

# ***Eksamens– og Evalueringsformer***

## **Medisinstudiet ved NTNU**

Innstilling fra Arbeidsgruppen for Eksamens- og Evalueringsformer  
April 2003

# Innhold

<b>Innledning</b> .....	<b>4</b>
<b>Sentrale begreper</b> .....	<b>4</b>
<b>Akronymer</b> .....	<b>5</b>
<b>Forslag til ny eksamens- og evalueringsordning</b> .....	<b>6</b>
<b>1 Dagens eksamensordning – en kortfattet oversikt</b> .....	<b>7</b>
A) Muntlige og skriftlige eksamener .....	7
B) Gjeldende kontinuasjonordninger .....	8
<b>2 Beskrivelse og vurdering av dagens eksamensformer</b> .....	<b>8</b>
A) Skriftlige integrerte eksamener.....	8
B) Bruk av PBL ved skriftlig eksamen Stadium I .....	9
C) Tre-stasjons muntlig/praktisk eksamen Stadium I A/B .....	9
D) ”Double Jump” muntlig/praktisk eksamen Stadium I C/D.....	9
E) Tre-stasjons muntlig/praktisk eksamen Stadium II .....	10
F) Tidligere forsøk på OSCE Stadium II .....	11
G) Konsultasjonseksamen Stadium IIC.....	11
H) Muntlig/Kliniske eksamener Stadium IID .....	12
<b>3 Alternative vurderingsformer</b> .....	<b>12</b>
A) MEQ - Modified Essay Questions.....	13
B) Flervalgstester / MCQ - Multiple Choice Questions .....	13
C) OSCE - Objective Structured Clinical Examination.....	15
D) Klinisk Avslutningseksamen IID .....	16
E) Mappeevaluering .....	17
<b>4 Forslag til endringer i eksamensordning</b> .....	<b>20</b>
A) Skriftlige integrerte flervalgstester .....	21
B) Bruk av PBL ved skriftlig eksamen Stadium I .....	21
C) OSCE Stadium I .....	21
D) Kommunikasjonseksamen I C.....	22
E) OSCE / Muntlig/Praktisk eksamen Stadium II.....	22
F) Konsultasjonsevaluering IIC.....	23
G) Klinisk Avslutningseksamen IID .....	23
H) Mappeevaluering .....	24
<b>5 Forslag til endring av eksamensavvikling og kontinuasjonordninger</b> .....	<b>24</b>
A) Integrering av forskerlinja .....	24
B) Endringsforslag til avvikling av konsultasjonseksamen .....	24
<b>6 Utplassering</b> .....	<b>24</b>
A) Lokalsykehus.....	25
B) Allmenpraksis.....	25
<b>7 Personlig oppfølging og veiledning</b> .....	<b>26</b>
A) PBL-veilederene.....	26
B) Mentor-ordning .....	27
<b>8 Karakterer</b> .....	<b>28</b>
A) Bakgrunn .....	28
B) Påstander og argumenter .....	29
C) Muligheter og krav ved innføring av karakterer .....	30
D) Ulike standpunkter .....	30
<b>Kilder</b> .....	<b>31</b>
<b>Vedlegg</b> .....	<b>31</b>

1 )

1 )

## Innledning

Arbeidsgruppen for Eksamens- og Evalueringsformer er oppnevnt av, har sitt mandat fra og leverer sin innstilling til prodekanus for undervisning Torsten Vik. Arbeidet ble påbegynt i november 2002, og avsluttet i april 2003. Arbeidsgruppens mandat har vært:

