Eksamen: MEDSEM_H12_ORD

Samlet vurdering:

Samlet vurdering af eksamenssættet er, at det er et acceptabelt sæt, men at der er rum for forbedring.

Styrkerne i sættet er, at der er et godt antal opgaver, at den emnesmæssige fordelingen mellem hovedområderne er god, at der er en acceptabel repræsentation af læringsmål, som udprøves og at det taxonomiske niveau er relativt lavt men i overensstemmelse med læringsmålene. Den kontekstuelle ophæng af spørgsmålene i de kliniske vignetter gør spørgsmålene relevante og interessante. Endelig er det godt med det supplerende materiale i form af fx EKG, billeder, laboratoriedata.

Svagheden i sættet er først og fremmest formuleringen af spørgsmålene og at flere af svarmulighederne i MCQ spørgsmålene vil belønne den testkloge studerende, som ikke også nødvendigvis er faglig dygtig. Desuden er det et generelt problem, at der i et ikke ubetydeligt antal spørgsmål indgår flere rigtige svar og ofte uden at det er specificeret, hvor mange rigtige svar, der er. Teknisk er der i sådant tilfælde tale om, at flere spørgsmål udmøntet i et.

Der skønnes at være et generelt behov for at træne underviserne i at udarbejde bedre spørgsmål. Noget kan gøres ved at konsultere for eksempel NBMEs item writing manual: http://www.nbme.org/publications/item-writing-manual.htm. Endnu bedre ville det være at afholde workshops og udvikle en peer-review kvalitetssikringsmekanisme af eksamensspørgsmålene.

Endelig er der truffet nogle valg, som det kunne være godt at diskutere igen:

- Skal så mage læringsmål være på så lavt taxonomisk niveau?
- Der testes generelt klinisk kundskaber, dvs. spørgsmål, der efterspørger klinisk diagnostik og behandling og i meget mindre grad grundlæggende basalbiologisk viden, herunder patofysiologisk indsigt. Man har altså en gruppe læringsmål, som generelt ikke udprøves.
- Når man folder ud bare én klinisk historie, begrænser man sig til at teste ét sygdomsområde. Generelt vil man anbefale at dække flere kliniske områder i en eksamen af denne type.
- En sygehistorie, der foldes ud som denne, giver den studerende svar på tidligere spørgsmål, efterhånden som der tilføjes informationer. Har man forhindret den studerende i at gå tilbage til tidligere svar?
- Ved at vælge forskellige typer opgaver, herunder åbne spørgsmål, reduceres mulighederne for digitalisering og signalet om forventningerne til de studerende kan blive utydeligt

I det efterfølgende foretages en nøjere analyse og diskussion af eksamenssættet

Spørgsmål 1:

- I hvilket omfang afspejler eksamensspørgsmålene læringsmålene (indholdsvaliditet)?
- I hvilket omfang testes der bredt i emneområderne (bredde er vigtigt, fordi der eksisterer stor casespecificitet mht. emneskundskaber)

Metode:

- Den relative fordelingen mellem de emnemæssige hovedområder (hjerte-lunge-nyre-traumatologi-urologi) undersøges i forhold til fordelingen i målbeskrivelsen for semesteret. Det forudsættes, at hvert underspørgsmål vægter lige både i den tid, det tager at svare på spørgsmålet og i den relative vægtning spørgsmålet gives i points.
- Indenfor hvert hovedområde afkrydses læringsmålene i målbeskrivelsen efter som de forekommer i eksamenssættet. Efterfølgende vurderes det, i hvilket omfang der er en spredning (tilfældighed). Det antages at alle læringsmål er lige vigtige

Analyse 1) relative fordeling mellem hovedområder:

- **Hjerte:** 4 delområder med henholdsvis 4, 7, 6 og 3 underspørgsmål: i alt 20 spørgsmål
- **Lunge:** 5 delområder med henholdsvis 3, 4, 4, 3 og 1 underspørgsmål: i alt 15 spørgsmål
- **Nyre:** 7 delområder med henholdsvis 1, 3, 4, 5,4,2 og 1 underspørgsmål: i alt 19 spørgsmål
- **Traumatologi:** 5 delområder med henholdsvis 4, 5, 2, 5 og 1 underspørgsmål: i alt 17 spørgsmål
- **Urologi**: 4 delområder med henholdsvis 6, 5, 6 og 3 underspørgsmål: i alt 20 spørgsmål

