-92.
15. Violato C, Lockyer JM, Fidler H. Changes in performance: a 5-year longitudinal study of participants in a multi-source feedback programme. *Med Educ* 2008;42:1007-13.
16. Archer JC, McAvoy P. Factors that might undermine the validity of patient and multi-source feedback. *Med Educ* 2011;45:886-93.
17. Bullock AD, Hassell A, Markham WA et al. How ratings vary by staff group in multi-source feedback assessment of junior doctors. *Med Educ* 2009;43:516-20.
18. Mazor KM, Canavan C, Farrell M et al. Collecting validity evidence for an assessment of professionalism: findings from think-aloud interviews. *Acad Med* 2008;83(suppl 10):S9-S12.
19. Archer J, McGraw M, Davies H. Assuring validity of multisource feedback in a national programme. *Arch Dis Child* 2010;95:330-5.
20. Whitehouse A, Hassell A, Wood L et al. Development and reliability testing of TAB a form for 360 degree assessment of Senior House Officers' professional behaviour, as specified by the General Medical Council. *Med Teach* 2005;27:252-8.
21. Seeberg J. 360°-evaluering i speciallæge-uddannelse, 2006. www.laeger.dk/portal/pls/portal/!PORTAL.wwwpob_page.show?_docname=2503359.PDF (7. okt 2014).
22. Hall W, Violato C, Lewkonian R et al. Assessment of physician performance in Alberta: the Physician Achievement Review. *CMAJ* 1999; 161:52-7.
23. Violato C, Lockyer JM, Fidler H. Assessment of pediatricians by a regulatory authority. *Pediatrics* 2006;117:796-802.
24. Violato C, Lockyer JM, Fidler H. Assessment of psychiatrists in practice through multisource feedback. *Can J Psychiatry* 2008;53:525-33.
25. Allerup P, Aspegren K, Ejlersen E et al. Use of 360-degree assessment of residents in internal medicine in a Danish setting: a feasibility study. *Med Teach* 2007;29:166-70.
26. Overeem K, Lombarts MJ, Arah OA et al. Three methods of multi-source feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach* 2010;32:141-7.
27. Canavan C, Holtman MC, Richmond M et al. The quality of written comments on professional behaviors in a developmental multisource feedback program. *Acad Med* 2010;85(suppl 10):S106-S109.
28. Cohen SN, Farrant PB, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol* 2009;161:34-9.
29. Burford B, Illing J, Kergon C et al. User perceptions of multi-source feedback tools for junior doctors. *Med Educ* 2010;44:165-76.
30. Holmboe ES, Ross K. Commentary: realizing the formative potential of multisource feedback in regulatory-based assessment programs. *Acad Med* 2012;87:1657-9.