

Kompetencevurde- ringsmetoder

- En oversigt

Udgave 19.04.2013

Titel (præcis samme som forside)

© Sundhedsstyrelsen, 2013. Publikationen kan frit refereres med tydelig kildeangivelse.

Sundhedsstyrelsen
Axel Heides Gade 1
2300 København S

URL: <http://www.sst.dk>

Emneord: amning, akkreditering, bosteder, it-sikkerhedspolitik, informationssikkerhed, sygehuse, ... (max fem til otte)

Sprog: Dansk

Kategori: (Regler eller Faglig rådgivning – tjek taksonomien på Sundnet)

Version: 1.0 (eller senere)

Versionsdato: (DD.MM.ÅÅÅÅ)

Format: pdf

Udgivet af Sundhedsstyrelsen, (måned og år).

Elektronisk ISBN: (Nummeret bliver tilføjet automatisk, når Word-dokumentet afleveres til Kommunikation)

Den trykte versions ISBN: (Nummeret bliver tilføjet automatisk, når Word-dokumentet afleveres til Kommunikation)

ISSN-nr: (Nummeret bliver tilføjet automatisk, når Word-dokumentet afleveres til Kommunikation)

Indhold

1	Forord	5
2	Introduktion	7
3	Generel del	9
3.1	Definition af kompetence	9
3.2	Formål med kompetencevurdering	9
3.3	Ansvar for og opfølgning på kompetencevurdering	10
3.4	Vurderingsskalaer	11
3.4.1	Definition af vurderingsskalaer	11
3.4.2	Skalaers anvendelighed	14
3.4.3	Fordele ved skalaer	15
3.4.4	Ulemper ved skalaer	16
3.4.5	Skalaers validitet, pålidelighed og gennemførlighed	17
3.5	Validering af metoder til kompetencevurdering - psykometriske /statistiske begreber	17
3.5.1	Pålidelighed (reliabilitet)	17
3.5.2	Validitet	18
4	Direkte observation	20
4.1	Objektiv vurdering af tekniske færdigheder	20
4.1.1	Definition af OSATS	20
4.1.2	OSATS anvendelighed	20
4.1.3	Fordele ved OSATS	21
4.1.4	Ulemper ved OSATS	21
4.1.5	OSATS validitet, pålidelighed og gennemførlighed	21
4.1.6	Praktisk anvendelse af OSATS	22
4.2	Objektiv vurdering af ikke-tekniske færdigheder	22
4.2.1	Definition af ikke-tekniske færdigheder og af NOTSS/ ANTS	22
4.2.2	NOTSS/ ANTS anvendelighed	23
4.2.3	Fordele ved NOTSS/ ANTS	23
4.2.4	Ulemper ved NOTSS/ ANTS	23
4.2.5	NOTSS/ ANTS validitet, pålidelighed og gennemførlighed	23
4.2.6	Praktisk anvendelse af NOTSS/ ANTS	24
4.3	Kompetencekort	24
4.3.1	Definition af kompetencekort	24
4.3.2	Kompetencekorts anvendelighed	24
4.3.3	Fordele ved kompetencekort	25
4.3.4	Ulemper ved kompetencekort	25
4.3.5	Kompetencekorts validitet, pålidelighed og gennemførlighed	25
4.3.6	Praktisk anvendelse af kompetencekort	26
4.4	Mini Clinical Evaluation Exercise (mini-CEX)	26
4.4.1	Definition af mini-CEX	26
4.4.2	Mini-CEX anvendelighed:	26
4.4.3	Fordele ved mini-CEX	26
4.4.4	Ulemper ved mini-CEX	26
4.4.5	Mini-CEX validitet, pålidelighed og gennemførlighed	26
4.4.6	Praktisk anvendelse af mini-CEX	27
5	Helhedsvurderinger	28
5.1	Generelle vurderinger	28
5.1.1	Definition af generelle vurderinger	28
5.1.2	Generelle vurderingers anvendelighed	28
5.1.3	Fordele ved generelle vurderinger	29
5.1.4	Ulemper ved generelle vurderinger	29
5.1.5	Generelle vurderingers validitet, pålidelighed og gennemførlighed	29

5.1.6	Praktisk anvendelse af generelle vurderinger	29
5.2	360 graders feedback	30
5.2.1	Definition af 360 graders feedback	30
5.2.2	360 graders feedbacks anvendelighed	30
5.2.3	Fordele ved 360 graders feedback	31
5.2.4	Ulemper ved 360 graders feedback	31
5.2.5	360 graders feedbacks validitet, pålidelighed og gennemførlighed	31
5.2.6	Praktisk anvendelse af 360 graders feedback	31
6	Retrospektive metoder	33
6.1	Audit	33
6.1.1	Definition af audit	33
6.1.2	Audits anvendelighed	33
6.1.3	Fordele ved audit	34
6.1.4	Ulemper ved audit	34
6.1.5	Audit, validitet, pålidelighed og gennemførlighed	34
6.1.6	Praktisk anvendelse af audit	34
6.2	Refleksive rapporter	35
6.2.1	Definition af refleksive rapporter	35
6.2.2	Refleksive rapporters anvendelighed	36
6.2.3	Fordele ved refleksive rapporter	36
6.2.4	Ulemper ved refleksive rapporter	36
6.2.5	Refleksive rapporter validitet, pålidelighed og gennemførlighed	36
6.2.6	Praktisk anvendelse af refleksive rapporter	37
6.3	Casebaseret diskussion	37
6.3.1	Definition af casebaseret diskussion	37
6.3.2	Casebaseret diskussion anvendelighed	37
6.3.3	Fordele ved casebaseret diskussion	38
6.3.4	Ulemper ved casebaseret diskussion	38
6.3.5	Casebaseret diskussion validitet, pålidelighed og gennemførlighed	38
6.3.6	Praktisk anvendelse af casebaseret diskussion	38
7	Kompetencevurdering med anvendelse af simulation	40
7.1	Simulation	40
7.1.1	Definition af simulation	40
7.1.2	Simulations anvendelighed	40
7.1.3	Fordele ved simulation	41
7.1.4	Ulemper ved simulation	41
7.1.5	Simulations validitet, pålidelighed og gennemførlighed	41
7.1.6	Praktisk anvendelse af simulation	41
8	Konklusion	42
9	Anbefalinger og perspektivering	44
10	Ordliste	46
11	Referencer	49
12	Bilag	

1 Forord

Kompetencevurderingen af den uddannelsessøgende læge er en central proces i speciallægeuddannelsen, både som formativ, dialogbaseret proces i det daglige arbejde og som en summativ, konsekvenshavende bedømmelse af om speciallægeuddannelsens mål nås.

I Status og Perspektiverings rapporten¹ blev det foreslået, at nedsætte en arbejdsgruppe bestående af medicinsk-pædagogisk uddannede læger til at udarbejde en oversigt over relevante kompetencevurderings-metoder. Et sådan analysearbejde skulle foretages i regi af Det Nationale Råd for Lægers Videreuddannelse.

Arbejdsgruppens opgave var at udarbejde en systematisk oversigt over metoder til kompetencevurdering, der fremadrettet kan anvendes af de videnskabelige selskaber i forbindelse med udarbejdelse af den specialespecifikke del af målbeskrivelsen og en konkret anbefaling af metode til vurdering af specifikke kompetencer. Rapporten supplerer således Rapporten om de 7 lægeroller², hvori rollerne beskrives og overordnede kompetencer uddybes på individ-, organisations- og samfundsniveau.

I nærværende rapport gives en introduktion til begrebet kompetencevurdering og en diskussion af betydende elementer for, at dette kan gennemføres optimalt. En systematisk oversigt over velbeskrevne og validerede metoder gives. Det skal dog pointeres, at der ikke eksisterer validerede metoder til vurdering af alle kompetencer indenfor alle specialer, hvilket gør det vigtigt for specialerne at udvikle og validere egne metoder til vurdering af kompetence. Dette kræver en betydelig indsats, og et nationalt samarbejde er nødvendigt såvel indenfor det enkelte speciale som mellem specialer. I den forbindelse vil den pædagogisk udviklende funktion i de respektive videreuddannelsesregioner kunne spille en central rolle.

Denne rapport kan endvidere anvendes af uddannelsesansvarlige overlæger og vejledere til at få et overblik over metodernes anvendelighed.

Rapporten er udarbejdet af repræsentanter fra de 3 uddannelsesregioner, heriblandt adskillige postgraduate kliniske lektorer. Arbejdsgruppen sammensætning:

Doris Østergaard, Postgraduat Klinisk Lektor, overlæge, dr.med. Formand for Arbejdsgruppen udpeget af Sundhedsstyrelsen.

Kari Mikines, Postgraduat Klinisk Lektor, uddannelsesansvarlig overlæge, dr.med. Udpeget af Videreuddannelsesregion Øst.

Jette Led Sørensen, Postgraduat Klinisk Lektor, uddannelsesansvarlig overlæge, MMed. Udpeget af Universiteterne.

Kirsten Bsted. Overlæge M.Sc. Udpeget af Videreuddannelsesregion Syd.

Bente Sørensen, Postgraduat Klinisk Lektor, overlæge. Udpeget af Videreuddannelsesregion Syd.

Hans Ri Jørgensen, Postgraduat Klinisk Lektor, ledende overlæge. Udpeget af Vi-

dereuddannelsesregion Syd.

Søren Prins. Praktiserende læge, uddannelseskoordinator. Udpeget af Lægeforeningen.

Gitte Valsted Eriksen, Postgraduat Klinisk Lektor, MPM, uddannelsesansvarlig overlæge, ph.d. Udpeget af Videreuddannelsesregion Nord.

Ulla Bartels, Lektor, ledende overlæge. Udpeget af Videreuddannelsesregion Nord.

Roar Maagard, Lektor, praktiserende læge. Udpeget af Lægevidenskabelige Selskaber (LVS).

Bente Malling, Postgraduat Klinisk Lektor, uddannelseskoordinerede overlæge, ph.d., MHPE. Udpeget af Videreuddannelsesregion Nord.

Anne Grethe Mølbak, uddannelsesansvarlig overlæge. Udpeget af Universiteterne

Ellen Holm, Postgraduat Klinisk Lektor, overlæge. Udpeget af Videreuddannelsesregion Øst.

Sekretariat:

Karen Geismar, Sundhedsstyrelsen

2 Introduktion

Kompetencevurderingen af den uddannelsessøgende læge er et af de bærende elementer i Fremtidens speciallæge, Betænkning fra Speciallægekommisionen fra 2001³.

I Status og Perspektiveringsrapporten¹ beskrives, at specialeselskaberne i store træk har valgt metoder til kompetencevurdering af den enkelte kompetence i målbeskrivelsen ud fra de metoder speciallægekommisionens betænkning anbefalede. Metoderne til kompetencevurdering er i mange specialer videreudviklet på baggrund af evidensbaseret medicinsk-pædagogisk viden. Denne positive udvikling skyldes et dynamisk samspil mellem de postgraduate kliniske lektorer og de videnskabelige selskabers uddannelsesudvalg¹. Interessen for formaliseret uddannelse i at varetage medicinsk uddannelse blandt uddannelsesansvarlige er stigende. En markant stigning ses også i antal forskningsprojekter, publikationer, masterafhandlinger og ph.d. afhandlinger inden for området.

Angivelse af kompetencevurderingsmetoder er ikke i sig selv en sikkerhed for at disse anvendes konsekvent i den kliniske hverdag¹. Trods samme ordlyd anvendes kompetencevurderingsmetoderne forskelligt i de forskellige specialer, men forskelligheder ses også indenfor de enkelte afdelinger i samme speciale. Det er tydeligt, at specialer, som tidligt fik uddannet personer eller fik assistance fra personer med viden om/erfaring vedrørende medicinsk uddannelse er kommet først i gang med implementering af kompetencevurdering.

En rapport viser dog, at kun en tredjedel af de uddannelsesgivende og en fjerdedel af de uddannelsessøgende oplyser om anvendelse af strukturerede kompetencevurderingsmetoder⁴. En stor spredning ses blandt specialerne i forhold til valg af kompetencevurderingsmetoder, kvaliteten af kompetencevurderingen og af vejledernes kendskab og holdning til kompetencevurdering. Rapporten konkluderer endvidere, at kompetencevurderingen overvejende relaterede sig til den medicinske ekspert rolle. De uddannelsessøgende giver også udtryk for manglende struktureret tilbagemelding (feedback). Dette kan ses som udtryk for, at vejledere/bedømmere ikke forstår betydningen af at foretage en sufficient kompetencevurdering og/eller ikke er blevet uddannet i at anvende metoderne. En uddannelse af bedømmere er vist at have en effekt⁵ og flere specialer har valgt at træne bedømmere på speciale-specifikke kurser. Manglende indarbejdelse af en konsekvent kompetencevurderingskultur er en af de væsentligste barrierer for en god kompetencevurdering.

Formål

Formålet med denne rapport er at præsentere en oversigt over metoder til kompetencevurdering, der kan anvendes af de videnskabelige selskaber i forbindelse med udarbejdelsen af målbeskrivelse og portefølje. I rapporten beskrives metoder til formativ vurdering i form af direkte observation, helhedsvurdering af individets progression samt retrospektive metoder som illustrerer refleksion. Samlet set er kompetencevurderingen en summativ, konsekvenshavende bedømmelse af om speciallægeuddannelsens mål nås.

Rapportens opbygning

Rapporten består af en generel introduktion til kompetencevurdering (afsnit 3) med en beskrivelse af hvorfor kompetencevurdering bør foretages, hvad der kræves af metoderne, en beskrivelse af betydningen af hvilken typer skala der anvendes samt en introduktion til validitet og pålidelighed. Herefter følger tre overordnede grupper af kompetencevurderingsmetoder: Direkte observation (afsnit 4), helhedsvurderinger (afsnit 5) og retrospektive metoder (afsnit 6).

Inden for hver gruppe er forskellige redskaber beskrevet efter samme skabelon, som er inspireret af ”The Good Assessment Guide”⁶: definition, anvendelighed, fordele og ulemper, validitet, pålidelighed og gennemførlighed samt praktiske eksempler på anvendelse. I afsnit 7 beskrives eksempler på hvordan kompetencevurdering kan foretages ved hjælp af simulation/simulatorer. En konklusion (afsnit 8) efterfulgt af anbefalinger og perspektivering (afsnit 9) afslutter rapporten, som endelig indeholder et referenceafsnit og et bilagsafsnit med talrige eksempler på kompetencevurderingsredskaber, som anvendes i forskellige specialer i Danmark.

Rapporten har visse steder valgt at bibeholde en engelsk terminologi, således at læseren lettere kan finde den relevante litteratur. En ordliste er udarbejdet som bilag.

3 Generel del

3.1 Definition af kompetence

I såvel rapporten om De 7 lægeroller² og denne rapport defineres begrebet kompetence, som lægens evne til at udføre de opgaver og udfylde de roller, der forventes på et givet tidspunkt i lægens professionelle udvikling⁷. Kompetencebegrebet omhandler viden, færdigheder og holdninger. Begrebet kompetence er beskrevet i detaljer i rapporten omhandlende de 7 lægeroller^{2, 8}.

I specialernes målbeskrivelse indgår en beskrivelse af kompetencemål, læringsmetoder samt en beskrivelse af hvilke metoder, der kan vælges til kompetencevurdering.

3.2 Formål med kompetencevurdering

Kompetencevurderingen af den uddannelsessøgende læge er et af de bærende elementer i Speciallægekommisionens betænkning ”Fremtidens speciallæge”³. Kommissionen anbefalede en fortløbende formativ kompetencevurdering af de uddannelsessøgende læger. En sådan formativ kompetencevurdering bør altid efterfølges af feedback. Endelig danner det samlede antal vurderinger baggrund for en summativ, konsekvenshavende bedømmelse af, om speciallægeuddannelsens mål nås.

I Danmark opnås speciallægeanerkendelse således, når alle obligatoriske uddannelseselementer er vurderet og godkendt. Kompetencevurderingen erstatter dermed en egentlig specialist eksamen. Kompetencevurderingen skal derfor foretages med høj kvalitet for at sikre, at lægen har de nødvendige kompetencer. Dette indebærer, at kompetencevurderingsmetoder er validerede, at bedømmerne er uddannede i at foretage kompetencevurdering samt at de nødvendige rammer for at foretage kompetencevurdering er til stede.

Kompetencevurderingen omfatter viden, færdigheder og holdninger – dvs. både kognitive og metakognitive færdigheder. Kompetencevurderingen foregår praksis nært – i relation til en arbejdsituation. Den enkelte kompetencevurdering kan derfor omfatte flere roller. Dette er i overensstemmelse med nyere evidens og tanker om kompetencevurdering. Flere specialer ser delelementer af kompetencevurderingen som et kørekort og dermed en tilladelse til at kunne udføre en opgave selvstændigt og har relateret dette til milepæle/delmål i uddannelsen. Som eksempel er der visse kompetencer en læge i introduktionsstilling i anæstesi skal mestre inden lægen kan gå i vagt. Dette er i overensstemmelse med begrebet, Entrustable Professional Activities (EPA)⁹ som udtrykker tillid til, at en læge kan udføre en aktivitet på egen hånd.