- Gruppen skal beskrive og vurdere, og eventuelt foreslå nødvendige endringer av de eksamensformene som per i dag benyttes i medisinstudiet. Der hvor det foreslås å erstatte eksamensformer, må det foreslås hvordan disse skal erstattes.
- I tråd med kvalitetsreformen i høyere utdanning, og føringene fra NTNU sentralt, gi råd om hvordan mappe-evaluering kan innføres, og hvordan studentene kan få individuell, jevnlig tilbakemelding på en mer systematisert måte enn i dag.
- Foreslå hvordan studentene kan evalueres under og /eller etter utplasseringen i sykehus.
- Spesielt vurdere om PBL-veilederne kan gi mer jevnlig tilbakemelding til den enkelte student.
- Spesielt er det ønskelig at man redegjør for hvilke overordnede læringsmål (teori, praktiske ferdigheter, kommunikasjon, holdninger) man ønsker å evaluere med de forskjellige evalueringsformene.
- I tråd med kvalitetsreformen gi råd om innføring av karakterer.
- Foreslå hvordan eksamensavviklingen kan tilpasses forskerlinjen

Grappa har vært sammensatt av følgende medlemmer, som har vært representanter for:

- Administrasjonen: May Britt Schjølberg
- Terminledere: Magne Børset (IA), Geir Bråthen (IIA), Kåre Bergh (IIID)
- Allmennpraktiker: Niels Bentzen
- Studenter: Erlend Prytz og Thomas Husby

Erlend Prytz har fungert som Arbeidsgruppens leder.

Grappa har hatt jevnlige, oftest ukentlige møter i perioden november 2002 til april 2003. Gruppemedlemmene har studert bakgrunns litteratur, invitert eksterne foredragsholdere, hentet inn erfaringer fra andre universiteter og deltatt på kurs/møter sammen med andre fagmiljøer ved universitetet. Spesielt nevnes Birger Bertheussen ved Pedagogisk Institutt, og Læringsarenaer NTNU, der gruppen har vært med i Arena 2: Læringsmål og Vurderingsformer.

## Sentrale begreper

### Validitet

I hvilken grad en måler det som er viktig/sentralt. Dette begrepet kan deles inn i to varianter:

- Ytre validitet: Å teste det som er viktig for yrket. Den vanligste betydningen av validitet.
- Indre validitet: Å teste det som er viktig/sentralt i pensum. Å teste det som blir undervist.

### Reliabilitet

Støhet/pålitelighet av resultatet. Avhenger av flere faktorer:

- *Oppgavevariasjon*: Vil studenten få samme resultat hvis han får ny eksamen?
- *Ulik sensor-variasjon*: Vil ulike sensorer gi en besvarelse samme resultat?
- *Samme sensor-variasjon*: Vil sensor gi samme resultat dersom han vurderer samme besvarelse flere ganger?

### Formativ evaluering

Skal styre studentens læring og videre arbeide.

### Summativ evaluering

Skal vurdere studentens nivå, innsats eller prestasjon.

Det kan umiddelbart være vanskelig å se den prisippielle forskjellen mellom formativ og summativ evaluering. Forskjellen er at en formativ evaluering ikke får konsekvenser for studenten ut over studiesituasjonen, mens en summativ evaluering får konsekvenser i form av karakter eller trussel om kontinuasjon. I en formativ evaluering ønsker studenten å gi et fullstendig bilde av sitt nivå, med flest mulig feil og mangler, for å kunne utbedre disse. I en summativ vurderingssituasjon vil studenten

framstille seg og opptre på en annen måte. Han vil her framheve sine sterke sider, og så godt som mulig dekke over sine mangler og svakheter. Dette er også studentens soleklare rett. En kan derfor ikke blande sammen formative og summative evalueringer. (Lauvås/Jakobsen).

### **Overordnede Læringsmål**

Dette begrepet ble brukt i arbeidsgruppens mandat, og er derfor også benyttet i instillingen. I mandatet ble begrepet eksemplifisert med teori, praktiske ferdigheter, kommunikasjon og holdninger. Arbeidsgruppen har valgt å tolke dette til å omfatte alle generelle egenskaper en medisinstudent skal utvikle i løpet av studiet.

## **Akronymer**

### **MEQ**

Modified Essay Questionnaire. En skriftlig eksamensform med korte svar, der spørsmålene skal knyttes til en kasuistikk. Eksamensformen drøftes i kapittel 3 A).