I henhold til semesterboken er den relative fordeling af undervisningen i semesteret for hjerte/cirkulation; lunge/respiration og nyre/urologi henholdsvis 6, 4 og 5 uger (eller (40 %, 27% og 33 %) . I eksamenssættet er fordelingen af spørgsmålene for hjerte/cirkulation; lunge/respiration, nyre/urologi og henholdsvis 22 %; 16%; 42% og 19%. Umiddelbart virker nyre/urologi overrepræsenteret i forhold til den tid, de har i undervisningen (eller undervisningen er relativt underrepræsenteret). Det er dog uklart, hvordan traumaotologien er tænkt ind i eksamenssættet.

Analyse 2) Spørgsmålenes afspejling af læringsmålene og det taxonomiske niveau

- **Hjerte:** Der stilles spørgsmål svarende til ca. 1/3 af læringsmålene. Hovedparten af spørgsmålene befinder sig lige som læringsmålene på relativt lavt taxonomisk niveau (beskrive og forklare).
- **Lunge:** Der stilles spørgsmål svarende til ca. 1/2 af læringsmålene. Spørgsmålene befinder sig både på lavere og lidt højere taxonomisk niveau.
- **Nyre:** Der stilles spørgsmål svarende til ca. 3/4 af læringsmålene. Spørgsmålene befinder sig både på lave og højere taxonomisk niveau.
- **Traumatologi:** Der stilles spørgsmål på både lave og højere taxonomisk niveau
- **Urologi**: Der stilles spørgsmål svarende til ca. 1/3 af læringsmålene. Hovedparten af spørgsmålene befinder sig lige som læringsmålene på relativt lavt taxonomisk niveau (beskrive og forklare).

Diskussion og konklusion:
Eksamenssættet består af i alt 25 dele med til sammen 91 spørgsmål, der er nogenlunde ligeligt fordelt på de 5 hovedområder. Det vurderes, at eksamenssættet har en fornuftig længde og en rimelig relativ fordeling mellem de emnemæssige delområder, som sættet skal teste. Det noteres specielt, at der, hvor der kun er et enkelt spørgsmål i et delområde (nyre og traumatologi), er der flere delområder (henholdsvis 7 og 5). Eksamenssættet skønnes at teste bredt emnemæssigt.
Det vurderes ikke at have været et egentligt problem, at der er forskellige vægtninger i henholdsvis undervisning og eksamen, men studieledelsen kunne overveje:

- Om vægtningen af undervisningen er hensigtsmæssig i den nuværende form og dermed afspejler den emnemæssige vægtning, som ønskes
- Hvis undervisningens relative fordeling skønnes hensigtsmæssig, om der da er andre faktorer, der nødvendiggør, at nyre og urologi gives relativ større vægtning i eksamenssættet. Hvis dette er tilfældet bør det være klart kommunikeret til de studerende
- Om vægtningen mellem de emnemæssige hovedområder i eksamenssættet fremover skal ændres

Spørgsmål 2:
Hvordan er kvaliteten af spørgsmålene

Metode:
Der er foretaget en global vurdering af hver hovedområdedel på en 5-punkt Likert skala (Udmærket, over middel, middel, under middel, uacceptabel). Vurderingen har taget udgangspunkt i blandt andet flg. underspørgsmål:

- I hvilket omfang er spørgsmålene kontekstualiseret
- Stilles der spørgsmål på varierende taxonomisk niveau?
- Er spørgsmålene klart og entydigt formuleret?
- Hvilken kvalitet har svarmulighederne
- Er der noget <u>ikke-</u>væsentligt (fx længde, bestemte ord), der leder mod svaret
- Er de anvendte svarmuligheder fornuftige i antal
- Er der noget <u>ikke</u>-væsentligt (fx længde, bestemte ord), der viser at det ikke er et korrekt svar
- I hvilket omfang er der brugt supplerende materiale