Et systematisk review har vist, at der ikke er evidens for, at de eksisterende kompetencevurderingsredskaber kan vurdere lægeroller uafhængigt af hinanden¹⁰. Forfatterne anbefaler derfor, at rollerne anvendes til at guide og koordinere kompetencevurderingen frem for at forsøge at måle rollen direkte. Dette er i overensstemmelse med EPA begrebet. Overordnet anbefales anvendelse af en bred vifte af kompetencevurderingsmetoder til at vurdere den enkelte rolle og at adressere en given rolle på forskellige måder¹⁰. Dette gælder også for et tema som patientsikkerhed, der internationalt ønskes integreret mere i roller og kompetencer¹¹. Dvs. en rolle kan godt

indgå i flere forskellige kompetencevurderinger – f.eks. kan professionalisme godt indgå både i en direkte observation og i en generel vurdering. På samme måde kan flere roller godt kompetencevurderes samtidigt. Et eksempel på dette er anvendelse af direkte observation, hvor såvel viden og færdigheder kan vurderes såfremt der anvendes en sammensat skala (f.eks. kompetencekort).

Nationalt anbefales det, at et givent speciale anvender de samme kompetencevurderingsmetoder, så en uddannelsessøgende læge kompetencevurderes ved hjælp af de samme redskaber i de 3 videreuddannelsesregioner. Dette reducerer udviklings- og valideringsarbejdet og erfaringer med den givne vurderingsmetode kan samles nationalt.

Formålet med kompetencevurderingen er også at få et helhedsperspektiv af den enkelte uddannelsessøgende – og at vurdere udvikling af lægen over tid. Kompetencevurderingen foretages primært for at styrke den enkelte læges udvikling, men muligheden for at identificere uhensigtsmæssige forløb er også væsentlig.

Betydning af kompetencevurdering for læring

Kompetencevurderingen er i meget høj grad styrende for læring, hvorfor viden om denne, etablering af relevante metoder og indlejring af en proaktiv, positiv kompetencevurderingskultur i de uddannelsesgivende afdelinger og praksis, er en forudsætning for en velfungerende speciallægeuddannelse. Såfremt det potentiale, som kompetencevurdering har til at kunne medvirke til at optimere læringen, skal udnyttes er det nødvendigt at både den uddannelsessøgende og de uddannelsesgivende personer forstår hvorfor kompetencevurdering er vigtig at den skal udføres med en vis kvalitet for at give mening og at udfald af vurderingen har en konsekvens for opgaveløsningen. Læringsudbyttet er afhængig af om bedømmeren er i stand til at give konstruktiv feedback og medvirke til at skabe refleksion hos den uddannelsessøgende omkring egen kompetence. Enhver evaluering skal efterfølges af feedback.

3.3 Ansvar for og opfølgning på kompetencevurdering

Den uddannelsesansvarlige overlæge har sammen med afdelingsledelsen det overordnede ansvar for videreuddannelsen. Uddannelsesprogrammet angiver, hvornår og hvordan de enkelte kompetencer kan vurderes. Den uddannelsessøgende læge siger til, når vedkommende er klar til kompetencevurdering og har også et ansvar for at kompetencevurderingen planlægges, så begge parter er forberedt på situationen. Vurderingen bør optimalt foretages af bedømmere, som er oplært i at anvende den givne kompetencevurderingsmetode og som kender til risiko for bedømmerbias mv¹². Hvilke speciallæger, som foretager kompetencevurdering besluttet lokalt og er afhængigt af lokal kontekst og praktiske forhold. Vurderingen kan foretages af andre end seniorer læger, men den skal foretages af en, der er mere kompetent inden for et givet område og som er på et højere uddannelsesstrin end den, der vurderes. Dog indgår andre faggruppers vurdering i en 360 graders feedback Vurderingen skal foretages så objektivt og omhyggeligt som muligt. For at øge objektivitet og pålidelighed anvendes en vifte af forskellige metoder og vurderingerne foretages flere gange og af forskellige bedømmere. Vurderingerne kan foretages op mod ekspertrollen (endemål) eller svarende til, hvad den uddannelsessøgende forventes at kunne på et givet tidspunkt (delmål/milepæle). Forankringen og organisering af hvornår kompetencevurderinger foretages i uddannelsesforløbet og hvem, der foretager denne, er afdelingens ansvar som ofte er uddelegeret til den uddan-

nelsesansvarlige overlæge. Kompetencevurderingsprogram planlægges i sammenhæng med uddannelsesprogram.

Resultat af kompetencevurderingen anvendes i vejledersamtalerne, og individuelle uddannelsesplaner for den fortsatte udvikling baseres på disse. Det skal pointeres, at de kompetencevurderinger, som er direkte observationer skal foretages i klinikken og underskrifterne gives umiddelbart efter. Godkendelse og attestering af opnåede mål sker ved vurdering i klinikken – ikke ved de formelle vejledersamtaler.

Der er forskel på tidsforbrug i forbindelse med forberedelse og anvendelse af de forskellige metoder – generelt bør man sikre, at det er gennemførligt at foretage kompetencevurderingen på en given måde inden den implementeres.

Kompetencevurderingerne opbevares af den uddannelsessøgende læge i hans/hendes portefølje, som bringes videre til næste afdeling i forløbet og præsenteres for hovedvejlederen / tutor til samtalerne. Porteføljen skal hjælpe den uddannelsessøgende til at udvikle sig over tid.

3.4 Vurderingsskalaer

De kompetencevurderingsmetoder, der anvendes i den lægelige videreuddannelse, indeholder i langt de fleste tilfælde en række observationspunkter, der vurderes ud fra en skala af en eller flere bedømmere. Skalaer indgår dels i procedurespecifikke checklister f.eks. i OSATS, NOTSS (se afsnit 4.1 og 4.2) og i helhedsvurderinger f.eks. i 360 graders feedback (se afsnit 5.2). Vurderingsskalaen er et vigtigt element i kompetencevurderingen, hvorfor det er nødvendigt at overveje om den valgte skala giver svar på det som ønskes vurderet, om bedømmere har det nødvendige kendskab til skalaen og om skalaen skal anvendes til summativ eller formativ evaluering. I dette afsnit gennemgås forskellige typer af skalaer og der gives eksempler på fordele, ulemper samt anvendelighed.

3.4.1 Definition af vurderingsskalaer

Som led i en kompetencevurdering anvendes en skala til vurderingen af den uddannelsessøgende læges færdigheder og eller adfærd. Der anvendes to typer skalaer; dikotomiske eller flertrinsskalaer.

I valget af, hvilken type af skala, der er den mest optimale til en given kompetencevurderingsmetode, skal der tages hensyn til formålet med at anvendelsen af skalaen:

- ved vurdering af ”kan/kan ikke” anvendes dikotomiske skalaer f.eks. ”ja/nej.
- ved vurderinger, hvor der skal være mulighed for mere end 2 svarmuligheder for hvert spørgsmål anvendes en flertrinsskala. I en flertrinsskala kan yderpunkterne og eventuelt midterpunkter være defineret og beskrevet således at de niveauer, der ønskes vurderet er tydelige og passer til de specifikke spørgsmål/områder, der ønskes vurderet. De enkelte trin kan være relateret til, hvor hyppigt den uddannelsessøgende læge udviser en ønsket adfærd/færdighed eller være relateret til et kompetenceniveau.

Valget af skala afhænger ligeledes af om det er simple procedurer eller komplekse færdigheder og adfærd, der skal vurderes. Flertrinsskalaer anvendes til at vurdere mere komplekse arbejdsituationer og kompetencer, som indeholder flere roller.

- dikotomiske skalaer
 - ”ja/nej”, ”kan/kan ikke” eller ”skal forbedres før godkendelse/er godkendt” der bruges i checklister.
- samlet vurdering af mere komplekse procedurer/arbejdssituationer inkl. vurdering af delelementer af proceduren
 - flertrinsskalaer
 - ”under forventet niveau” til ”over forventet niveau”, der f.eks. bruges i Mini-CEX.
 - ”ikke udført” til ”udført struktureret og kompetent” fra den procedure specifikke del af OSATS.
- helhedsvurderinger af den uddannelsessøgende læge
 - flertrinsskalaer
 - ”næsten aldrig” til ”altid”, ”aldrig” til ”altid”, ”ikke kompetent” til ”kompetent” eksempler på skalaer fra 360 feedback.
 - ”stort set aldrig” til ”stort set altid” i generelle vurderinger i anæstesi.
 - ”meget dårlig til meget god” fra observation af ikke-tekniske færdigheder som bruges i NOTSSdk vurderingsskemaerne.
 - ”mange unødige bevægelser” til ”tydelig økonomi med bevægelser” fra den generelle del af OSATS.

Til den specifikke kompetencevurderingsmetode vælges den type af skala, der bedst kan understøtte det, som metoden skal vurdere. Dog er valget af skala desuden afhængig af, om skalaen skal anvendes til at vurdere om den uddannelsessøgende læge har opnået kompetencen (summativ kompetencevurdering) eller skal bruges til at give den yngre læge feedback på dennes kompetence udvikling (formativ kompetencevurdering), eller til vurdering af simple og entydige færdigheder eller procedurer.

De fleste skemaer til brug for kompetencevurdering har, udover observationspunkter med tilhørende skala, plads til skriftlige enten frie eller strukturerede kommentarer fra bedømmerne. Dette gælder f.eks. i checklister, hvor der ofte spørges; ”angiv hvad der skal forbedres”. I en 360 graders feedback efterspørges ”supplerende kommentarer” eller svar på hvad ”denne læge skal begynde med”.

I det følgende beskrives de to typer af skalaer samt gives eksempler på deres udformning og brug i kompetencevurdering i speciallægeuddannelsen.

3.4.1.1 Dikotomiske skalaer - simple og entydige færdigheder eller procedurer

Dikotomiske skalaer anvendes i checklister til struktureret observation eller optælling af aktiviteter.

Checkliste ved struktureret observation

Checklisten er et redskab til at systematisere observationer af procedurer, som opdeles i en række adskilte trin. Disse afspejler de successive steps, som tilsammen beskriver en fuldt ud gennemført procedure. Bedømmeren observerer lægens gennemførelse af proceduren og noterer, hvilke trin, der er udført og ikke er udført. Bedømmeren observerer, men tolker ikke, hvordan proceduren gennemføres dvs. foretager ikke vurderinger i forhold til ”godt / mindre godt / dårligt”, hvilket teore-

tisk fjerner subjektiviteten i evalueringsprocessen¹³. Checklisten indeholder færre grader af bedømmelse og er oftest dikotomisk begrænset (*gør/gør ikke, ja/nej*) med værdien 0/1. Til sidst samles score i en overordnet vurdering som konkludere "*kan / kan ikke udføre proceduren selvstændigt*". Scoren bliver lig med antal svar, der har fået værdien 1. Rent psykometrisk dvs. ved statistiske beregninger/valideringer giver en dikotomisk skala kun mening såfremt alle spørgsmål tilhører samme dimension dvs. er enten relateret til de tekniske færdigheder eller til kommunikation.

Checkliste til optælling af gennemført aktivitet

Checklisten er en fortegnelse over gennemførte aktiviteter samt antallet af disse. Checklisten er velegnet til formativ evaluering, fordi den gør uddannelsessøgende bevidst om, hvilke aktiviteter, der kræves, og hvilke prioriteringer, der skal gøres. Checklister anvendes i samtaler med hovedvejlederen og vil kunne danne baggrund for dokumentation i logbogen.

3.4.1.2 Flertrinsskalaer - samlet vurdering af komplekse arbejdsituationer/procedurer og helhedsvurdering af lægen

Flertrinsskalaer er ifølge den angelsaksiske litteratur enten "*globale*" dvs. vurderer den overordnede performance eller er flere skalaer, der måler forskellige aspekter/delementer af performance^{13,14}.

Flertrinsskalaer svarer til den engelske betegnelse "*global rating scale*". Ud fra hvert ord defineres, hvad skalaen anvendes til

- "*global*" betyder "alt inkluderet" eller overordnet
- "*rating*" det at kunne placere en persons performance på den kontinuerlige skala eller på et specifikt niveau
- "*scale*" beskriver en lineær analog oftest numerisk, der anvendes til at skelne mellem niveauer eller trin i en klinisk arbejdsituation.

Flertrinsskalaer til brug for samlede vurderinger af en kompleks procedure/arbejdsituation og til helhedsvurderinger af hele lægen skal være konstrueret af ekspert bedømmere. Dette er i overensstemmelse med, at den type af vurderinger, der foretages ud fra en flertrinsskala svarer til, at eksperter anvender overordnede subjektive vurderinger^{15,16}.

Der er udformet forskellige typer af flertrinsskalaer som beskrives kort nedenfor

1. *Likert skalaer (ækvivalent scoring)*; en skala, hvor der er mere end 2 svarmuligheder for hvert spørgsmål. Skalaens trin kan være "*helt uenig, delvis enig, neutral, delvis enig, helt enig*" som evt. kan angives numerisk fra 1 = *helt uenig* til f.eks. 5 eller 7 = *helt enig*.
2. *Kvantitativ vurderingsskala*; bruges til at bedømme adfærd og komplekse færdigheder, hvor der ikke findes mere objektive måleinstrumenter. En ordnat skala, som oftest er delt i 5 trin anvendes.
3. *BARS – Behaviorally Anchored Rating Scales*; American Board of Internal Medicine anvender en 9 punkts skala; *utilfredsstillende* (1-3), *tilfredsstillende* (4-6) og *superior* (7-9). En alternativ skala er; *under niveau/ikke kompetent til at udføre færdigheden/proceduren* (1-2), *borderline level* (3), *svarende til forventet kompetence niveau* (4), *over forventet niveau/over gennemsnit ifht at kunne udføre proceduren/færdigheden selvstændigt* (5-6)¹⁴.

4. OSATS – *Objective Structured Assessment of Technical Skills* indeholder dels en flertrinsskala (5 punkts) til de generelle kirurgiske kompetencer og dels en checkliste til de procedurespecifikke kompetencer/færdigheder. Sidstnævnte var oprindeligt en dikotomisk skala, men efterhånden udelukkende en 5-punkts skala baseret på adfærdsindikatorer¹⁷.

3.4.1.3 Brug af sammensatte skalaer

I OSATS, NOTSS og kompetencekort (se afsnit 4.1, 4.2, 4.3) anvendes nogle gange en kombination af dikotomiske skalaer og flertrinsskaler samtidig med, at flere domæner vurderes. Herudover foretages til sidst en samlet summativ vurdering med besvarelse af spørgsmål som ”denne kompetence er samlet vurderet som godkendt” eller ”kompetencekortet er gennemført tilfredsstillende”. Det er i valideringen af disse metoder vigtigt at sikre, at de relevante psykometriske analyser gennemføres før det konkluderes om en skala/metode er valid eller ej. Helhedsvurdering af den uddannelsessøgende læge foretages som et selvstændigt vurdering og ikke ved at addere de enkelte vurderinger af delkompetencerne.

3.4.2 Skalaers anvendelighed

Skalaer kan anvendes som led i formativ eller summativ kompetencevurdering. Skalaens udformning har betydning for dens anvendelighed til brug for feedback, hvor adfærdsbaserede skalaer som f.eks. BARS eller OSATS vurderes til i højere grad at kunne understøtte feedback sammenlignet med en dikotomisk skala^{18,19}.

En flertrinsskala kan konverteres til en dikotomisk skala f.eks. at beslutte at ≤ 3 svarer til ”ikke kompetent til at udføre” og > 3 ”kompetent til at udføre” færdigheden/arbejdssituationen” ved brug af en 5-trins skala. Det samme opnås ved at bruge en 4-trins-skala, der indeholder trinene; ”utilfredsstillende, næsten tilfredsstillende, tilfredsstillende og yderst tilfredsstillende”. Opbygges flertrinsskalaer på denne måde fungerer de i praksis som dikotomisk skalaer og er derfor ikke understøttende for formativ evaluering/feedback^{14,19}.

Oftest er der ikke lige lang afstand mellem de forskellige svarmuligheder i flertrinsskalaer. Det optimale numeriske antal i en skala er ligeledes til diskussion, men en skala indeholder fra 4 til en 9 punkter anbefales. Oftest vælges en 4-5 trins skala, da det er svært at adskille mere end 4-5 niveauer i vurderingen af den enkelte læge.

Anvendes en skala som led i en summativ kompetencevurdering, dvs. til at beslutte om en ”person er bestået eller dumpet” svarende til en eksamen, skal der forud være defineret en grænseværdi for, hvornår en kompetence vurderes som værende på det ønskede niveau. Valg af grænseværdi kan gøres arbitrært eller efter konsensus. Generelt stilles strengere krav til metodens evne til at diskriminere mellem novicer og eksperter (*construct validity*) og pålidelighed (*reliability*) når testen skal anvendes til summativ vurdering. Ideelt set skal sensitivitet og specificitet af grænseværdien være testet.