### **MCQ**

Multiple Choice Questions, også kalt flervalgsspørsmål. En skriftlig eksamensform der en krysser av for de på forhånd oppgitte alternativer som oppfyller kriteriene gitt i spørsmålet. Eksamensformen drøftes i kapittel 3 B).

### **OSCE**

Objective Structured Clinical Examination. En muntlig/praktisk eksamensform der en skal løse konkrete problemstillinger i løpet av et avgrenset tidsrom, i form av ca. 20 poster på ca. 5-10 minutter hver. Eksamensformen drøftes i kapittel 3 D).

## Forslag til ny eksamens- og evalueringsordning

Stadium Eksamenstidspunkt	Eksamens/evalueringsform
<b>Stadium I A/B</b> Mai/Juni	Mappeevaluering
	Første integrerte flervalgstest
<b>Stadium IC</b> Desember	Kommunikasjonseksamen
<b>Stadium ID</b> Mai/Juni	Mappeevaluering
	OSCE Stadium I
	Andre integrerte flervalgstest
<b>Stadium II A/B</b> Mai/Juni (Forskerlinje: Januar)	Mappeevaluering
	OSCE / Tre-stasjons muntlig/praktisk eksamen
	Tredje integrerte flervalgstest
<b>Stadium II C/D</b> Mai/Juni (Forskerlinje: Januar)	Mappeevaluering
	OSCE / Tre-stasjons muntlig/praktisk eksamen
	Fjerde integrerte flervalgstest
<b>Stadium IIIA</b>	Hovedoppgave
<b>Stadium IIIB</b>	Mappeevaluering
<b>Stadium IIIC</b> Desember	Mappeevaluering
	Konsultasjonsevaluering
	Femte integrerte flervalgstest
<b>Stadium IIID</b> Mai/Juni	Klinisk Avslutningseksamen
	Sjette integrerte flervalgstest

Arbeidsutvalgets forslag må betraktes som en helhet. Det er forsøkt å lage et opplegg der hver enkelt del har sin avgrensede funksjon Samlet skal de gi en bred, god vurdering av alle egenskaper en medisinstudent skal tilegne seg i løpet av studiet.

I hovedsak er funksjonen tiltenkt de ulike eksamens- og evalueringsformene som følger:

### Integrerte flervalgstester:

- Kunnskapsnivå
- Resonnement

### OSCE / Tre-stasjons Muntlig/praktisk eksamen:

- Kliniske og parakliniske praktiske ferdigheter

### Kommunikasjonseksamen IC og Konsultasjonsevaluering IIIC

- Lege-pasient-kommunikasjon
- Konsultasjonsprosess

### Klinisk Avslutningseksamen IIID

- Alle kunnskaper og ferdigheter en nyutdannet lege skal ha
- Evne til å håndtere de ulike situasjoner en turnuslege vil møte
- Sette pasienten og konsultasjonen i en sammenheng, ut fra rammene i det norske helsevesen

### Mappeevaluering

Det kreves videre planlegging før mappeevalueringens rolle i studiet er avklart, spesielt når det gjelder balanse mellom formativ og summativ evaluering. Mappeevaluering vil være spesielt nyttig i forbindelse med utplasseringsperiodene. I andre deler av studiet kan mappeevaluering best utnyttes ved å evaluere de overordnede læringsmål som ikke lar seg teste på en god måte ved en eksamen. Dette gjelder egenskaper som skriftlig framstilling, bruk av fagterminologi, klinisk og basalfaglig resonnement, dypere forståelse/analytiske evner, selvrefleksjon/evaluering, holdninger, læringsstrategier, informasjonssøk, kritisk kildebehandling, litteraturforståelse mm.

## 1 Dagens eksamensordning – en kortfattet oversikt

### A) Muntlige og skriftlige eksamener

Alle ordinære eksamener avholdes i mai/juni. Unntak er eksamener for IIIC som avholdes i desember. Den innbyrdes rekkefølgen mellom skriftlig og muntlig/praktisk eksamen fastsettes for hver enkelt eksamen.