Analyse af spørgsmålenes kvalitet:
Hjertedel: Alle spørgsmål er hængt op på én klinisk historie og således at den kliniske vignet udfoldes gradvist, startende med diagnostik, dernæst behandling og sidst rekonvalescens. Kvaliteten af spørgsmålene er middels. Der er en række "flaws", som der let kunne gøres noget ved, hvorved kvaliteten kunne øges betydeligt:

- Der bruges mange negationer ("ikke typisk", "ikke rutinemæssigt", fejl udsagn"… osv. Det er vanskeligt for den studerende og unødvendigt (tester ikke relevant viden)
- Flere gange gives det rigtige svar i et senere spørgsmål: Fx viser 2.1. at svar på 1.1. er ekkokardiografi og oplysningen i delvignet 3, at svaret på 2.4 er biologisk AVR

- Enkelte svarmuligheder er dårlige og belønner "test-visdom" (fx begge svarmuligheder i 2.2 og svarmulighederne 4 og 5 i 2.7)
- Enkelte spørgsmål virker som overflødig udenadslære, fx 4.2

Lungedel: Alle spørgsmål er hængt op på én klinisk historie og således at den kliniske vignet udfoldes gradvist. Spørgsmålene er af tre typer – kortsvars med fri formulering, MCQ og sandt/falsk. Der efterspørges generelt et højere taxonomiske niveau end i forrige hovedområde (hjertedelen), hvilket anses for at være positivt. Kvaliteten af spørgsmålene vurderes at være under middel, fordi der er en del problemer i såvel spørgsmål som forslag til pointgivning:

- Flere gange gives det rigtige svar i et senere spørgsmål. For eksempel får studenten alle svarene på del 1 i indledningen til del 2
- Enkelte svarmuligheder er dårlige og belønner "test-visdom" (fx kunne svarmulighed 3 + 4 udelades i 4.2, da de alene fra en overfladisk betragtning er anderledes end de øvrige svarmuligheder)
- Nogle spørgsmål burde have testet forståelse (man kan have ret svar men fejl begrundelse), fx spørgsmål 3.2
- Problematisk forslag til vægtning af ret svar (fx i hele del 1, hvor a + b ikke er det samme som (a) + (b))
- Formuleringen af visse spørgsmål burde være mere præcis, fx 3.3, hvor det er indforstået, at der bliver brugt kontrast
- Kan spørgsmål 3.1 have to rette svar? (er ikke astma en obstruktiv lungesygdom?)
- To rette svar i 4.3 burde af hensyn til retfærdig pointgivning burde have givet anledning til at der var blevet stillet to spørgsmål i stedet
- Nogle spørgsmål (og forslaget til pointgivning) giver ikke klare signaler om, hvad der forventes af den studerende (fx 2.2, hvor man synes at ville belønne den studerende for noget ekstra, der ikke er direkte spurgt om)

Nyredel: Alle spørgsmål er hængt op på én klinisk historie og således at den kliniske vignet udfoldes gradvist. Spørgsmålene er af tre typer – kortsvars med fri formulering, MCQ og sandt/falsk. Ligesom i lungedelen efterspørges der generelt et højere taxonomiske niveau end i hjertedelen, hvilket anses for at være positivt (studerende bør ikke bare kunne beskrive men også fx kunne ræssonere). Kvaliteten af spørgsmålene vurderes at være over middel. Der findes flere rigtig fine spørgsmål, fx 6.2 ; 7.1. og 4.3 4.3 (sidste svarmulighed dog dårlig) De er dog også flere problematiske forhold, som giver anledning til "flaws" og som relativt let kan forbedres og dermed forhøje kvaliteten af opgaverne:

- Flere spørgsmål har mere end et rigtigt svar. Det gør det svært for de studerende at vide, hvad der forventes (yderligere problematisk, fordi antallet af rette svar ikke opgives). Det giver desuden problemer med ordentlig pointgivning.
- Der bruges ord som "altid" (fx 8.2). Studerende ved godt, at der stort set ikke findes noget, som "altid" hænder og dermed at svaret er fejl (hermed belønnes en studerende for et svar uden nødvendigvis at kende ret sammenhæng)
- Det bør overvejes om nogle af spørgsmålene er formuleret, så de bliver for lette/trivielle? (fx 3.1 og 3.3)
- Nogle spørgsmål burde formuleres bedre rent tekstmæssigt. Principielt er ret svar på fx 3.2, hvor der spørges om "Ved du hva", enten et "ja" eller et nej" og ikke et åbent svar.
- Flere gange givet det rigtige svar i et senere spørgsmål: Fx giver teskten i del 2 svaret på 1.