Ved score ”under niveau” anbefales det at supplere med specifikke kommentarer eller konstruktive anbefalinger som led i den formative vurdering jf. det skema som anvendes til vurdering af intern medicinske uddannelsessøgende læger under American Board of Internal Medicine¹⁴. Studier har vist, at hvis kvaliteten af de narrative kommentarer øges ændrer de uddannelsessøgende læger hyppigere adfærd omkring klinisk praksis og tillægger feedbacken større betydning²⁰.

Et review fra 2011 (Ahmed el al)²¹, der systematisk gennemgår kompetencevurderingsredskaber til brug for vurdering af tekniske færdigheder, konkluderer, at en kombination af flertrinsskalaer og procedurespecifikke checklister *per se* er at anbefale til vurdering af tekniske færdigheder. Det er dog vigtigt at være bevidst om at anvendelsen af skalaer er afhængige af kontekst f.eks. kan ikke-tekniske færdigheder ikke fuldt ud skilles fra tekniske og kliniske færdigheder dvs. typen af procedure og de gældende omstændigheder vil indgå i og vanskeliggøre kompetencevurderingen.

Ved anvendelse af skalaer er det vigtigt at fokusere på¹³:

- Bedømmerne
 - at træne bedømmere i at anvende skalaen, dvs. en formel uddannelse med krav til bedømmerne vil øge validitet
 - at bruge flere bedømmere (forbedre objektivitet og pålidelighed)
 - at give feedback til bedømmerne således at de anvender skalaen optimalt og har kendskab til om de er ”høge eller duer”²²
- Skalaen
 - at kende skalaens begrænsninger
 - at definere enderne af skalaen entydigt og præcist
 - at tvinge bedømmere til at bruge hele skalaen f.eks. varierer antal punkter på skalaen, klare definitionen af yderpunkter, anvende en 7 punkts skala
- Rammerne
 - at sikre ro og tilstrækkelig tid til at bedømmeren og den uddannelsessøgende læge kan være sammen under vurdering og efterfølgende feedback

3.4.3 Fordele ved skalaer

Checklister

Fordelen ved at anvende procedure relaterede checklister er, at bedømmelsen bliver meget præcis¹⁴. Checklisten kan med fordel bruges til at give struktureret feedback på såvel gennemførelsen som på den korrekte rækkefølge af de forskellige delelementer i en færdighed²¹. Checklisten skal designes med omhu og det skal sikres, at delfærdigheder tydelig kan observeres. Den skal afspejle god lægelig praksis i det pågældende speciale, Checklister har høj inter-bedømmer pålidelighed (*reliability*), når der måles på enkelt-færdigheder dikotomisk (ja/nej). Der er ingen pålidelige data vedrørende, hvor mange observationer på checklister af en færdighed, der kræves i postgraduat lægeuddannelse. Et rimeligt skøn er 5 – 10 gange, før bedømmelsen kan anses for valid.

Flertrinsskalaer

Fordelen ved at anvende flertrinsskalaer er, at de kan anvendes til at vurdere komplekse færdigheder/adfærd og at de tillader bedømmere at vurdere forskellige aspekter af den yngre læges praksis i den konkrete kliniske situation²³. Flertrinsskalaer er fleksible og de samme skalaer kan anvendes til vurdering af flere forskellige kliniske situationer. De er lette at anvende og ikke tidskrævende²¹. De er velegnede til efterfølgende at give feedback, specielt hvis yderpunkterne er observerbare og præcist/entydigt definerede¹³. De kan med fordel bruges til dokumentere og monitorere progression over tid og kan, såfremt det er muligt at definere en grænseværdi

for acceptabelt niveau, også anvendes som led i en summativ kompetencevurdering.

3.4.4 Ulemper ved skalaer

Checklister

Checklister anerkender ikke alternative tilgange til en problemløsning og afspejler dermed en novices tilgang til problemløsning. I OSCE (*Objective Structured Clinical Examination*) er det vist, at checklister ikke differentierer mellem novicer og eksperter dvs. har lav begrebs validitet²⁴. Baggrunden herfor er, at når ”ekspertbedømmere” anvender checklister kan de ikke anvende deres holistiske/erfaringsbaserede evne til at danne sig et overordnet billede af, hvor godt den yngre læge udfører proceduren og kan specielt ikke sætte fokus på de vigtigste trin i proceduren. Bedømmelsen bliver snæver, hvilket betyder, at checklister er uegnede til bedømmelse af mere komplekse færdigheder som f.eks. klinisk beslutningsproces, samarbejde med patienter og kollegaer, mv.¹⁴. Checklister er således ikke et optimalt redskab til brug for kompetencevurdering i den kliniske hverdag^{14,19}. Mange checklister vurderer et begrænset antal kompetencer med fokus på de tekniske færdigheder og uden at vurdere om proceduren er udført under hensyntagen til patientsikkerhed, humanistisk og teambaseret²⁵.

Flertrinsskalaer

Idet skalaen anvendes på flere forskellige kliniske situationer, kan yderpunkterne til tider være for løst formulerede og dermed vanskelige at anvende til at give feedback efter²¹. Svagheden ved flertrinsskalaer kan være, at de medfører en subjektiv vurdering og dermed har lav inter-observatør pålidelighed. Denne kan dog forbedres ved at instruere og træne bedømmerne¹³. Pålideligheden kan endvidere øges ved, at den læge, der skal bedømmes bliver bedømt i den samme situation og med den samme skala af forskellige og uafhængige bedømmere, enten samtidigt eller ved forskellige tidspunkter. Data indikerer, at der kræves mindst 11 forskellige observatører²⁶.

De største problemer med anvendelse af flertrinsskalaer er, hvordan bedømmere bruger dem og ikke, at der er større defekter i den måde skalaerne er designet/opbygget på. Problemerne er primært relateret til^{13,14}:

- problemer med at anvende hele skalaen; enten går vurderingen mod middel eller, at det kun er den øvre del af skalaen, der anvendes.
- ”due og høj” med en tendens til, at det er de mest ”flinke”, der vurderer højest²².
- at bedømmerens sympati for den yngre læge påvirker vurderingen således, at der enten gives højere eller lavere score.
- ”halo effekten”, der dækker over, at vurderingen af et delelement påvirker vurderingen af et andet delelement og dermed en tendens til, at bedømmeren vurderer de enkelte delelementer ens. Der angives flere årsager til ”halo effekten” bl.a., at det overordnede/globalt indtryk har stor betydning i vurderingen af delelementer, at der kan være en modvilje mod at give en lav score rating samt mangel på tilstrækkelig observation af den yngre læges performance. Øges længde og kompleksiteten med flere domæner/kriterier og/eller flere spørgsmål vil dette forstærke ”halo effekten”.

3.4.5 Skalaers validitet, pålidelighed og gennemførlighed

En høj grad af inter-observatør pålidelighed er ønskelig specielt hvis vurdering foretages over tid af flere bedømmere. Er bedømmerne ikke trænet i at anvende skalaer /metoder eller anvender de dem sjældent, falder pålideligheden. For at vurdere validiteten fuldt ud kræves en "guld standard", men denne findes ikke på områder som f.eks. professionalisme, holdninger og klinisk vurdering^{14,15, 27}.

Checklister er tidligere blevet opfattet som værende mere objektive og dermed mere reliable (træfsikre) end flertrinsskalaer. Dette er udfordret af Norman²⁸, der påpeger, at pålideligheden ikke øges *per se* ved at anvende checklister. Anvender ekspertbedømmere flertrinns skalaer øges pålideligheds- og validitetsparametre sammenlignet med checklister^{29, 30}.

Anvendes flertrinsskalaer som led i en formativ kompetencevurdering dvs. med det mål at identificere dem, der har behov for yderligere læring kan en vis grad af falsk positive i forhold til at identificere inkompetence accepteres. Derimod kan falsk negative ikke accepteres. I et studie af Ma et al medførte anvendelse af flertrinsskala frem for checklister identifikation af personer, der var inkompetente i en grad, der ville betyde skade på patienter og som var vurderet som kompetente ud fra checkliste score. Ma et al anbefaler således brug af flertrinsskalaer¹⁹.

3.5 Validering af metoder til kompetencevurdering - psykometriske /statistiske begreber

Der findes forskellige analyserammer for, hvordan en kompetencevurderingsmetode skal valideres. Van der Vleuten beskriver en model, der inkluderer evaluering/vurdering af realibilitet og validitet samt gennemførlighed, acceptabilitet, uddannelsesmæssig effekt og cost-effectiveness³¹. Problemstillingen omkring validering af metoder til kompetencevurdering er, at en validering af en metode/test kræver, at der kan defineres en "guld standard". Dette er ikke muligt indenfor områder som professionalisme, adfærd og evne til klinisk beslutningstagning (Dansk psykolog forening)^{14,21,32}.

3.5.1 Pålidelighed (reliabilitet)

Pålidelighed (træfsikkerhed) er metodens evne til at nå det samme måleresultat ved gentagne målinger under samme betingelser. Det udtrykkes tillige som konsistens eller reproducerbarhed og er mål for unøjagtigheden i en metode. Pålidelighed udtrykker korrelationen mellem et sæt af observationer. Realibiliteten kan være 1 uden at bedømmerne er enige, hvis der er tale om systematisk bias hos en bedømmer. Til formativ kompetencevurdering kan en rimelig fejlmargen accepteres, men ved en summativ vurdering (= afsluttende test) bør graden af pålidelighed være høj³³. Jo flere observationer/spørgsmål jo mere pålidelig vil testen/metoden være. Desuden øges pålideligheden ved at bruge flere bedømmere, der hver især bedømmer det samme delement.

De 4 mest anvendte pålidelighedsmål er

- Test-retest (intra-rater pålidelighed)
 - Testens evne til at give samme resultat hvis den anvendes af samme person på 2 forskellige tidspunkter dvs. testens stabilitet over tid. Værdien udtrykker korrelationen mellem test og re-test.

- *Testes ved at samme bedømmer vurderer 2 gange f.eks. ser en video med 2 dages interval*
- Intern konsistens (inter-item pålidelighed)
 - Afspejler korrelationen mellem de forskellige spørgsmål (items) i en test og hvordan hvert item bidrager til testens resultat. Det er et mål for i hvilken grad alle items i testen måler det samme underliggende begreb eller om der er konvergens på det begreb, der måles. Det angives som Cronbach´s alfa koefficienten (0-1) som stiger jo mere homogene items bliver og jo flere items, der medtages i testen. Cronbach´s alfa kan kun anvendes for multiitems Likert skalaer. Skal generelt være > 0,8.
 - *Testes ved at sammenligne resultatet fra forskellige spørgsmål.*
- Inter-rater pålidelighed
 - Graden af enighed mellem ≥ 2 observatørers vurdering (rating) af den samme seance (f.eks. ved at se samme video).
 - *Testes bedst ved at anvende 2 uafhængige og blindede observatører.*
- Reproducerbarhedskoefficient
 - Graden af hvordan en persons respons på et item kan forudsiges ud fra svar på tidligere items dvs. at en respondent, der har svaret positivt på et item ligeledes må svare positivt på et mindre vanskeligt spørgsmål. Pålidelighedskoefficient angives fra 0-1 og grænsen er generelt 0,8.

3.5.2 Validitet

Validitet udtrykker den grad, hvormed et instrument/en metode faktisk måler det, som den er tænkt at skulle måle og omvendt ikke måler det, som ikke ønskes målt. Validitetsbegrebet knytter sig ikke alene til instrumentet/metoden eller skalaen alene, men knytter sig også til, om den er gyldigt for et specifikt formål og ved brug overfor en specifik gruppe f.eks. studenter, yngre læger eller speciallæger. At en metode er valideret er ikke i sig selv tilstrækkelig, idet der skal tages hensyn til såvel formål som målgruppe samt viden om hvordan validering er foretaget. Der findes en lang række validitetsmål.

- Face validity
 - I hvilken grad afspejler testen situationen i det virkelige liv? Virker testen umiddelbar fornuftig i forhold til sit formål?
 - Hvordan vurderer brugerne funktionaliteten og realismen af en given test?
 - *Testes ved at spørge, hvordan yngre læger og bedømmere vurderer metoden?*
- Indholds validitet (*content validity*)
 - Er alle væsentlige aspekter set ud fra en ekspert-vurdering taget med?
 - Syntes de medtagne items umiddelbart at dække det, der ønskes testet? Er de repræsentative i forhold til den kompetence, der skal vurderes?
 - Er indholdet af testen egnet til at måle det, som den formodes at skulle måle?

- *Testes ved brug af litteraturstudier, ekspertgruppediskussioner (Delphi, fokusgrupper mm).*
- Begrebsvaliditet (construct validity)
 - Hvorvidt en test i virkeligheden måler det, den er designet til at skulle måle f.eks. tekniske eller kommunikative færdigheder?
 - Kan testen diskriminere mellem forskellige eksperter eller novicer fra eksperter? (*discriminative validity*)
 - *Testes ved at vurdere novicer, intermediære og eksperter.*
- Kriterievaliditet (*criterion validity*)
 - I hvilken grad testresultatet stemmer overens med ydre kriterier? Er testen relateret til den virkelige verden? Korrelerer instrumentet med et andet mål f.eks. "the good doctors test" som guld standard?
 - samtidig kriterievaliditet (*concurrent validity*); stemmer resultatet overens med resultatet ved andre etablerede tests?
 - prädiktiv eller fremtidig kriterievaliditet; instrumentets evne til at forudsige noget centralt vedrørende det fænomen man ønsker målt
 - *Testes ved at sammenligne med en anden test. Dette er ofte svært indenfor medicinsk uddannelse og specielt ved metoder, der bruges til at vurdere adfærd, hvor guld standarden ikke er entydig svarende til at der ikke eksisterer en "the good doctor test"!*

4 Direkte observation

Direkte observation i forbindelse med en arbejdssituation anvendes i forskellige former af adskillige specialer. Observationen kan være af tekniske færdigheder som f.eks. kirurgiske tekniske færdigheder eller af ikke-tekniske færdigheder som kommunikation, ledelse, samarbejde, situationsbevidsthed og beslutningstagen. Hertil anvendes checklister og skalaer, hvor udformningen er afhængig af formålet (se kapitel 3). Direkte observation af en kombination af viden, færdigheder og holdninger anvendes ligeledes af mange specialer. Eksempler på sådanne strukturerede observationer er kompetence kort eller miniCEX.

4.1 Objektiv vurdering af tekniske færdigheder

Der findes forskellige muligheder for kompetencevurdering af tekniske færdigheder. En af de mest undersøgte og udbredte er en fem punkts flertrinsskala *Objective Structured Assessment of Technical Skills (OSATS)*¹⁷. OSATS er udviklet i Canada, oprindeligt på simple opstillede modeller, såkaldte *bench-models* (Martin et al)¹⁷ og er siden modificeret til at kunne bruges til kirurgiske procedurer og andre procedurer. Senere er der udviklet en række procedurespecifikke OSATS skemaer og skalaer til andet end kirurgiske tekniske færdigheder.

Et andet eksempel på globale vurderingsskemaer til kompetencevurdering af tekniske færdigheder er den amerikanske *Fundamentals of Laparoscopic Skills (FLS)*, som er en videreudvikling af *McGill Inanimate Systeme for Training and Evaluation of Laparoscopic skills (MISTELS)*. FLS indgår som en test i curriculum for kirurgisk træning i USA, men er ikke et egentligt specifikt system til vurdering af kirurgisk kompetence. Et andet vurderingsskemaer er *Global Operative Assessment of Laparoscopic skills (GOALS)*, som vurderer generelle laparoskopiske kompetencer. GOALS er testet og fundet brugbar, men er ikke så udbredt som OSATS^{34, 35, 36, 37, 38}, hvorfor det ikke vil blive yderligere omtalt her.

4.1.1 Definition af OSATS

OSATS kan oversættes til objektiv struktureret vurdering af tekniske færdigheder¹⁷.

4.1.2 OSATS anvendelighed

OSATS er oprindeligt brugt til kompetencevurdering af abdominal kirurgiske færdigheder¹⁷, men der er siden beskrevet anvendelse af og udviklet skalaer til brug indenfor gynækologi og obstetrik¹⁸, karkirurgi, urologi, og oftalmologi.

OSATS består af en 5-punkts flertrinsskala, hvis yderpunkter er modsætninger og de mellemliggende trin er grader imellem.

Kategorierne i den generelle OSATS til kirurgisk kompetencevurdering er oversat til dansk. Andre skala-eksempler kan findes i litteraturen, se f.eks. Larsen CR et al, 2012³⁸ og Maagaard M et al, 2012¹⁸.