#### Termin IA/B

1. Integreert skriftlig eksamen 6 timer- PBL-møte i forkant av eksamen, inntil 25% av spørsmålene til eksamen skal være fra PBL-oppgaven
2. Muntlig/praktisk eksamen - Stasjonseksamen: 3 stasjoner á 15 min. med sensor og eksaminator på hver stasjon.
  - Praktiske ferdigheter er et element som ønskes testet.

#### Termin IC/D

1. Integreert skriftlig eksamen 6 timer - PBL-møte i forkant av eksamen, inntil 25% av spørsmålene til eksamen skal være fra PBL-oppgaven
2. Muntlig/praktisk eksamen – Double jump (Kommunikasjonseksamen): kjøres parallelt i 5 rom, sensor og eksaminator i hvert rom, ca. 40 min. per kandidat.
  - Eksamen består av to elementer:
    - 1) Studentene eksaminerer pasienter (skuespillere) omkring et problem og gjennomfører en enkel klinisk undersøkelse. Kommunikasjonsferdigheter og pasientkontakt og ferdigheter evalueres.
    - 2) Studentene formulerer læringsmål knyttet til konsultasjonen og angir forslag til hvordan en vil gå frem for å oppfylle læringsmål.

#### Termin IIA/B

1. Integreert skriftlig eksamen (hvorav 1/6 MEQ) 6 timer
2. Muntlig/praktisk eksamen – Stasjonseksamen: 3 stasjoner á 15 min. med sensor og eksaminator på hver stasjon.
  - Oppgaver av praktisk art, gjerne med pasient

#### Termin IIC/D

1. Integreert skriftlig eksamen (hvorav 1/6 MEQ) 6 timer
2. Muntlig/praktisk eksamen - Stasjonseksamen: 3 stasjoner á 15 min. med sensor og eksaminator på hver stasjon.
  - Oppgaver av praktisk art, gjerne med pasient

#### Termin IIIA/B

- Hovedoppgavetermin – må være bestått før studenten kan gå videre i studiet.
- Utplassering i lokalsykehus – ferdighetslister må være godkjent før studenten kan gå videre i studiet.

#### Termin IIIC (desember)

1. Skriftlig eksamen 6 timer
2. Muntlig/praktisk eksamen – Konsultasjonseksamen - kjøres parallelt i 5 rom, sensor og eksaminator i hvert rom, ca. 40 min. per kandidat.
  - Ressurskrevende å skaffe pasienter, derfor bringer studenten med seg en video han/hun har tatt opp av en konsultasjon med en pasient. Videoen danner basis for eksaminasjonen. Man vurderer gjennomføring av konsultasjonen, kommunikasjonsferdigheter, viten om konsultasjonsprosessen, og tilstrekkelige medisinske og basalfaglige kunnskaper til å kunne gjennomføre en ryddig konsultasjon.

#### Termin IIID

1. Skriftlig eksamen 6 timer. Vitenskapelig artikkel deles ut 72 timer før eksamen og spørsmål knyttet til denne utgjør 15% av skriftlig eksamen.
2. Muntlig/kliniske eksamener. Det eksamineres i indremedisin eller kirurgi samt et "lite" fag.

- Anamnese, full klinisk undersøkelse med eksaminasjon i relevante problemstillinger knyttet til aktuelle kasus. Basis i samlede læringsmål for hele studiet, bred vinkling.

### **B) Gjeldende kontinuasjonensordninger**

Tidspunkt for kontinuasjonseksamen og utsatt eksamen i henhold til studiehandboka for medisin: "Kontinuasjonseksamen etter ordinær eksamen for vårsemesteret holdes i uke 32. Etter ordinær eksamen for høstsemesteret holdes kontinuasjonseksamen i siste uke i februar. Utsatt eksamen holdes etter individuell avtale med de studentene det gjelder. Vanligvis vil man forsøke å få avvirket den utsatte prøven så snart det er praktisk mulig".

Tidspunkt for kontinuasjonseksamen for avgangskullet (termin IIID) vedtatt i F-sak 26-02:

"1. Dersom en student stryker i et muntlig fag vil kontinuasjonseksamen holdes innen tre uker. Det foretas ny trekning av fag og studenten får vite hvilket fag han/hun skal opp i en uke før eksamen, på samme måte som til ordinær eksamen.