Traumatologi: Alle spørgsmål er hængt op på én klinisk historie og således at den kliniske vignet udfoldes gradvist. Halvdelen af spørgsmålene er af MCQ-type, den anden halvdel er kortsvarsspørgsmål. Endelig er der et spørgsmål (5.1, som teknisk er 4 spørgsmål af sandt/falsk format). Kvaliteten af spørgsmålene vurderes at være over middel med nogle rigtig gode spørgsmål, fx 4.5, andre som har en god bagvedliggende tankegang, fx 5.1 og hvor der testes bred taxonomisk. Også på dette hovedområde er der svagheder, som relativt let kunne rettes op og dermed være med til at øge denne eksamens kvalitet:

- Kortsvarsopgaverne forudsætter at studenten skal genkalde sig viden. Der er mange af dem i dette sæt, fx det meste af del 2. Hvis man vil teste åbent, bør det overvejes, at lave et enkelt spørgsmål og bede studenterne om at levere mere sammenhængende forklaringer. Den del af patienthistorien, om går på costafrakturer/pneumo-hæmothorax kunne herved have afsløret også , om den studerende havde en dybtgående forståelse af emnet.
- Tilsvarende får man for lidt udbytte af de gode informationer i del 2 – her kunne man have testet bedre for forståelsen af patofysiologien, hvis der blev efterspurgt begrundelser/ræssonering
- Også de gode informationer i del 3 bruges kun til at efterspørge mere triviel viden
- Der er flere spørgsmål, hvor der er flere rigtige svar (for begrundelse for, hvorfor det er uheldigt, henvises til de andre hovedområder).
- I flere spørgsmål betyder et enkelt ord meget for hvad der er ret svar (fx "bakgrund" i 1.1, "påfølgende" i 1.4 og "første" i 2.3). Det kan diskuteres, om det er hensigtsmæssigt eller om man ved reformulering kunne være klarere.

Urologi: Alle spørgsmål er hængt op på én klinisk historie og således at den kliniske vignet udfoldes gradvist. Spørgsmålene er dels MCQ og dels kortsvars. Kvaliteten af spørgsmålene vurderes samlet at være under middel. Trods enkelte gode spørgsmål, fx 4.2 er der overvejende tale om test af paratviden (og ikke af fx evnen til at ræssonere eller have kendskab til patofysiologi). Samtidigt er der adskillige spørgsmål med indlagte "flaws" (som hvis de rettes op, kan give meget bedre kvalitet):

- Flere spørgsmål har mere end èt rigtigt svar. Det gør det svært for de studerende at vide, hvad der forventes (yderligere problematisk, fordi antallet af rette svar ikke altid opgives). Det giver desuden problemer med ordentlig pointgivning.
- Flere spørgsmål efterspørger relevant klinisk viden men der savnes spørgsmål, som også efterspørger begrundelser (klinisk ræssonering), fx 1.3; 1.4;
- Der bruges ord som "altid" (1.6) – det belønner den testkloge studerende, som ikke nødvendigvis har god faglig indsigt
- I hvert fald et spørgsmål bruger negation ("ikke", 3.3), hvilket gør spørgsmålet unødvendigt svært at læse (unødvendigt, fordi vanskeligheden ikke er relateret til faglig viden)
- Nogle spørgsmål ville blive bedre, hvis oplagte fejl svarmuligheder blev udeladt (fx svarmulighed 1, 2 og 5 i 3.6)

Konklusion:

Samlet er spørgsmålskvaliteten acceptabel, men der er variationer i kvaliteten af spørgsmålene mellem de fem hovedområder. På samtlige hovedområder er der basale svagheder i mange spørgsmål, i mange svarmuligheder og derfor er der rum for betydelig forbedring.