OSATS er efterfølgende udviklet i modificerede former til specifikke operative procedurer som f.eks. laparoskopisk cholecystektomi og salpingektomi, suturering af bristinger efter fødsler samt til ikke-kirurgiske procedurer som kopforløsning. OSATS som princip har også givet inspiration til et vurderingsskema, der kan bru-

ges til vurdering af generelle færdigheder ved ultralydsscanning benævnt *OSAUS*, dvs. *objective structured assessment of ultrasounds skills*³⁹.

4.1.3 Fordele ved OSATS

OSATS vurderingsskemaer kan både bruges ved direkte observation på færdigheder udført på fantomer samt på færdigheder udført i klinisk praksis. Det er muligt, at anvende OSATS til vurdering af videooptagelser, dvs. vurderinger kan foregå uafhængigt af, om en færdighed udøves i klinisk praksis.

Systematiske reviews viser, at OSATS især er velegnede til vurdering af færdigheder, hvor der samtidig skal foretages en vurdering af progressionen af læringen og hvor der efterfølgende skal gives feedback med det mål at skabe yderligere læring^{35,38}. Flertrinsskalaer som 5 punktsskaler har formentlig fordele frem for dikotomiske skaler og er formentlig mere præcise i vurderingerne, men yderligere undersøgelser deraf er nødvendige¹⁹.

4.1.4 Ulemper ved OSATS

OSATS og lignende flertrinsskalaer har den ulempe, at de primært er udviklet og testet på modeller, dyreoperationer samt fantomer og ikke til humane operationer og klinisk praksis.

Der er ved anvendelsen tvivl om hvilke grænseværdier, der er gældende for, at en præstation ikke kan accepteres. Vurderingsskemaerne er derfor vanskelige at anvende til summativ kompetencevurdering ved en eksamen.

4.1.5 OSATS validitet, pålidelighed og gennemførlighed

OSATS har demonstreret pålidelighed og validitet ved blindet evaluering af uddannelsessøgendes læges færdigheder. Der er god *interrater reliability* (IRR), hvilket udtrykker forholdet mellem antallet af overensstemmende observationer og det totale antal observationer. *Cronbach's alfa coefficient* (α), der udtrykker konsistens observatørerne imellem (*internal consistency*) er høj.

Vurdering af *indholds validitet* af OSATS er gjort ved Delphi audit⁴⁰.

Flere af de procedurespecifikke vurderingsskemaer har vist god begrebsvaliditet dvs. at de kan skelne mellem novicer og læger med forskelligt kompetenceniveau.

Yderligere udvikling og testning af procedurespecifikke vurderingsskemaer vil muligvis give bedre muligheder for at kunne anvende OSATS til brug for summativ vurdering.

Udvikling af procedurespecifikke vurderingsskemaer er dog krævende og stiller krav om, at der findes en bedste evidensbaseret klinisk praksis indenfor en operation, hvis et procedurespecifikt vurderingsskema skal kunne udvikles og anvendes over tid og på tværs af institutioner.

4.1.6 Praktisk anvendelse af OSATS

Supervisor eller observatør ved et kirurgisk indgreb kan score på OSATS skemaet og give feedback direkte til den uddannelsessøgende læge ud fra registreringerne på OSATS skemaet.

Et OSATS skema kan med fordel anvendes på videooptagelser. Dette er tidskrævende, men eliminerer den vanskelighed der kan være ved at skulle give og modtage feedback på en operationsstue. Det kan også gøres blindet, således at interaktion mellem uddannelsessøgende og bedømmer/supervisor elimineres.

Indenfor de kirurgiske områder er der solide data, der viser at OSATS kan anvendes i forbindelse med kirurgisk træning af laparoskopiske og endoskopiske færdigheder på simulatorer^{35,38,41,42,40}.

Eksempler på OSATS skemaer kan ses i bilagene 4.1.7.1, 4.1.7.2 og 4.1.7.3.

4.2 Objektiv vurdering af ikke-tekniske færdigheder

4.2.1 Definition af ikke-tekniske færdigheder og af NOTSS/ ANTS

Rollerne som kommunikator, samarbejder, leder og professionel repræsenterer kompetencer, der i andre domæner som f.eks. luftfart er blevet beskrevet som ikke-tekniske færdigheder. Disse kan defineres som kognitive og interpersonelle færdigheder, der underbygger de kliniske og tekniske kompetencer. Beskrivelsen tager udgangspunkt i betydningen af menneskelige faktorer (human factors) tankegangen fra andre højrisiko domæner. I UK beskrives disse færdigheder som kernekompetencer indenfor professionalism, kommunikation, ledelse og team samarbejde, patientsikkerhed mv⁴³.

Et af de områder i sundhedsvæsnet, hvor denne tankegang er appliceret, er for team medlemmerne på operationsstuen. Her har studier vist, at nedsat situationsbevidsthed og utilstrækkeligt samarbejde medfører en øget risiko for komplikationer⁴⁴. Forskelle i præsentationer er forklaret ud fra kirurgens udvisning af god objektivt, observerbar adfærd, såsom tilpasningsevne, kognitiv fleksibilitet og kommunikationsstil hos kirurgen⁴⁵.

Flere forskergrupper har udviklet *behavioural marker* systemer (taxonomier) til at beskrive de ikke-tekniske færdigheder på individniveau. Disse systemer beskriver observerbar adfærd på individ niveau og anvendes til at strukturere læring og kompetencevurdering af ikke tekniske færdigheder. Sådanne systemer er udviklet for kirurgi og anæstesiologi. For kirurger findes eksempelvis the Oxford Non-Technical Skills (NOTECHS) og Non Technical Skills for Surgeons (NOTSS)^{46, 47}. For anæstesiologen er Anaesthesia Non Technical Skills (ANTS) system udviklet⁴⁸. Lignende systemer er udviklet for anæstesisygeplejersker og operationssygeplejersker.

4.2.2 NOTSS/ ANTS anvendelighed

De nævnte systemer til observation af færdigheder på operationsstuen er udviklet i UK. NOTSS er udviklet og valideret til anvendelse i kirurgi, men efterfølgende er redskabet anvendt i andre skærende specialer. ANTS er udviklet til anæstesiologer.

NOTSS og ANTS er efterfølgende oversat og valideret til dansk kontekst^{49, 50}. Det danske eksempel på NOTSS.dk kan ses på www.regionh.dk/DIMS/menu/notss.dk

Taxonomien består af overordnede kategorier, elementer og eksempler på god eller dårlig adfærd. Til taksonomien er tilknyttet en kvantitativ vurderingsskala, som kan ses i bilag 4.2.6.1.

En dansk håndbog er tilgængelig, til brug for både uddannelsessøgende og uddannelsesansvarlige læger. Håndbogen beskriver hvordan NOTSS kan anvendes i den kliniske uddannelse, og kan ses på [www.regionh.dk/DIMS/ menu/ notssdk.dk](http://www.regionh.dk/DIMS/menu/notssdk.dk). En lignende håndbog vil blive udviklet til ANTS.

4.2.3 Fordele ved NOTSS/ ANTS

Kommende speciallæger forventes at udvise gode ikke-tekniske færdigheder, men oplæring og vurdering af disse er ikke tidligere blevet gennemført systematisk. Taxonomier, som ANTS og NOTSS gør det mere tydeligt for den uddannelsessøgende, hvad der forventes og beskriver flere af lægerrollerne og betydningen af disse for den medicinske ekspertrolle. De uddannelsesgivende får ved de givne eksempler nemmere ved at foretage kompetencevurdering i kliniske situationer og derved kan en mere nuanceret og struktureret feedback gives til den uddannelsessøgende.

Gentagne målinger over tid kan medvirke til at den uddannelsessøgende kan følge egen progression.

4.2.4 Ulemper ved NOTSS/ ANTS

NOTSS og ANTS er udviklet af forskere til brug i klinikken. Udfordringen kan være at få systemerne implementeret i en travl klinisk hverdag. Introduktion til redskabet er nødvendig, mest pga. den manglende viden om betydningen af ikke tekniske færdigheder og betydningen af at fungere som et team.

4.2.5 NOTSS/ ANTS validitet, pålidelighed og gennemførlighed

Systemerne er baseret på grundigt forarbejde bestående af interview, observationer og spørgeskema undersøgelser. Pålidelighed og validitet af de oprindelige UK udgaver af NOTSS og ANTS er publiceret^{46,48,47}. Pålidelighed af den danske udgave af NOTSS er undersøgt ved anvendelse af simulationsbaserede video optagelser^{49,50}.

Hvem skal foretage vurderingen? Umiddelbart giver det mening, at kompetencevurderingen foretages af den, der superviserer den uddannelsessøgende på operationsstuen. En anden mulighed er, at lade en observatør foretage kompetencevurderingen. Den førstnævnte løsning er den mest realistiske og gennemførlige. For kirurger vil det således være en supervisor eller observatør, som følger operationen. For anæstesiologer i DK kan det være vanskeligt at supervisere under hele indgrebet, da uddannelsessøgende sjældent superviseres under hele det kirurgiske ind-

greb, men oftest under induktion og ved afslutningen. Anæstesiologen vil ofte blive kaldt til at deltage, hvis der opstår problemer såsom blødning. At sætte fokus på induktion og afslutning ved kompetencevurderingen, vil måske være lettere at få til at fungere. I fremtiden kunne man overveje om andre team medlemmer kunne give feedback.

Det optimale tidspunkt for at give feedback kendes ikke. Umiddelbart virker det mest sandsynligt, at en feedback givet lige efter operationen vil være mest operationelt, da teamet ikke nødvendigvis skal være et team ved efterfølgende operation.

4.2.6 Praktisk anvendelse af NOTSS/ ANTS

Supervisor eller observatør anvender redskabet til at score den uddannelsessøgendes kompetence og til at strukturere feedback. Kompetencevurderingssituationen planlægges således, at begge parter er bevidst om at dette skal foregå. Kirurgen har mulighed for at give løbende feedback under operationen, men udfyldelse af skema og anvendelse af vurderingsskalaen kan først foretages efter afslutningen af operationen.

Redskabet anvendes flere gange under et uddannelsesforløb, f.eks. i alt 10 gange under den samlede introduktions- og hoveduddannelse. Jo flere gange redskabet anvendes, jo mere valide er målingerne. Redskabet anvendes desuden under de obligatoriske kurser i hoveduddannelsen, hvor kursisterne introduceret til begreberne og træner disse egenskaber i workshops og under den simulationsbaserede træning.

Eksempler på NOTSS skema kan ses i bilag 4.2.6.1.

4.3 Kompetencekort

4.3.1 Definition af kompetencekort

Kompetencekortet er en checkliste, som anvendes i forbindelse med direkte observation af en uddannelsessøgende læge. I intern medicin anvendes ordet kompetencekort endvidere for en systematisk opbygget casebaseret diskussion, da det er udformet som en checkliste. Af kompetencekortet fremgår hvilke elementer og kriterier for bedømmelsen, der indgår i kompetencevurderingen. Ofte vurderes en kombination af viden og færdigheder, idet der er en stigende bevidsthed om, at der i forbindelse med udførelsen af en procedure indgår tænkning både før, under og efter udførelsen af proceduren. En vurdering af flere dimensioner foretages, hvilket betyder at der er spørgsmål vedrørende de tekniske procedurer og f.eks. spørgsmål vedrørende information af patienten/samarbejdspartnere. Ofte anvendes en kombination af dikotomiske og flertrinsskalaer.

I de initiale danske studier er begrebet checkliste blevet anvendt af nogle specialer⁵¹. I andre specialer som f.eks. ortopædiske kirurgi og anæstesiologi, anvendes ordet kompetencekort. Internationalt anvendes begrebet ”competence cards” eller ”clinical encounter cards”.

4.3.2 Kompetencekorts anvendelighed

Kompetencekortene består af forskellige delelementer, som i forbindelse med kompetencevurderingen kan krydses af (kryds for at kompetencen er godkendt). Et kort kan bestå af vidensspørgsmål, færdigheder i relation til procedurer eller kom-

munikation med patient eller teammedlemmer. Kompetencekortene er overvejende specifikke for den givne arbejdssituation.

Kompetencekortet kan anvendes på forskellige måder. Nogle specialer har udvalgt kompetencekort, som skal bestå inden den uddannelsessøgende må foretage en given procedure selvstændigt. I anæstesiologi blev den form for anvendelse kaldt ”kørekort”⁵² og eksempelvis skulle kompetencerne på 4 kort bestå inden lægen i introduktionsstilling måtte gå i vagt (kørekortsmodellen). Alle delelementer skal bestå ved en observation. Nogle specialer anvender en samlet vurdering til slut.

4.3.3 Fordele ved kompetencekort

Fordelene ved kompetencekortene er, at de adresserer flere lægeroller som f.eks. den medicinske ekspert/lægefaglig, professionel og kommunikator rolle. En anden fordel er, at de ved at linke teori til praksis kan anvendes til at vurdere elementer af klinisk beslutningstagen i forbindelse med en arbejdssituation. Helheden i opgaveløsning er af betydning, og delelementerne skal ses som en hjælp til både den uddannelsessøgende og bedømmeren. Det betyder, at det er mere tydeligt, hvad der skal bedømmes og gives feedback på.

Indholdet i et kompetencekort skal gennemtænkes, således at kompetencevurderingen kan foregå i forbindelse med en given arbejdssituation – alle punkter skal kunne vurderes i et sammenhængende klinisk forløb. Metoden er derfor ikke specielt tidskrævende, idet det kun tager den tid som den kliniske opgaver tager samt den tid, det efterfølgende tager at giver feedback.

Kørekortsmodellen gør det tydeligt for både den uddannelsessøgende og speciallægen, hvilke opgaver, der kan løses uden supervision og hvilke, der skal løses i samarbejde med speciallægen. Dette kan være en hjælp specielt i vagtsituationen hvor læger, som ikke så ofte arbejder sammen mødes om opgaver.

4.3.4 Ulemper ved kompetencekort

Ulempen ved kompetencekortene er, at den enkelte speciallæge ikke nødvendigvis føler sig rustet til at foretage kompetencevurderingen. Planlægning af hvordan en sådan skal foregå er derfor vigtig. Uddannelse af bedømmerne til at foretage kompetencevurdering og at give specifik feedback er også nødvendig.

4.3.5 Kompetencekorts validitet, pålidelighed og gennemførlighed

Indholds- og begrebsvaliditet af et kompetencekort til stuegang for læger i intern medicin er høj⁵¹. Ideelt set burde hvert enkelt kompetencekort i et givent speciale valideres, men pragmatisk set handler det lige så meget om, at det giver mening, skaber læring og gør det nemmere at vurdere kompetence. I anæstesiologi er den overordnede tilfredshed med en portefølje indeholdende adskillige kompetencekort vurderet høj⁵³.

En undersøgelse af anvendelighed og gennemførlighed viste størst tilfredshed med checklisteformat (check af hvert enkelt punkt på kompetencekortet) frem for en samlet vurdering⁵⁴.

4.3.6 Praktisk anvendelse af kompetencekort

Kompetencekort er anvendt i intern medicin i form af et stuegangskort, i ortopædkirurgi til knæundersøgelse og i anæstesiologi til f.eks. akut indledning, spinal anæstesi mv.

Eksempler på kompetencekort kan ses i bilag 4.3.6.1 og bilag 4.3.6.2.

4.4 Mini Clinical Evaluation Exercise (mini-CEX)

4.4.1 Definition af mini-CEX

Mini-CEX er en metode til vurdering af uddannelsessøgende lægers kliniske kompetencer indenfor anamnese, objektiv undersøgelse, klinisk dømmekraft og vurdering, patientvejledning og rådgivning, organisation og samarbejde, empati og professionel adfærd og generel klinisk kompetence. Metoden er udviklet i USA. Norcini et al viste, at scoren af den uddannelsessøgende læges kliniske dømmekraft, organisation og samarbejde udviklede sig over et år ved anvendelse af mini-CEX⁵⁵.

4.4.2 Mini-CEX anvendelighed:

Metoden kan anvendes i stort set alle kliniske situationer. Med andre fokuspunkter kan metoden også anvendes i parakliniske fag.

4.4.3 Fordele ved mini-CEX

Mini-CEX kan let passes ind i klinikken, da det kun behøver tage observatøren den tid som ”konsultationen” varer samt tid til efterfølgende feedback. Selve observationen er afhængig af formålet med observationen. Mini-CEX kan anvendes i flere kliniske situationer og er en ”arbejdssituationsvurdering”. Mini-CEX kan integreres i den daglige drift, og kræver ikke særlige foranstaltninger eller forberedelser. Øvrige fordele er, at den uddannelsessøgende kan kompetencevurderes flere gange med miniCEX og dermed får metoden både øget validitet og pålidelighed. Der skal være mindst 8-10 mini-CEX vurderinger for at metoden har pålidelighed²⁶.

4.4.4 Ulemper ved mini-CEX

Der er altid en risiko for bias ved direkte observation – både i forhold til bedømmeren og den uddannelsessøgende læge²². Bedømmeren kan synes at en klinisk situation er kompleks og derved give en højere ’karakter’ til uddannelsessøgende læge. Der findes både ”strenge” og ”blide” bedømmere, men bias vil formindskes, hvis den enkelte læge bliver vurderet flere gange svarende til de anbefalede 8-10 mini-CEX.