2. Stryker studenten i både lite fag og stort fag skal det kontinueres i begge fag etter ny trekning. Studenten går opp etter tre uker med en uke mellom fagene, og får vite faget en uke før eksamen.

3. Ved stryk til skriftlig eksamen eventuelt stryk både til skriftlig og muntlig eksamen arrangeres kontinuasjonseksamen i oktober (som tidligere)."

## **2 Beskrivelse og vurdering av dagens eksamensformer**

Arbeidsutvalget vil i dette kapittelet gi en nærmere beskrivelse og vurdering av de ulike eksamnene som er i bruk per i dag. Det er forsøkt å belyse ressursbruk, validitet, reliabilitet og hvilke temaer og overordnede læringsmål som testes ved de ulike eksamener.

### **A) Skriftlige integrerte eksamener**

#### **Beskrivelse:**

Det er ingen krav til hvilken eksaminasjonsform som skal brukes ved dagens skriftlige eksamener. Unntak: For stadium II skal MEQ utgjøre 1/6 av eksamen både med hensyn på tidsbruk ved eksamen og poeng. Resultatet er at eksamnene blir tildels ulike i utforming, både mellom terminene og fra år til år. Det er ofte innslag av tildels "improviserte" flervalgsspørsmål og en tilsynelatende vilkårlig sammensetning av kort- og langvarsoppgaver. Det er generelle føringer for hvilke overordnede læringsmål som skal testes ved skriftlig eksamen. Erfaringsmessig brukes mye tid og ressurser både før og etter eksamen. Eksaminasjonstid er seks timer.

#### **Fordeler:**

- Stor frihet til eksaminator gir fleksibilitet
- Mulighet for å teste mange overordnede læringsmål

#### **Ulemper:**

- Lite forutsigbarhet for studentene
- Begrenset omfang av testing gir lav indre validitet
- Usikker reliabilitet
  - skriveferdigheter innvirker på prestasjon
  - sensur basert på skjønn
  - tester begrenset del av læringsmål for terminen
- Ressurskrevende både før og etter eksamen

#### **Konklusjon:**



Dagens tradisjonelle skriftlige eksamen er en uklart definert og ressurskrevende eksamensform med usikker reliabilitet.

### **B) *Bruk av PBL ved skriftlig eksamen Stadium I***

#### **Beskrivelse:**

Fem dager i forkant av skriftlig eksamen Stadium I går hver PBL-gruppe gjennom en "eksamens-PBL-oppgave". Inntil 25% av spørsmål og poeng på eksamen kan være i relasjon til PBL-oppgaven. Denne løsningen ble utarbeidet etter fakultetsstyrets vedtak om at PBL-elementet i studiet skulle gjenspeiles ved eksamen. Gruppene som var "dyktige" i PBL-metoden skulle få belønning for dette på eksamen.

#### **Vurdering:**

Erfaringen har vist at PBL-oppgaven ikke har hatt særlig påvirkning på utforming av eksamen, og at den for studentene bare fungerte som "eksamenstips" og en ekstra stressor som styrte lesingen de siste dagene i forkant av eksamen.

#### **Konklusjon:**

PBL er en læringsmetode. Det tjener ingen formål å blande læringsmetode med eksamensmetode. Integreerte eksamener er ikke ment å teste de kunnskaper studentene klarer å tilegne seg de siste hektiske dagene før eksamen.

### **C) *Tre-stasjons muntlig/praktisk eksamen Stadium I A/B***

#### **Beskrivelse:**

Eksamen består av tre stasjoner på 10-15 min. med muntlig eksaminasjon og et innslag av praktisk oppgave, i form av demonstrasjon/bruk av utstyr, begrenset klinisk undersøkelse eller beskrivelse og diskusjon av mikro-/makropreparat. Alle læringsmål for terminene kan testes. Oppgavene bør være utformet slik at fagintegrasjon og resonnement kommer til anvendelse. Kliniske emner skal bare eksamineres der basalfag kommer til direkte anvendelse. Muntlig/praktisk eksamen og skriftlig eksamen skal planlegges i felleskap. Disse kravene oppfylles i varierende grad.