4.4.5 Mini-CEX validitet, pålidelighed og gennemførlighed

Mini-CEX er den kompetencevurderingsmetode som har den bedste evidens for validitet²⁶, inklusive begrebsvaliditet⁵⁶. Holmboe et al har bedømt videooptagelser af læger i kliniske situationer med mini-CEX skemaet og påvist at læger, som klarede sig dårligt, middel og udmærket, havde forskellig scores⁵⁶. Sammenlignet med brug af standardiserede patienter har metoden en højere nøjagtighed.

Metoden har været afprøvet i en dansk onkologisk afdeling, med gode resultater⁵⁷. Både de uddannelsessøgende læger og bedømmere synes, at det er en god metode til kompetencevurdering.

4.4.6 Praktisk anvendelse af mini-CEX

Bedømmeren observerer den uddannelsessøgende læge i den kliniske situation f.eks. i ambulatorium, ved stuegang, i den akutte modtagelse, ved journalskrivning i behandlingsrum mv.

Bedømmeren udfylder et skema og tager stilling til fokus og kompleksitetsgrad. Skemaet tager udgangspunkt i en flertrinsskala med definerede yderpunkter. Vurderingen kan tage udgangspunkt i den uddannelsessøgende lægers kompetenceni-veau på nuværende uddannelsestrin eller en vurdering op imod specialiststandard. Hvis man taget udgangspunkt i specialiststandard kan den uddannelsessøgendes progression følges over tid.

Der bør være givet fokusområder (f.eks. objektive undersøgelse/klinisk dømmekraft) i forbindelse med den enkelte kompetencevurdering. Bedømmeren giver efterfølgende specifik og konstruktiv feedback på situationen. I feedback situationen bør den uddannelsessøgende læge også selv give en vurdering af den kliniske situation. En plan for hvorledes lægen kan forbedre sig bør fremlægges af lægen – eventuelt hjulpet af anbefalinger fra bedømmeren.

Eksempel på mini-CEX skema kan ses i bilag 4.4.6.3.

5 Helhedsvurderinger

Begrebet helhedsvurdering beskriver en mere overordnet vurdering af lægens samlede performance. Denne vurdering foretages på baggrund af en eller flere samarbejdssituationer. Vurderingen kan foretages over tid, hvor man følger lægen gennem flere arbejdssituationer eller ved at lave vurderingen ud fra porteføljeoptegnelser, hvor man f.eks vurderer summen af flere enkeltstående kompetencevurderinger og giver en samlet overordnet score på lægens performance.

En analog skala (ofte numerisk) anvendes fra dårlig / uacceptabel til excellent på forskellige områder / lægelige arbejdssituationer. Oftest er skalaen defineret ud fra nærmere fastsatte mål for ønsket adfærd (performance). Skalaen bygger på Dreyfuss' udviklingsbegreb, hvor lægen bevæger sig fra novice → avanceret begynder → kompetent → kyndig (proficient) → intuitiv ekspert → mestring⁵⁸. Men samtidig rummer skalaerne Blooms taxonomi for udvikling af klinisk ræsonnement, hvor man bevæger sig fra viden om → forståelse af → applicering af viden → analyse → syntese → evaluering⁵⁹.

Ved at bruge denne form for skalaer kan man vise, hvordan lægens faglige formåen stadig øges idet lægens score på skalaen gerne skulle flytte sig i positiv retning i takt med, at lægen bevæger sig fra at være ny i faget til at være ekspert. Anvendelsen af disse skalaer kan gøres mere eller mindre eksplicit med hensyn til hvor lægen befinder sig i sit uddannelsesforløb ved f.eks at definere standarder for opnået score på forskellige uddannelsesstrin (baislæge, introduktionslæge, tidlig hoveduddannelse, sen hoveduddannelse, speciallæge)⁶⁰. Dermed er de fleste skemaer, til brug for helhedsvurdering af den uddannelsessøgende læge, specialespecifikke.

Der er en glidende overgang mellem brugen af helhedsvurdering af samlet performance til brugen af enkeltvurderinger i form af strukturerede observationer.

5.1 Generelle vurderinger

5.1.1 Definition af generelle vurderinger

Til vurderinger af komplekse kliniske færdigheder eller vurdering af adfærd, f.eks. indenfor samarbejde, kommunikation, ledelse eller evne til at tage beslutninger samt til at påvise progression, anvendes ofte generelle vurderinger, der i litteraturen benævnes *Global Rating Scales*. "Global" betyder her overordnet/"alt inklusive" og "rating scale" står for en skala til indplacering af personens performance med veldefinerede målepunkter/yderpunkter i hver ende af skalaen, som beskriver niveauet for den observerede præstation. Generelle vurderinger har vist sig at være bedre egnet til at vurdere overordnede funktioner end simple checklister²⁴.

5.1.2 Generelle vurderingers anvendelighed

Der er tale om en overordnet vurdering designet/konstrueret af en ekspert hvor der kan diskrimineres mellem novice-expert, og hvor der ved flere målinger kan vises udvikling/progression^{15,16}. Metoden kan anvendes i kliniske situationer, hvor der er mulighed for direkte observation. Vurdering kan gentages i lignende situationer af både samme observatør eller en anden, og kan udføres i komplekse og pressede kliniske situationer. Vurderingen giver god mulighed for feedback⁶⁰. Såfremt man på forhånd definerer en grænseværdi for, hvornår en kompetence er på det ønskede

niveau, kan man anvende metoden summativt³⁴. Dette er særligt vigtigt ved anvendelse til skelsættende vurderinger som en eksamen.

Der er udviklet forskellige skemaer til generel vurdering. De fleste bygger på R.I.M.E *Student Evaluation Framework*^{61,62}, som oprindeligt blev udviklet til at vurdere studenter i kliniske ophold. Dette har dannet grundlag for udvikling af forskellige skemaer til generel vurdering af den uddannelsessøgende lægers udvikling i klinisk arbejde. R.I.M.E. står for Reporter ("hvad": anamneseoptagelse og objektiv undersøgelse), Interpreter ("hvorfor": diagnose), Manager ("Hvad så": plan og overvejelser) samt Educator (opfølgning).

Det amerikanske Accreditation Council for Graduate Medical Education (ACGME) har udviklet et skema til vurdering af kernekompetencer (svarende til lægeroller), som anvendes til at vurdere yngre lægers kompetence i forhold til kernekompetencerne (se eksempel i bilag 5.1.7.2). I Danmark anvendes ligeledes et skema til vurdering af generelle kompetencer ud fra de 7 lægeroller i anæstesiologi (bilag 5.1.7.1.). Erfaringen fra flere studier tyder på, at metoden anvendes bedst i kombination med andre metoder. Metoden er velegnet til at udpege indsatsområder indenfor de forskellige roller / kompetencer¹⁰.

Et dansk studie viser, at R.I.M.E. validt kan påvise udviklingen i kompetencer⁶³.

5.1.3 Fordele ved generelle vurderinger

Generelle vurderinger er velegnet og fleksible til vurdering af mange typer klinisk adfærd eller kompetencer. Det er hurtigt og nemt at udføre for bedømmeren, som anvender sin holistiske/erfaringsbaserede ekspertviden. Generelle vurderinger kan anvendes i næsten alle kliniske situationer og i en efterfølgende feedback.

5.1.4 Ulemper ved generelle vurderinger

Bedømmeren anvender en subjektiv og holistisk tilgang, der ikke nødvendigvis er pålidelig. Det kan være vanskeligt at udforme skemaerne, idet der ikke altid findes en guld standard.

5.1.5 Generelle vurderingers validitet, pålidelighed og gennemførlighed

Såfremt der er veldefinerede yderpunkter (guld standard) har metoden god indholdsvaliditet. Pålidelighed afhænger af bedømmernes kendskab til metoden og kan bedres ved at træne bedømmere og ved at anvende mere end én bedømmer i samme situation – eller ved at anvende gentagne målinger. Metoden har vist sig gennemførlig i klinisk praksis⁶⁴.

5.1.6 Praktisk anvendelse af generelle vurderinger

De generelle vurderingsskalaer skal udarbejdes af eksperter. Bedømmer udfylder skemaet i forbindelse med observation af klinisk arbejdsgang eller ved porteføljevurdering. Der er mulighed for at give frie kommentarer på skemaet. Seancen afsluttes med mundtlig feedback og dialog over hensigtsmæssige forbedringsmuligheder. Den uddannelsessøgende læge kan opbevare sin vurdering i porteføljen og lade den indgå i vejledersamtaler efterfølgende.

Eksempler på skema til anvendelse i forbindelse med generelle vurderinger kan ses i bilagene 5.1.7.1. og 5.1.7.2.

5.2 360 graders feedback

5.2.1 Definition af 360 graders feedback

360 graders feedback er en kompetencevurderingsmetode, hvor den uddannelsessøgende lægers samarbejdspartneres vurdering af lægens adfærd i forskellige samarbejdsrelationer sættes i relation til yngre læges vurdering af egen adfærd. Metoden indebærer således indhentning af feedback fra lægens samarbejdspartnere^{65,66}.

5.2.2 360 graders feedbacks anvendelighed

Metoden anvendes i Danmark overvejende til at foretage en helhedsvurdering af den uddannelsessøgende læges kompetence indenfor rollerne samarbejder, kommunikator, leder/administrator og professionel. Principielt kan alle lægeroller vurderes med en 360 graders feedback, som det f.eks. er gjort i en model udviklet i Videreuddannelsesregion Syd⁶⁷. Nogle af de bedst validerede internationale skemaer anvendes også til de øvrige roller.

Skemaer/spørgeramme

Mange forskellige skemaer har været anvendt, afhængigt af formålet med 360 graders feedback (summativ, formativ) samt af hvilke kompetencer, der ønskes vurderet. De fleste skemaer som anvendes, indeholder en række spørgsmål til lægens adfærd, som respondenterne og lægen selv besvarer på en flertrinsskala. De fleste skemaer har plads til, at respondenterne kan skrive åbne kommentarer. Der kan også anvendes en model for 360 graders feedback, hvor skemaerne er helt uden "skala-spørgsmål" og udelukkende består af kommentarer til en række kompetencer, en såkaldt narrativ model.

I de 2 sammenlignende undersøgelser som foreligger af en skala model versus en narrativ model foretrækker uddannelsessøgende læger narrative kommentarer, mens respondenterne helst vil nøjes med at sætte et kryds på en skala^{68,69}.

Hvis der anvendes et skema, som indeholder bedømmelse på en skala, er det nødvendigt med minimum ca. 8 respondenter, mens antallet af nødvendige respondenter ved anvendelse af en ren narrativ model ikke kendes. Besvarelserne kan indhentes i et elektronisk format eller på papir. Svarrapport udarbejdes typisk i et format hvor den uddannelsessøgende læges selvevaluering kan sammenstilles med svarene fra de forskellige respondentgrupper.

Respondenterne.

Hvordan respondenterne udvælges vil ofte afhænge af de logistiske forhold og praktiske hensyn. Det væsentligste er, at respondenterne har et tilstrækkeligt kendskab til den uddannelsessøgende læge, men også andre forhold kan have betydning⁷⁰.

Respondenterne udvælges oftest af lægen selv, men man kan anvende en model hvor andre udvælger respondenterne eller et fast respondentpanel. Der er fordele og ulemper ved begge metoder. Respondenterne kan være anonyme eller ikke anonyme⁶⁵.

Feedbacksamtale efter indhentning af data.

En vejleder, som er uddannet i feedbackgivning efter 360 graders feedback, afholder en feedbacksamtale med uddannelseslægen⁷¹. Vejlederen skal kunne forholde sig objektivt til data og er ikke selv respondent. Samtalen skal give den uddannelsessøgende læge lejlighed til at reflektere over de indkomne vurderinger herunder eventuel diskrepans mellem lægens selvsvurdering og respondenternes vurdering. Samtalen skal indeholde en stillingtagen til udviklingsmuligheder samt en aftale om hvorvidt og i så fald hvornår 360 graders feedback skal gentages. Oftest vil uddannelsesprogrammet for et givent forløb indeholde en oversigt over, hvornår 360 graders feedback skal gentages. Indsatsområderne kan med fordel formuleres på skrift, så de kan indgå i vejledersamtaler.

5.2.3 Fordele ved 360 graders feedback

Lægen vurderes i det virkelige liv og af mange forskellige personer. Mange kompetencer, som ellers kan være vanskelige at bedømme, kan evalueres samtidig med et relativt beskedent tidsforbrug. Feedbacksamtalen, som altid skal være en del af processen, giver mulighed for at drøfte problemstillinger som – måske fordi de er vævet sammen med personlige karaktertræk – ellers kan være vanskelige at nærme sig.

5.2.4 Ulemper ved 360 graders feedback

Metodens validitet er ikke veldokumenteret. Det diskuteres, om metoden faktisk måler den kompetence vi ønsker at måle eller om den (også) måler andet - uafhængigt af egentlige lægeprofessionelle kompetencer^{70,72,73}.

5.2.5 360 graders feedbacks validitet, pålidelighed og gennemførlighed

I et systematisk review af kompetencevurderingsmetoder fandtes 360 graders vurdering at være den eneste metode, som er vist at kunne føre til ændring i adfærd⁷⁴. Det betyder dog ikke, at andre metoder ikke kan have effekt. Der er i litteraturen udbredt enighed om at validiteten er så usikker, at metoden ikke bør anvendes som summativt evalueringsredskab.

5.2.6 Praktisk anvendelse af 360 graders feedback

- Definer hvilke kompetencer, der skal vurderes.
- Brug et allerede valideret skema eller valider det skema, der skal anvendes.
- Beslut hvordan respondenter udvælges. Hvem udvælger? Er det den uddannelsessøgende læge, uddannelsesansvarlig overlæge/vejleder, eller et fast respondentpanel?
- Distribuer skemaer til respondenterne.
- Analyse af data.
- Feedbackgivende vejleder udleverer svar-rapporten til uddannelseslægen umiddelbart før feedbacksamtalen.

- Feedbacksamtalen skal give plads til en fælles refleksion over resultatet herunder en eventuel diskrepans mellem lægens selv-vurdering og respondenternes vurdering.
- Feedbacksamtalen skal resultere i en udviklingsplan, som skal indgå i den uddannelsessøgendes individuelle uddannelsesplan.
- Foretages en beslutning om hvorvidt vurderingen skal gentages.

I Region Nord anvendes en elektronisk model, der er baseret på en valideret spørgeskræmme til brug for alle specialer^{75,76,77}. Videreuddannelsesregion Nord har udarbejdet en praktisk vejledning i gennemførelse af denne model⁷⁵.

Eksempler på internationalt anvendte skemaer, skema fra Uddannelsesregion Nord samt skema fra forskellige specialer ses i bilagene 5.2.6.1., 5.2.6.2., 5.2.6.3., 5.2.6.4., 5.2.6.5., 5.2.6.6., 5.2.6.7. og 5.2.6.8.

6 Retrospektive metoder

6.1 Audit

6.1.1 Definition af audit

En audit er en struktureret bedømmelse af kvaliteten ud fra i forvejen opstillede kriterier. Den traditionelle auditproces bygger på en sekvens inkluderende: fastsættelse af standarder, review, implementering af forandringer og reevaluering⁷⁸. Audit betyder revision eller høring.

Audit bruges i mange sammenhænge også indenfor det danske sundhedsvæsen. Det anvendes ofte sammen med kvalitetssikring og kvalitetsudvikling. Audit er velkendt i evalueringsprocesser i kontinuerlig medicinsk uddannelse. Audit i denne henseende har hovedvægten på sammenligning af egen arbejdspræstation mod givne standarder og efterfølgende analyse for at afdække områder for personlig faglig udvikling⁷⁹.

6.1.2 Audits anvendelighed

I dansk kontekst anvendes metoden i den medicinske kompetencevurdering både i en summativ og formativ sammenhæng. Metoden er egnet til at bedømme teoretisk kontekstuel viden, dvs. bedømme viden i den sammenhæng den indgår i. Metoden er egnet til identifikation af læringsmål i den givne kontekst⁸⁰.

Audit kan bruges i flere forskellige opsæt i klinikken:

Ekstern bedømmelse af den uddannelsessøgende læge, oplæg til læringsramme gives af vejleder:

- Auditprocessen styres af vejlederen, hvor denne gennemgår et antal ens elementer den uddannelsessøgende læge både har haft procesledelsen for og dokumenteret (eks. journal, obduktionsrapporter, erklæringer). Der foretages vurdering mod givne standarder og derefter tilbagemelding til lægen enten mundtlig eller skriftligt.

Den uddannelsessøgende læge bedømmer egen arbejdspræstation mod givne standard, oplæg til læringsramme for egen udvikling gives af lægen efter selvrefleksion i samarbejde med vejleder:

- Auditprocessen styres af den uddannelsessøgende læge, hvor denne samler et antal egne elementer til en auditproces omhandlende en given problemstilling. Lægen gennemgår materialet mod givne standard, udarbejder en refleksion mundtlig eller skriftligt over sin egen faglige adfærd i henhold til den givne problemstilling f. eks relaterende sig til kliniske retningslinjer eller organisatoriske forhold. Dette oplæg danner grundlag for efterfølgende refleksiv struktureret samtale og vurdering. I evalueringsprocessen kan der være udarbejdet standard for vejlederens vurderingsproces.

Den uddannelsessøgende læge bedømmer andres og egne arbejdspræstationer mod givne standard, oplæg til læringsramme for egen og organisationen gives af den uddannelsessøgende læge og vejleder i samarbejde⁸¹.

- Auditprocessen styres af den uddannelsessøgende læge, som udvælger udarbejdede elementer, omfattende egne (fåtal) og kollegers. Auditanalysen udføres i forhold til formuleret standard for hele auditprocessen. På baggrund af analyse, udfører lægen en rapport/et mundtligt oplæg. Her forholder lægen sig til kvalitet i kontekst af egen viden og organisatoriske forhold efter given evalueringsramme. Rapport/oplæg udmunder i konklusion og forslag til forbedring for afdelingen – inkl. den uddannelsessøgende læge selv. Rapport/oplæg evalueres af vejleder efter en given standard for indhold. Resultatet gennemgås derefter ved struktureret samtale⁸².

6.1.3 Fordele ved audit

Audit er let anvendelig, kan tilrettelægges i den kliniske hverdag, når det er relevant i forhold til uddannelsesprogram. Audit kræver ikke nødvendigvis mødetid mellem vejleder og den uddannelsessøgende læge ved selve vurderingen af materialet. Feedback kan om nødvendigt gives skriftligt. Bedømmelsesprocessen er klar og intuitiv og med øget kendskab til evalueringsforløbet øges accept af metoden.

6.1.4 Ulemper ved audit

Audit processen afspejler ikke nødvendigvis den udførte kliniske praksis, der ligger til grund for auditmaterialet (journaler, indstillinger, rapporter).

Ved peer review alene kan den uddannelsessøgende læges bevidsthed overfor egne mulige lærings-potentialer mangle og dermed mindskes udbyttet af auditprocessen⁸².

De enkelte faser i auditprocessen bør være klart beskrevne, så den uddannelsessøgende læge får et tilstrækkeligt udbytte. Ofte mangler der klare indikatorer for godkendte og ikke godkendt forløb.

6.1.5 Audit, validitet, pålidelighed og gennemførlighed

Pålidelighed og validitet er vurderet i enkelte studier som værende tilfredsstillende. Pålidelighed og validitet øges ved standardisering af processen og oplæring af bedømmerne i auditprocessen⁸⁰. Noget tyder på, at det antal af journaler, der gennemgås er af betydning, men dette punkt er uklart. I akkrediterings-processer vælges oftest mellem 15 og 20, da dette antal menes, at kunne afdække mangler⁷⁹.

6.1.6 Praktisk anvendelse af audit

Et praktisk eksempel på anvendelse af audit er eksemplet Medicinsk gennemgang af patient fra det geriatriske speciale, hvor audit af vurdering af en enkelt fælles intern medicinsk kompetence foretages (bilag 6.1.6.1.). Den uddannelsessøgende læge vælger 5 journaler, og vurderer de indeholdte ”gennemgange” ved brug af dette skema. I mindst 2 af journalerne skal lægen selv have foretaget ”gennemgang af patienten”, som er den kompetence, der vurderes.

Den uddannelsessøgende læge forbereder en mundtlig fremlæggelse eller udfærdiger en kort skriftlig rapport over de 5 journaler til den speciallæge, der skal vurdere. Hvad enten man vælger en mundtlig fremstilling eller en skriftlig rapport, vil audit være udgangspunkt for en struktureret vejledersamtale. Strukturen i vejledersamtalen følger auditskemaets punkter.

Rapporten skal indeholde:

1. Kort beskrivelse af resultat af audit resultater.
2. Diskussion af resultatet: er det tilfredsstillende, lever det op til afdelingens instruks/praksis. Har organisatoriske forhold haft indflydelse på resultatet?
3. Konklusion og evt. forslag til forbedringer.

Rapporten vurderes af kompetencevurderende speciallæge på nedenstående kriterier:

1. Klar og fyldestgørende beskrivelse af resultat af audit.
2. Diskussion af resultaterne i relation til god medicinsk praksis og organisatoriske overvejelser.
3. Klar konklusion i overensstemmelse med fund.

Et eksempel på udformning af audit opgave kan ses i bilag 6.1.6.2.

6.2 Refleksive rapporter

6.2.1 Definition af refleksive rapporter

Refleksion kan defineres som bevidst kritisk analyse af viden og erfaring. Analysen udføres for at opnå dybere forståelse⁸³. Refleksion kan beskrives som en gentagne proces hvor egen praksis vurderes medens den udføres, eller umiddelbart efter, bedst gentaget over lignende praksis. En dybere refleksionsproces kræver evnen til association til andre processer, integration og vurdering af denne refleksive viden, og evne til bevidst at tilegne sig denne viden⁸⁴.

I postgraduat medicinsk uddannelse anvendes refleksion med tre hovedformål⁸⁵:

1. Refleksion over læring: Erfaring fra klinisk arbejde fortolkes og inddrages i den viden den uddannelsessøgende har opnået og danner grundlag for videnssøgning forud for videre klinisk arbejde (Kolbs læringcyklus 1984).
2. Refleksion over læge-patient forhold: Læge-patient forhold er en væsentlig del af patientbehandling og har betydning for patienttilfredshed, forløb af kroniske tilstande og behandlingsgennemførelse⁸⁵. Udvikling af evne til at opbygge tilfredsstillende læge-patient forhold er essentiel i postgraduat uddannelse og kræver oftest guidet refleksion med involvering af vejleder/mentor.
3. Refleksion over udvikling af professionel praksis: Læger agerer i dagligdagen i mange forskellige, komplekse og ikke veldefinerede opgaver. Evnen til at agere og tage beslutninger i denne hverdag er en væsentlig del af den uddannelsessøgendes professionelle udvikling. Guidet refleksion kan optimere læring fra de gentagne komplekse arbejdsopgaver.

Refleksionens hovedformål er, at være en bevidst proces der understøtter en forståelse for forbedret handling i kommende arbejdsopgaver.

Refleksive rapporter er skriftlige opgaver relevante for den uddannelsessøgende læges aktuelle niveau og kontekst, dvs. omhandler arbejdsopgaver med passende sværhedsgrad og aktualitet. Opgaveformulering stimulerer til refleksion over et eller flere af de ovennævnte hovedformål.

Eksempler på typer af refleksive rapporter:

- Skriftlig rapport over arbejdssituation
- Kritisk hændelsesrapport
- Læringsdagbog
- Personlige refleksive uddannelsesplaner

6.2.2 Refleksive rapporters anvendelighed

Refleksive rapporter understøtter den uddannelsessøgende læges evne til at vurdere egen-udvikling og refleksion over handlinger. Et væsentlig element i brug af refleksive rapporter er vejlederens indsats ved vurdering af rapporten, især feedback på opgaven er essentiel for værdien af opgaveskrivningen⁸³.

Rapporten kan vise den uddannelsessøgende læges specielle potentiale indenfor selektive områder af det pågældende fag/område og dermed danne grundlag for videre karrierevejledning. Rapporten kan også afdække manglende kompetencer hos den uddannelsessøgende læge, og dermed give mulighed for korrektion. Refleksive rapporter stimulerer til fortsat refleksion over fagligt udviklingsbehov, til postgraduat kontinuerlig udvikling. Rapporten kan desuden påpege behov for revision eller udfærdigelse af behandlingsretningslinjer samt behov for forskning i specifikke områder, og kan dermed have organisatoriske konsekvenser.

6.2.3 Fordele ved refleksive rapporter

Refleksive rapporter sikrer en egen vurdering hos den uddannelsessøgende læge over de tre elementer i refleksion som beskrevet herover. Rapporten giver et sikkert grundlag for faglig gennemgang af konkrete emner når vejlederen giver den uddannelsessøgende læge feedback. Opgaveformuleringen er afgørende for hvilke af lægerollerne, der fokuseres på. Undersøgelser viser, at uddannelsessøgende læger finder rapporterne værdifulde når opgaveformuleringen passer til uddannelseskonteksten. Erfaringen viser, at rapporten kan føre til faglig udvikling i en afdeling/praksis i form af undervisning baseret på rapporten, udarbejdelse/revision af retningslinjer, forskning eller audits⁶⁴.

6.2.4 Ulemper ved refleksive rapporter

Rapporten er tidskrævende for den uddannelsessøgende læge. Gennemgang og feedback fra vejleder er tidskrævende og kræver fastsatte tidsrammer.

6.2.5 Refleksive rapporter validitet, pålidelighed og gennemførlighed

Validitet af refleksive rapporter er afhængig af opgaveformulering. De få studier, der foreligger på anvendelse af skriftlige rapporter i postgraduat uddannelse viser, at tilfredsstillende validitetsvurdering kan opnås^{64,86}. Konstruktions- og indholdsvaliditet øges ved at flere eksperter på aktuelle område deltager i opgaveformulering. Pålidelighed er afhængig af vejlederne, hvorfor uddannelse af vejledere til denne opgave og diskussion af standard for godkendelse af opgave anbefales^{22,86}. Gennemførlighed er afhængig af om den uddannelsessøgende og vejleder forstår formålet med anvendelse af rapporten. Gennemførligheden fremmes når niveau og

kontekst i opgaven er i overensstemmelse med den uddannelsessøgende læges niveau, og når opgaven er planlagt som obligatorisk kompetencevurdering.

6.2.6 Praktisk anvendelse af reflektive rapporter

En skriftlig opgave af begrænset omfang fastlægges til et specifikt tidsrum i speciallægeuddannelsen. Opgaven bør omhandle en praksis som den uddannelsessøgende læge møder i dette tidsrum og hvor opgaven kan løses ved forskellige metoder. En opgavebeskrivelse udformes, hvor den uddannelsessøgende skal beskrive egen erfaring og reflektere over andre metoder til løsning, herunder tage stilling til evidensniveau.

Eksempler på reflektive opgavebeskrivelser kan ses i bilagene 6.2.6.1., 6.2.6.2., 6.2.6.3., 6.2.6.4., 6.2.6.5., 6.2.6.6., 6.2.6.7. og 6.2.6.8.

6.3 Casebaseret diskussion

6.3.1 Definition af casebaseret diskussion

Casebaseret diskussion er en struktureret vejledersamtale med henblik på kompetencevurdering af, og feedback til en uddannelsessøgende læge inden for områder som klinisk ræsonnement, beslutningstagning og anvendelse af medicinsk viden i relation til patientbehandlingen^{87,88}.

Casebaseret diskussion fungerer også som en metode til at dokumentere samtaler om uddannelsessøgende læges præsentationer af kliniske tilfælde. Det skal basere sig på et skriftligt materiale (journalnotat, epikrise eller lignende).

Casebaseret diskussion foregår mellem en uddannelsessøgende læge og en vejleder – og svarer altså ikke til en normal klinisk konference.

”*Case based discussion*” er den angelsaksiske version af den amerikanske ”*Chart-Stimulated Recall*”, som anvendes i Nordamerika og lande, der læner sig op af en amerikansk uddannelsestradition. Publikationer vedr. metoden skal primært søges under ”*Chart-Stimulated Recall*”.

Casebaseret diskussion er blevet meget udbredt i UK, hvor det siden 2007 har indgået som én af metoderne i deres kliniske basisuddannelse (”*Foundation years*”) og som én af fire obligatoriske metoder i alle specialers speciallægeuddannelse. De andre 3 metoder er DOPS (*Directly Observed Procedural Skills*), mini-CEX, og 360 graders feedback).

6.3.2 Casebaseret diskussion anvendelighed

Casebaseret diskussion kan anvendes på alle uddannelsesstrin og - med de britiske erfaringer in mente - i alle specialer.

Da de fleste cases indebærer mulighed for vurdering af en flæthed af kompetencer og lægeroller, kan casebaseret diskussion vinkles i forhold til dette. Denne vinkling kan vælges af uddannelsessøgende læge, af vejlederen eller i fællesskab.

Metoden kræver ikke meget forberedelse: det skriftlige materiale er allerede til stede, og den uddannelsessøgende læge vælger blot nogle få cases på forhånd. Vejle-

deren skal anvende tid til diskussionen sammen med den uddannelsessøgende læge.

6.3.3 Fordele ved casebaseret diskussion

Metoden er umiddelbart forståelig for såvel uddannelseslæge som vejleder. Den kræver minimal forberedelse. Metoden vurderes at give god mening for såvel den uddannelsessøgende læge som vejleder. Metoden tager sit afsæt i handlinger allerede udført af lægen – og ikke i hvordan lægen hypotetisk ville agere i fremtiden.

Metoden kan anvendes i alle specialer og på alle uddannelsestrin.

6.3.4 Ulemper ved casebaseret diskussion

Diskussionen baserer sig på skriftlige notater og den uddannelsessøgende læges opfattelse af casen, hvilket kan medføre en risiko for bias, idet vejleder ikke direkte har observeret omstændighederne i den kliniske case.

6.3.5 Casebaseret diskussion validitet, pålidelighed og gennemførlighed

Adskillige studier har undersøgt metodens validitet og pålidelighed^{89,90,91} og fundet disse tilfredsstillende. Casebaseret diskussion blev i et studie også opfattet som den mest valide vurderingsmetode blandt de deltagende udøvende læger⁹¹. Et studie påpeger, at anvendelse af begrebsvaliderede skalaer ved casebaseret diskussion øger pålideligheden af metoden⁹².

I blandt andet almen medicin i UK anbefales det, at vejlederen anvender en struktureret spørgeguide – og at vejlederen under samtalen i et skema noterer dokumentation for uddannelseslægens kompetenceniveau inden for de forskellige områder, der søges afdækket.

Casebaseret diskussion er gennemførlig i alle specialer, hvis man vil investere den fornødne tid til feedback. I såvel USA som i Canada anvendes metoden bredt i mange specialer. I UK anvendes metoden i alle specialer – og der er publikationer om metodens anvendelse fra ortopædkirurgi over psykiatri til almen medicin. I Danmark indgår metoden i tillempet form i kompetencevurderingsmetoden ”Den strukturerede vejledersamtale” (bl.a. i almen medicin).

Som eksempel på hvilket antal gennemførte casebaserede diskussioner, der anbefales i et uddannelsesforløb kan anføres, at man i almen medicin i UK i begyndelsen af uddannelsesforløbet har 6 per år – og i det sidste uddannelsesår 12 casebaserede diskussioner.

6.3.6 Praktisk anvendelse af casebaseret diskussion

Den uddannelsessøgende læge udvælger 2-4 cases til diskussion med vejlederen; vejlederen udvælger én af disse cases, som så anvendes. Vejleder eksplorerer så denne case i løbet af 20 minutter inden for ét eller flere områder som:

- klinisk beslutningstagning
- undersøgelse/henvisning
- behandling
- opfølgning

- professionalisme
- skriftlig dokumentation

Dernæst anvendes 5-10 minutter til konstruktiv feedback.

Et emne fra dagligdagen kunne for en speciallæge i pædiatri være indlæggelse af barn med astmatisk bronchitis eller for almen medicin konsultation med en patient med nykonstateret type 2 diabetes.

Et eksempel på casebaseret diskussion kan ses i bilag 6.3.6.1.

7 Kompetencevurdering med anvendelse af simulation

7.1 Simulation

7.1.1 Definition af simulation

Simulationsbaseret medicinsk uddannelse kan i sin bredeste form defineres som enhver uddannelsesaktivitet som anvender simulationsbaserede udstyr og metoder med det formål at skabe en læringsmulighed for deltagerne⁹³. Den hurtige teknologiske udvikling gør det muligt at udvikle ”værktøj” som kan anvendes i alle stadier af den medicinske uddannelse – den professionelle udvikling.

- Færdighedstrænere
- Simulerede patienter
- VR simulatorer
- Kirurgiske eller procedurale simulatorer
- Patientsimulatorer eller (mannikin baserede)

Simulation har i Danmark overvejende været anvendt til træning, men både de simple og avancerede kirurgiske simulatorer gør det muligt at træne og vurdere kompetence inden den uddannelsessøgende møder patienten. Andre lande har tidligt interesseret sig for at anvende simulerede omgivelser til at vurdere kompetence.