Sensor og eksaminator på hver post vurderer studentens prestasjon til bestått/ikke bestått. Sensorer og eksaminatorer fra alle tre postene møtes og avgjør eksamensresultatet i felleskap. For å få bestått kan maksimalt en post være vurdert til ikke bestått.

#### **Fordeler:**

- Mulighet til å teste mange overordnede læringsmål

#### **Ulemper:**

- Varierende validitet
- Lav reliabilitet
  - tester et lite utvalg av praktiske ferdigheter
  - sensur basert på sensors og eksaminators skjønn

#### **Konklusjon:**

Muntlig/praktisk eksamen stadium I A/B er en uklart definert eksamen med varierende validitet og lav reliabilitet. Den tester et smalt spekter av de praktiske ferdigheter studentene skal tilegne seg i løpet av det første studieåret.

### **D) *"Double Jump" muntlig/praktisk eksamen Stadium I C/D***

#### **Beskrivelse:**

Studenten gjennomfører en samtale med en pasient med sensor og eksaminator til stede. Deretter skal studenten utføre en begrenset klinisk undersøkelse. Samtale og undersøkelse skal ta maksimalt 20 min. Studenten eksamineres så omkring lege-pasientkommunikasjon og undersøkelse, og skal til slutt definere læringsmål ut fra pasientens problemer, begrunne valget og skissere hvordan han/hun vil gå fram for å oppfylle læringsmålet. Sensor og eksaminator bedømmer studentens prestasjon til bestått/ikke bestått. Det benyttes et skjema med vurderingskategorier for pasientsamtalen.

#### **Fordeler:**

- Gir studenten god tilbakemelding på sine kommunikasjonsferdigheter
- Tilfredsstillende validitet

**Ulemper:**

- Tester ikke de andre praktiske ferdigheter studentene skal tilegne seg i andre studieår
- Lav reliabilitet

**Konklusjon:**

"Double Jump" muntlig/praktisk eksamen er en god eksamen for å teste lege-pasientkommunikasjon. Den har forbedringspotensiale når det gjelder validitet og reliabilitet. Alene fyller den ikke behovet for testing av de praktiske ferdigheter studentene skal tilegne seg i andre studieår.

**E) Tre-stasjons muntlig/praktisk eksamen Stadium II**

Arbeidsgruppen har valgt å beskrive og drøfte denne eksamenen mer inngående, slik den har gjort med de alternative eksamensformene i kapittel 3.

**Formål:**

Muntlig/praktisk stasjonseksamen skal teste kunnskaper som ikke kan testes ved skriftlig eksamen. Spesielt testes studentens praktiske kliniske ferdigheter innen de fag som er undervist i terminen, med hovedvekt på å teste oppvurde kunnskaper i klinisk undersøkelse og tolkning av funn.

**Beskrivelse:**

Muntlig/praktisk eksamen gjennomføres i stadium II felles for to terminer (II a og b, eller II c og d). Dette er terminer som har helt ulikt faglig innhold. Eksamensformen tillater ikke at alle fag som er undervist kan testes. Studentene "kommer opp" i tre fagområder, mens de øvrige fag dermed kun testes skriftlig.

Terminlederne har på forhånd blitt enige om hvilke fag som skal testes, og i hvilken kombinasjon dette skal skje. Det er ingen skrevne regler eller føringer for dette, men i praksis tilstreber man en likelig fordeling av eksaminasjoner fra fagene innen hver termin, og de største fagene gis gjerne større plass enn små fag. Som oftest vil en student dermed bli testet i ett sentralt fag fra hver termin, pluss et lite klinisk fag, et paraklinisk fag eller et basalfag.