7.1.2 Simulations anvendelighed

Interessen og behovet for træning i færdighedslaboratorier vil være stigende de kommende år (S&P) og konceptet¹ ”alt hvad der kan trænes skal trænes” vender større og større indpas. Fordelen for den uddannelsessøgende læge er, at uddannelsen her kan have højere prioritet end patientbehandlingen – speciallægen er ikke nødt til at tage over i risikofyldte situationer. En fordel ved simulation er også, at det er muligt at give konstruktiv feedback i den efterfølgende feedback, hvor klip af evt. suppleret videooptagelser kan illustrere læringsmål. Den meget strukturerede feedback efter scenarie baseret træning stimulerer til refleksion over egne kompetencer og samspil med andre. Det betyder, at både rollen som medicinske ekspert/lægefaglig og andre roller som kommunikator, samarbejder og leder/administrator/organisator kan vurderes.

I Danmark anvendes kirurgiske simulatorer mere og mere til formativ vurdering af kirurgiske kompetencer.

Patientsimulatorer anvendes i anæstesiologi til formativ vurdering af kompetencer efter visse af de specialespecifikke kurser. Her er udviklet en objektiv struktureret clinical examination (OSCE), som afholdes ved afslutning af kurset i luftvejshåndtering. Ved afslutningen af det sidste kursus i uddannelsen afholdes en team baseret OSCE, hvor 2 deltagere sammen gennemgår stationer, som repræsenterer specialtets fire søjler (anæstesiologi, intensiv terapi, smerte og præhospital behandling). I Danmark indgår simulationerne primært i en formativ vurdering, men på sigt kan en summativ vurdering overvejes.

I andre lande, som f.eks. UK og Israel anvendes patientsimulation i stigende grad i forbindelse med summativ vurdering af den uddannelsessøgende. Således indgår simulationsscenerier i del 1 af den engelske anæstesiologiske specialist eksamen og i sluteksamen for anæstesiologer i Israel^{94,95,96}.

7.1.3 Fordele ved simulation

Fordelen ved at anvende simulation til kompetencevurdering er, at casen og kontekst i meget høj grad kan standardiseres. Patientcasen kan udvikle sig efter et standardiseret script, så den udvikler sig på samme måde over tid, hvilket betyder at de uddannelsessøgende får samme situation/samme betingelser. Medhjælperne kan agere på samme måde, tidsaspektet er det samme, så tid til behandling mv. kan registreres. Simulationen gør det muligt at ”kopiere” en arbejdssituation og gør det muligt at foretage kompetencevurderingen på højt niveau (kan vise frem for at have viden om). Casen kan videoptages, og kompetencevurderingen kan foretages af flere forskellige bedømmere. Checklister udformes og valideres med henblik på at gøre det nemmere for bedømmerne. Hvis der er tale om summativ vurdering (eksamen), skal der foretages test af såvel pålidelighed som validitet. Bedømmerne skal trænes indtil en acceptabel pålidelighed mellem bedømmere og over tid, er opnået.

7.1.4 Ulemper ved simulation

De væsentligste ulemper er udgifterne til udstyr og de menneskelige ressourcer, der er nødvendige for at kunne afvikle casen.

7.1.5 Simulations validitet, pålidelighed og gennemførlighed

Kompetencevurderingsredskaberne skal på samme måde som i klinikken valideres og bedømmerne skal trænes. Ofte valideres indhold af checkliste ved en Delphi metode. Validerede redskaber fra klinikken (som f.eks. NOTSS) kan også anvendes i simulerede omgivelser. Ved anvendelse af NOTSS er en højere pålidelighed vist, når bedømmere skal vurdere video optagede scenarier end af virkelige cases i klinikken.

7.1.6 Praktisk anvendelse af simulation

En objektiv struktureret klinisk eksamination (OSCE) er udviklet i anæstesiologi, som afholdes ved afslutning af kurset i luftvejshåndtering. I alt 10 stationerne er udviklet, hvor viden og færdigheder vurderes ved forskellige metoder, hvoraf simulation er en af disse. Kursisterne har 7 minutter ved hver station. Kursisterne får umiddelbar feedback på præstationen og får til slut svar på, om de er bestået. Kursisterne udfylder ved afslutningen af kurset en læringsplan, som de efterfølgende præsenterer for den uddannelsesansvarlige overlæge i egen afdeling. Herved kan der følges op i klinikken.

8 Konklusion

Kompetencevurderingsmetoder

Denne rapport giver en systematisk oversigt over anbefalede metoder til kompetencevurdering, der fremadrettet kan anvendes af de videnskabelige selskaber. De udvalgte metoder er evidensbaserede eller valgt ud fra bedste praksis på området. Det er således ikke en fuldstændig oversigt over eksisterende kompetencevurderingsmetoder. Rapporten understreger vigtigheden af, at differentiere mellem vurdering af enkelte trin i en procedure, en samlet vurdering af komplekse arbejdssituationer eller en helhedsvurdering af den uddannelsessøgende læge. Rapporten er i sin opbygning orienteret mod praktisk anvendelse og giver mulighed for at specialeselskaberne kan hente inspiration til at udvikle og validere kompetencevurderingsmetoder. Et afsnit om skalaer og et om validitet, pålidelighed og gennemførlighed vil også være en hjælp til dette.

Overordnet set har specialeselskaberne valgt anbefalede metoder til kompetencevurdering og videreudviklet disse på baggrund af evidensbaseret medicinsk-pædagogisk viden i samarbejde med den pædagogisk udviklende funktion og andre eksperter. Flere af disse metoder er også valideret til dansk kontekst. Der er imidlertid en stor forskel på specialernes valg af metoder og i hvilken grad kompetencevurdering er implementeret. Nogle specialer er længere fremme end andre. Dette kan skyldes, at disse har haft bedre pædagogisk/organisatorisk hjælp til opgaven, men det kan også skyldes en anden kultur og organisering i de specialer som ikke har implementeret kompetencevurdering. En særlig indsats er nødvendig overfor disse specialer.

Det konkluderes endvidere, at kompetencevurderingen i stigende grad vil inkludere vurdering af flere lægeroller og ikke alene den medicinske ekspert rolle. Dette vil medføre en mere holistisk tilgang til kompetence, der formentlig er mere i tråd med lægeligt arbejde end den til tider lidt kunstige fragmentering af lægeligt arbejde i de 7 lægeroller, som har været anvendt tidligere. I princippet erstatter dette tidligere tiders detaljerede mål for kompetence og integrerer medicinsk ekspertise med de øvrige 6 lægeroller. Såvel nationalt som internationalt har der været en tendens til at fragmentere mål for kompetence i stedet for at se mere holistisk på kompetence. Trenden går nu mod at udvikle metoder, som kan vurdere kompetence mere overordnet.

Betydning af kultur og kontekst

Valg af kompetencevurderingsmetode er dog ikke i sig selv en sikkerhed for at disse anvendes konsekvent i den kliniske hverdag. Kontekst og kultur er af væsentlig betydning. Konteksten er en travl klinisk hverdag hvor patientbehandlingen er i fokus. Det er derfor nødvendigt også at forholde sig til praktiske forhold, når metoden skal udvikles eller udvælges. Metoder, der kan appliceres omkring en klinisk situation og som kan gøres uden et væsentligt tidsforbrug, er at foretrække.

Kompetencevurderingen er i meget høj grad styrende for læring. En positiv kompetencevurderingskultur i afdelingerne og praksis er en forudsætning for en velfungerende speciallægeuddannelse. Ledelse, uddannelsesansvarlige og vejledere må forstå betydningen af, at foretage en omhyggelig kompetencevurdering og efterfølgende af, at give konstruktiv feedback til den uddannelsessøgende læge. Tid til feedback skal indbefattes. Det er afdelingens ansvar, at rammerne for at opnå kom-

petencer og blive kompetencevurderet er til stede. Den uddannelsessøgende læge har ansvar for at forberede sig til kompetencevurderingerne og at medvirke til at planlægge disse – og et generelt ansvar for egen læring.

Kendskab til de forskellige metoder, forståelse af rollen som bedømmer samt evne til at kunne give konstruktiv feedback er en forudsætning for at kompetencevurderingen kan foregå på en acceptabel måde. En formel uddannelse af uddannelsesansvarlige nøglepersoner og uddannelse af bedømmerne vil styrke kompetencevurderingen. De generelle principper for forståelse for uddannelse og kompetencevurdering kan foregå på et kursus, hvor uddannelsesansvarlige fra flere specialer kan deltage sammen og dele erfaringer mellem specialer. Men specialespecifikke kurser for bedømmere vil være en fordel, så disse kan trænes i at anvende kompetencevurderingsmetoder fra eget speciale. Koncepter for sådanne kurser kan være nationale.

Erfaringer fra den meget strukturerede feedback efter simulationsbaseret træning kan med fordel forsøges anvendt i oplærings- og kompetencevurderingssituationer. Det stigende antal vejledere, der har gennemgået en instruktøruddannelse for at kunne medvirke til simulationsbaseret træning, vil kunne medvirke til dette.

9 anbefalinger og perspektivering

Fremtidige valg af kompetencevurderingsmetoder kan med fordel udvikles i en proces bestående af 3 trin:

- Design (indhold og *face* validitet)
- Validering (begrebs- og kriterievaliditet samt pålidelighed)
- Implementering (gennemførlighed, acceptabilitet, effekt på uddannelse, cost-effectivitet samt prediktiv validitet)²¹

Internationalt er der enighed om, at værktøjskassen skal indeholde flere metoder og at ny viden om vurdering i arbejdssituationer og tankegangen omkring Entrustable Professional Activities (EPA) i højere grad skal anvendes. Kompetencevurdering er mere end design og validering. Fokus på aspekter af kompetencevurdering knyttet til implementering og praktisk gennemførlighed såsom rammer, muligheder og acceptabilitet er nødvendigt. Metoden kan være valideret i eksperimentel kontekst, men spørgsmålet er nu om den er generelt brugbar i klinisk kontekst. Translational forskning ("fra laboratorium til klinisk kontekst") er nødvendig til at besvare spørgsmål som: Opfatter brugeren (uddannelsessøgende og bedømmer) metoden som brugbar? Giver den uddannelsesmæssig værdi således, at der skabes læring? Betyder fokus på visse kompetencer, at andre ikke læres? Er den økonomisk rentabel? Et eksempel herpå kan være, hvorvidt træning på simpel bænkmødel vs. træning på mere avanceret simulator kan forudsige, hvordan den uddannelsessøgende læge performer på sigt.

Desuden er der behov for at undersøge, om de samlede modeller som specialeselskaberne har opstillet for kompetencevurdering og som angives i målbeskrivelserne, er pålidelige og valide i forhold til at sikre, at den nyuddannede speciallæge er kompetent og kan arbejde på speciallægeniveau efter endt uddannelse. Det anbefales derfor, at Sundhedsstyrelsen sammen med Lægevidenskabelige Selskaber (LVS) sikrer, at specialeselskaberne har adgang til den nødvendige viden herunder mulighed for konsulentbistand fra særlig uddannede indenfor medicinsk pædagogik.

På baggrund af erfaringerne fra denne rapport anbefales følgende indsatsområder:

- En kombination af faglig og medicinsk pædagogisk indsigt er en forudsætning for udarbejdelse af målbeskrivelser og udvikling/udvælgelse af kompetencevurderingsredskaber. Pædagogisk vejledning til specialeselskaberne er nødvendig specielt til mindre specialer og specialer, som ikke i tilstrækkelig grad har implementeret kompetencevurdering. Specialerne bør kunne trække på ekspertise fra andre specialer.
- Udvikling af nye metoder og videreudvikling af eksisterende metoder, specielt med hensyn til kompetencevurdering bør være mere holistisk. Herunder bør der udvikles metoder, som kan fungere i en travl klinisk hverdag. Ligeledes bør det sikres, at flere af de 7 lægeroller integreres i kompetencevurderingsredskaberne.
- Validering af kompetencevurderingsmetoder i dansk kontekst og involverende en bredere palet af specialer.

- Forskning i kompetencevurdering og feedback, herunder dennes effekt på læring, samt translationel forskning, som kan medvirke til at give viden om udfordringer ved anvendelsen af metoderne, og medvirke til en hurtigere implementering af kompetencevurdering i speciallægeuddannelsen. Udvikling af specialespecifikke vejleder- og bedømmerkurser regionalt, gerne baseret på nationalt koncept.
- Styrkelse af den medicinsk-pædagogiske kompetence i specialerne.
- Sikring af de nødvendige rammer for kompetencevurdering på de enkelte afdelinger/praksis, hvilket kræver en ledelsesmæssig prioritering af området.
- Styrkelse af de eksisterende netværk mellem uddannelsesansvarlige nøglepersoner på regionalt og nationalt niveau både i og mellem specialer.

10 Ordliste

Begreb	Begrebets anvendelse i denne rapport
Kompetencevurdering	<p>En vurdering af den enkelte læges kompetence på et givet område, som omfatter alle 3 elementer af kompetencebegrebet: 1) viden og færdigheder, 2) evne til at omsætte viden og færdigheder samt udnytte personlige evner i praksis og 3) personlige evner og holdninger.</p> <p>Kompetencevurdering kan anvendes summativt eller formativt (s.d.), men opfølges altid af feedback.</p>
Kompetence	<p>Den lægelige videreuddannelse bygger på målbeskrivelser, der beskriver de kompetencer, som en speciallæge skal have. Begrebet kompetence er lægens evne til at udføre de opgaver, og udfylde de roller, som forventes på et givet tidspunkt i lægens professionelle udvikling. I rapporten anvendes kompetence i forståelsen: viden, færdigheder og holdning. Kompetence omfatter derfor</p> <ol style="list-style-type: none">1. Viden og færdigheder2. Evne til at omsætte viden og færdigheder samt udnytte personlige evner i praksis. Dette refererer til det, der de facto gøres i en konkret situation.3. Personlige evner og holdninger <p>Kompetence indeholder dermed ikke kun det en person ved og kan gøre under ideelle omstændigheder, men hvad en læge gør i det daglige arbejde. Kompetence er således kontekstafhængig og bygger på lægens erfaring.</p> <p>En kompetence er således ikke blot en lægefaglig færdighed (medicinsk ekspert), men indeholder elementer af de andre roller. På samme vis kan man ikke blot se på f.eks. samarbejder rollen, idet denne ikke har mening, uden at den vurderes i sammenhæng med udøvelse af lægefaget. Rollerne bliver således integreret i hinanden på en række måder.</p>

Summativ kompetencevurdering	Kompetencevurdering foretaget m.h.p. at afgøre, om en læge har opnået en given kompetence, og evt. på hvilket niveau. Med andre ord dækker det over en form for eksamen.
Formativ kompetencevurdering	Kompetencevurdering foretaget m.h.p. at styrke læringen, og ved hjælp af feedback at gøre den uddannelsessøgende opmærksom på egne styrker og svagheder indenfor den observerede del af arbejdet.
Bedømmer	Den person, der observerer og bedømmer den uddannelsessøgende under kompetencevurdering, og efterfølgende giver feedback. Dette vil typisk være den uddannelsessøgendes kliniske vejleder i den givne situation.
Uddannelsessøgende læge (US)	Den læge, hvis kompetence(r) bliver vurderet, benævnes i rapporten den "uddannelsessøgende læge"
Samlet vurdering	I visse kompetencevurderingsmetoder indgår der en samlet bedømmelse af den observerede aktivitet, udover bedømmelse af delelementer. Eksempelvis kan der ved vurdering af en udført konsultation indgå vurdering af anamneseoptagelse, objektiv undersøgelse og flere andre delelementer, og til sidst en samlet vurdering af lægens præstation ved konsultationen.
Helhedsvurdering	Med en helhedsvurdering menes en vurdering af "den hele læge", hvor så mange aspekter af lægens samlede kompetencer som muligt indgår. Dette kan f.eks. ske ved en 360 graders vurdering.
Færdighed	Bruges i rapporten i samme betydning som det engelske ord <i>skill</i> .
Dikotomisk skala	Skala med udelukkende to trin, f.eks. ja/nej, bestået/ikke bestået.
Vurderingsskala	Anvendes i rapporten i samme betydning som det, der i litteraturen hedder <i>rating scale</i> .
Skærende specialer	Lægelige specialer, hvori der udføres kirurgiske indgreb.
Portefølje	En samling af materiale, der vedrører den enkelte læges uddannelse, hvori der løbende gemmes information om, og dokumentation for, hvad lægen har lært, samt om planlagte læringstiltag. I porteføljen indgår derfor en logbog med dokumentation for opnåede kompetencer i målbeskrivelsen, kursusbeviser, operationslister mv. Information

	<p>om lærerige begivenheder, opnåede delkompetencer m.v. kan gemmes systematisk, så de senere kan bruges i forbindelse med kompetencevurdering, vejledningssamtaler eller kompetencegodkendelse, f.eks. ved mulighed for at tilføje noter eller dokumenter relateret til de enkelte kompetencer. Resultater af formativ kompetencevurdering kan også indgå i porteføljen. Den fremadrettede del indeholder f.eks., læringskontrakter, uddannelsesplaner og lign.</p> <p>Porteføljen kan med fordel bruges til at registrere information om al uddannelse og læring for den enkelte, dvs. også efteruddannelse mv.</p> <p>Porteføljen kan være elektronisk eller i papirform.</p>
--	--

Forkortelser:

S & P	Status og perspektiveringsrapporten vedr. den lægelige videreuddannelse.
US	Uddannelsessøgende læge
OSATS	Objective structured assessment of technical skills
NOTSS	Non technical skills for surgeons
ANTS	Anaesthesia non technical skills
MiniCEX	Mini clinical evaluation exercise
EPA	Entrustable Professional Activities

11 Referencer

- 1 Sundhedsstyrelsen. Speciallægeuddannelsen – status og perspektivering. 2012. www.sst.dk.
- 2 Sundhedsstyrelsen. Rapport om de 7 lægeroller. 2013. www.sst.dk.
- 3 Sundhedsministeriet. Fremtidens speciallæge. Betænkning fra speciallægekommisionen. Betænkning nr. 1384. København, 2000. www.sum.dk.
- 4 Søjnæs C, Lillevang G, Jørgensen RL, Ringsted C. Kvalitet i videreuddannelsen i Region Hovedstaden. CEKU publikation 2011.
- 5 Malling B, Bested KM, Skjelsager K, Ostergaard HT, Ringsted C. Long-term effect of a course on in-training assessment in postgraduate specialist education. *Med Teach*. 2007 Nov;29(9):966-71.
- 6 Jolly B, Grant J, Joint Centre for Education in Medicine. The good assessment guide: a practical guide to assessment and appraisal for higher specialist training. London. 1997.
- 7 Erault M & Du Boulay B. Developing the attributes of medical professional judgement and competence (Report to the department of Health). Department of Health London, UK, 2000.
- 8 Malling B, Mikines K, Hilsted L, Sorensen JL. De syv lægeroller anno 2012 – tid for revision? UFL 2013.
- 9 Ten Cate O & Scheele F. Competency-Based Postgraduate Training: Can We Bridge the Gap between Theory and Clinical Practice? *Acad Med*. 2007; 82:542–547.
- 10 Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med*. 2009 Mar;84(3):301-9.
- 11 Okuyama A, Martowiriono K, Bijnin B. Assessing the patient safety competencies of healthcare professionals: a systematic review. *BMJ Qual Saf* 2011;20:991-1000.
- 12 Bested KM, Malling B, Skjelsager K et al. Bedømmerbias i speciallægeuddannelsen. *Ugeskr Læger* 2011;173(44):2788.
- 13 Davies et al. Workplace-based assessment, sec 45, p 341-348. *Dent & Harden; A practical Guide for Medical Teachers*, 2009 (3rd edition), Churchill Livingstone.

14 Pangora L & Holmboe ES. "Evaluation Forms and Global Rating Scales", sec 3 p 24-41 in Holmboe & Hawkins; "A practical Guide to the Evaluation of Clinical Competence". 2008. Elsevier.

15 Reznick R , Regehr G, MacRae H, Martin J, McCulloch W. "Testing technical skill via an innovative "bench station" examination". *Amer J of Surg* 1997;173(3):226-230.

16 Dauphinee WD, Blackmore DE, Smee S, Rothman AI, Reznick R. Using judgements of physician examiners in setting the standards for a national multicenter high stakes OSCE. *Adv Health Sci Educ* 1997;2:201-211.

17 Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997 Feb;84(2):273-8.

18 Maagaard M, Oestergaard J, Johansen M, Andersen LL, Ringsted C, Ottesen B, et al. Vacuum extraction: development and test of a procedure-specific rating scale. *Acta Obstet Gynecol Scand* 2012 Aug 12.

19 Ma IW, Zalunardo N, Pachev G, Beran T, Brown M, Hatala R, et al. Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Adv Health Sci Educ Theory Pract* 2012 Oct;17(4):457-70.

20 Holmboe ES, Flebach NF, Galaty L, Huot S. The effectiveness of a focused educational intervention on resident evaluations from faculty: A randomized controlled trial. *J Gen Intern Med* 2001; 16(7):427-434.

21 Ahmed K, Miskovic D, Darzi A, Athansiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Amer J of Surg* 2011;202:469-480.

22 Bested KM, Malling B, Skjelsager K, Østergaard D, Østergaard HT, Ringsted C. Bedømmerbias i speciallægeuddannelsen. 2011 Ugeskrift for Læger.

23 Holmboe ES. "Direct Observation of Faculty", sec 9 p 119-129 in Holmboe & Hawkins; *A practical Guide to the Evaluation of Clinical Competence*. 2008. Elsevier.

24 Hodges et al. OSCE checklists do not capture increasing level of expertise. *Acad Medicine* 1999; 74(10), 1129-1134.

25 McKinley RK, Strand J, Ward L, Grav T, Alun-Jones T, Miller H. Checklists for assessment and certification of clinical procedural skills omit essential competencies: a systematic review. *Med Educ* 2008;42;338-349.

26 Kogan JR, Holmboe ES, Hauer KE. Tools for Direct Observation and Assessment of Clinical Skills of Medical Trainees. A systematic review. *JAMA* 2009;(302)30:1316-1326.

-
- 27 Gray JD, Ruedy J. Undergraduate and postgraduate medical education in Canada. *CMAJ*. 1998 Apr 21;158(8):1047-50.
- 28 Norman G. Editorial-Checklist vs ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. *Adv in Health Sci* 2005;10:1-3.
- 29 Regehr G, Mac R, Reznick RK, Szaley D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med* 1998;73(9):993-7.
- 30 Hodges B, McIlroy JH. Analytical global OSCE ratings are sensitive to level of training. *Med Educ* 2003;37:1012-16.
- 31 Van der Vleuten. The assessment of professional competence; developments, research and practical implications. *Adv Health Sci Educ Theory Pract* 1996;1:41-67.
- 32 Dansk Psykolog forening. Introduktion til psykometriske begreber – vejledning vedrørende krav til konstruktion af erhvervstest. 2004.
- 33 Holm AL. 2006. "Overvejelser om prøver" s 13-17. *Dansk Universitetspædagogisk Tidsskrift Netværk*.
- 34 van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg* 2010 Jul;97(7):972-87.
- 35 Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database Syst Rev* 2009;(1):CD006575.
- 36 Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *BMJ* 2003 Nov 1;327(7422):1032-7.
- 37 Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Assessment of technical surgical skills. *Eur J Surg* 2002;168(3):139-44.
- 38 Larsen CR, Oestergaard J, Ottesen BS, Soerensen JL. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. *Acta Obstet Gynecol Scand* 2012 Sep;91(9):1015-28.
- 39 Tolsgaard M, Todsén T, Sørensen JL, Ringsted C, Lorentzen T, Ottesen B, Tabor A. International Multispeciality Consensus on How to Evaluate Ultrasound Competence: A Delphi Consensus Survey. *PLOS* 2013; 8 (2): e587687.
- 40 Palter VN, Grantcharov TP. Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. *Ann Surg* 2012 Jul;256(1):25-32.
- 41 Sarker SK, Chang A, Vincent C, Darzi SA. Development of assessing generic and specific technical skills in laparoscopic surgery. *Am J Surg* 2006 Feb;191(2):238-44.

-
- 42 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 Jun;115(7):908-16.
- 43 Malling B. Managing work-based postgraduate medical education in clinical departments. Ph.d.-afhandling, Maastricht Universitet. 2011. Chapter 6, p 90.
- 44 Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS system: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality and Safety in Health Care* 2009;18:104-108.
- 45 Carthey J, de Leval MR, Wright DJ, Farewell VT, Reason JT and all UK paediatric cardiac centres. Behavioural markers of surgical excellence. *Safety Science* 2003;41:409-425.
- 46 Yule S, Flin R, Paterson-Brown S, Maran N. Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery* 2006;139:140-9.
- 47 Yule S, Flin R, Maran N, Rowley DR, Youngson GG, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavioural rating system. *World Journal of Surgery* 2008;32:548-556.
- 48 Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' non-technical skills (ANTS): evaluation of a behavioural marker system. *Br Journal of Anaesth* 2002;88:418-429.
- 49 Spanager L, Lyk-Jensen HT, Dieckmann P, Wettergran A, Rosenberg J, Østergaard D. Customization of a tool to assess Danish Surgeons' non-technical skills in the operating room. *Dan Med J* 2012;59(11):A4526.
- 50 Spanager L, Beier-Holgersen R, Dieckmann P, Konge L, Rosenberg J, Østergaard D. Reliable assessment of general surgeons' non technical skills based on video recordings of patient simulated scenarios. *Indsendt til Am J Surg* 2012. (Abstract).
- 51 Norgaard K, Ringsted C, Dolmans D. Validation of a checklist to assess ward round performance in internal medicine. *Med Educ* 2004; 38:700-7. *Sekundærpublikation Ugeskr Laeger* 2004;166:2027-31.
- 52 Ravn LI, Lund CM. A clinical skills training program--a structured, accelerated introduction. *Ugeskr Laeger*. 2004 May 17;166(21):2014-7.
- 53 Ringsted C, Østergaard D, Van der Vleuten CPM. Implementation of a formal in-training assessment programme in anaesthesiology and preliminary results of acceptability. *Acta Anaesthesiol Scand* 2003;47:1196-1203.
- 54 Ringsted C, Østergaard D, Ravn L, Pedersen JA, Berlac P, Van der Vleuten CPM. A feasibility study comparing checklists and global rating forms to assess resident performance in clinical skills. *Medical Teacher* 2003;25:654-8.

-
- 55 Norcini, J. J., Blank, L. L., Duffy, F. D., and Fortna, G. S.: The mini-CEX: a method for assessing clinical skills. *Ann Intern Med*, 138: 476, 2003.
- 56 Holmboe, E. S., Huot, S., Chung, J., Norcini, J., and Hawkins, R. E.: Construct validity of the miniclinical evaluation exercise (miniCEX). *Acad Med*, 78: 826, 2003.
- 57 Eriksen, J. G., Simonsen, D., Bastholt, L., Aspegren, K., Vinther, C., Kruse, K. et al.: (Mini clinical evaluation exercise as evaluation tool of communicative and cooperative skills in the outpatient clinic). *Ugeskr Laeger*, 171: 1003, 2009.
- 58 Dreyfus SE & Dreyfus HL. *Mind over machine*. New York, Free Press, Macmillian, 1986; pp 16-51.
- 59 Bloom BS. *Taxonomi of educational objectives. Handbook I, Cognitive domain*. New York, Longman, 1956.
- 60 Holmboe ES & Hawkins RE. Eds. *Practical guide to the Evaluation of clinical competence*. Mosby inc. 2008. ISBN 978-0-323-04709-8.
- 61 Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med* 1999;74 (11):1203–7.
- 62 DeWitt D, Carline J, Paauw D, Pangaru L. Pilot study of a “RIME”-based tool for giving feedback in a multispecialty longitudinal clerkship. *Med. Educ* .2008; 42(12): 1205-9.
- 63 Tolsgaard MG, Arendrup H, Lindhardt BO, Hillingsø JG & Ringsted C. Construct validity of the reporter-interpreter-manager-educator structure for assessing students’ patient skills. *Acad Med* 2012; 87(6) :799-806.
- 64 Skjelsager K, Malling B, Bsted KM, Østergaard HT, Ravn L, Østergaard D & Ringsted C. Implementering af nationalt kompetenceprogram i anæstesiologi. *Ugeskr Læger* 2008; 170(44): 3557-3561.
- 65 Whitehouse A, Hassell A, Wood L, Wall D, Walzman M, Campbell I. “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”. *Medical Teacher*. 2005;27(3):252-8.
- 66 Archer JC, Norcini J, Davies HA. Use of SPRAT for peer review of paediatricians in training. *BMJ*. 2005 May 28;330(7502):1251-4.
- 67 Allerup P, Aspegren K, Ejlersen E, Jørgensen G, Malchow-Møller A, Møller MK, Pedersen KK, Rasmussen OB, Rohold A, Sørensen B. Use of 360-degree assessment of residents in internal medicine in a Danish setting: a feasibility study. *Med Teach*. 2007 Mar;29(2-3):166-70.
- 68 Burford B, Illing J, Kergon C, Morrow G, Livingston M. User perceptions of multi-source feedback tools for junior doctors. *Med Educ*. 2010 Feb;44(2):165-76.

-
- 69 Overeem K, Lombarts MJMH, Arah OA, Klazinga NS, Grol RPTM, Woltersheim HC. Three methods of multi-source feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach*. 2010;32(2):141-7.
- 70 Archer JC, McAvoy P. Factors that might undermine the validity of patient and multi-source feedback. *Med Educ*. 2011 Sep;45(9):886-93.
- 71 Overeem K, Woltersheim H, Driessen E, Lombarts K, Van der Ven G, Grol R & Arah O. Doctor's perceptions of why 360-degree feedback does (not) work: a qualitative study. *Medical Education* 2009; 43(2): 874-882.
- 72 Campbell J, Hill J, Hobart J, Narayanan A, Norman G, Richards S, et al. GMC Multi-Source Feedback Study. Scientific report of the Main Survey (2008-10). Executive summary. http://www.gmc-uk.org/executive_summary_of_research.pdf 48212169.pdf Accessed on 13th of April 2012.
- 73 Wright C, Richards SH, Hill JJ, Roberts MJ, Norman GR, Greco M, et al. Multisource Feedback in Evaluating the Performance of Doctors: The Example of the UK General Medical Council Patient and Colleague Questionnaires. *Acad Med*. 2012 87 (12): 1668-78.
- 74 Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ*. 2010;341:c5064.
- 75 Guidelines til indførelse af 360 graders feedback i den lægelige videreuddannelse, Uddannelsesregion Nord. <http://www.auh.dk/fagfolk/forskning>.
- 76 Seeberg, J. (2006) 360° - evaluering i speciallægeuddannelse. Evalueringsrapport af KULpuljeprojektet: "360° evaluering i speciallægeuddannelsen". 2006. www.lvu.auh.dk.
- 77 Eriksen G & Malling B. 360 graders feedback af yngre læger; psykometriske egenskaber og udvikling af optimeret spørgeramme KUL-pulje projekt 2007-2008. Afrapportering. 2009. www.lvu.auh.dk.
- 78 Dent JA, Harden RM. *A practical guide for Medical Teachers*. Churchill Livingstone Elsevier 2009.
- 79 Petersen T. *Medical Audit Ugeskrift for Læger* 2000 3, 405.
- 80 Bennett IJ, Hayden J. Audit as part of summative assessment of vocational training. *Br J Gen Pract* 1995; 45(390):47-49.
- 81 Lough JRM, McKay J, Murray TS. *Br J Gen Pract* 1995; 607-9.
- 82 Schön D. *The reflective practitioner – How Professionals Think in Action*. New York: Basic Books 1983.
- 83 Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract*. 2009 Oct;14(4):595-621.

-
- 84 Schön D, 1983, *The reflective practitioner*. San Francisco: Jossey_Bass.
- 85 Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach*. 2009 Aug;31(8):685-95.
- 86 Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. *Med Teach*. 2009 Apr;31(4):299-318.
- 87 Alle specialer i UK: <http://www.jrcptb.org.uk/assessment/Pages/Workplace-Based-Assessment.aspx>.
- 88 Almen medicin, UK: <http://www.rcgp.org.uk/gp-training-and-exams/mrcgp-workplace-based-assessment-wpba/~media/Files/GP-training-and-exams/WPBA/Planning-and-conducting-the-CbD-interview.ashx>.
- 89 Munger BS et al. The certification examination in emergency medicine: An update. *Ann Emerg Med* 1982; 11: 91-96.
- 90 Jennett P, Affleck L. Chart audit and chart stimulated recall as methods of assessment in continuing professional health education. *J Cont Educ Health Prof* 1998; 18: 163-171.
- 91 Norcini J., Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31, *Med Teacher*, 2007; 29: 855-871.
- 92 Crossley J et al. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ*; 2011; 45: 560-9.
- 93 Gaba DM. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*. 13: 2-10.
- 94 Berkenstadt H, Ben-Menachem E, Dach R, Ezri T, Ziv A, Rubin O, Keidan I. Deficits in the provision of cardiopulmonary resuscitation during simulated obstetric crises: results from the Israeli Board of Anesthesiologists. *Anesth Analg*. 2012 Nov;115(5):1122-6. doi: 10.1213/ANE.0b013e3182691977. Epub 2012 Aug 2.
- 95 Ben-Menachem E, Ezri T, Ziv A, Sidi A, Brill S, Berkenstadt H. Objective Structured Clinical Examination-based assessment of regional anesthesia skills: the Israeli National Board Examination in Anesthesiology experience. *Anesth Analg*. 2011 Jan;112(1):242-5. doi: 10.1213/ANE.0b013e3181fc3e42. Epub 2010 Oct 21.
- 96 Ben-Menachem E, Ezri T, Ziv A, Sidi A, Brill S, Berkenstadt H. Objective Structured Clinical Examination-based assessment of regional anesthesia skills: the Israeli National Board Examination in Anesthesiology experience. *Anesth Analg*. 2011 Jan;112(1):242-5. doi: 10.1213/ANE.0b013e3181fc3e42. Epub 2010 Oct 21.