**Ressurser:**

Muntlig/praktisk eksamen er satt opp med eksaminator (faglærer) og sensor (ekstern lege) på hver stasjon, i alt 6 leger. Det kreves eksamensvakter, og nærvær av representant fra fakultetet eller terminleder. Kapasiteten per eksamenslokale er tre studenter per time. For hver student krever selve eksamen dermed ca. tre timers arbeid, hvorav to er legetimer. I tillegg kommer medgått tid til forberedelser. Pasienter som deltar i eksamen får et beskjedent honorar og dekning av reiseutgifter.

**Praktiske forhold:**

Eksamen planlegges og administreres av terminleder og fakultetsadministrasjonen i fellesskap. Terminlederne bestemmer fagkombinasjoner. Eksamen avvikles på to steder parallelt, og tar to hele dager (100 studenter). Hver faglærer lager oppgaver og klargjør utstyr, bilder, preparater etc, eller finner pasient(er) som er egnet og villig til å delta.

Studentene møter tre og tre, og føres samtidig inn i hvert sitt eksamenslokale, med eksaminator og sensor. Studenten får en klar, avgrenset oppgave som skal løses på få minutter. Eksaminator har på forhånd utarbeidet en plan og sjekklister, som sikrer at hver student eksamineres på samme måte og bedømmes ut fra de samme krav og kriterier. I den grad det blir tid ut over denne planen testes studentens resonnerende evner og kliniske kunnskaper relatert til oppgaven. Etter 13 minutter ringer en bjelle, og studentene går ut til eksamensvakten. Eksaminator og sensor bruker da to minutter på å diskutere studentens prestasjoner. Deretter ringer bjellen igjen, og studentene går inn til neste eksaminasjon. Etter tre kvarter har de tre studentene kommet igjennom. De sluses til et venterom, mens de seks eksaminatorer og sensorer møtes til felles sensur. Deretter fortsetter eksamen på samme måte for de neste tre studentene.

**Konklusjon:**

Mange vil hevde at 13 minutter er for kort tid til å teste en student tilstrekkelig godt i ferdigheter innenfor et fagområde. Dette er riktig i den forstand at den dypere forståelse for faget, evnen til klinisk resonnerement, strukturert anamneseopptak eller grundig klinisk undersøkelse, i begrenset grad lar seg

teste. Eksamensformen er egnet til å teste hvilke praktiske ferdigheter studentene har opparbeidet seg i et lite utvalg av de ulike fag, f.eks. klinisk organundersøkelse, undersøkelse av patologiske preparater eller gjennomføring av laboratorieprosedyrer. Eksamensformen gir også til en viss grad rom for å teste den kliniske kunnskap og forståelse som er knyttet til disse spesifikke oppgavene. Muntlig/praktisk trestasjonseksamen har svakheter, men er et kompromiss som gjør det mulig å teste avgrensede problemstillinger innenfor tre spesifikke fagområder. Den er ressurskrevende, har god validitet, men lav reliabilitet.

### **F) Tidligere forsøk på OSCE Stadium II**

Det har vært forsøkt å gjennomføre OSCE-eksamen i Stadium II. Arbeidsutvalget har valgt å ta med erfaringene fra dette forsøket i sin vurdering.

#### **Bakgrunn:**

Etter vedtak i Fakultetsstyret ble praktisk/muntlig eksamen i stadium II som en prøveordning gjennomført som en modifisert OSCE i juni 2001. Man ønsket å prøve ut OSCE ut fra en antagelse om at dette kunne teste praktiske ferdigheter ved fagene i terminen bedre enn en tre-stasjonseksamen. Gjennom bruk av mange små stasjoner ville man kunne inkludere oppgaver fra mange av terminens fag og dermed teste langt bredere enn i dag, men da med avgrensede oppgaver.

#### **Beskrivelse:**

Av kapasitetshensyn valgte man å gå bort fra den egentlige OSCE (om lag 20 stasjoner), og forsøkte en modifisert versjon av OSCE. Man valgte å beholde en klinisk stasjon med eksaminator og sensor, og erstattet de to andre stasjonene med seks små (OSCE-)stasjoner med begrensede oppgaver som skulle besvares skriftlig.

#### **Vurdering:**

Den praktiske gjennomføringen av eksamen gikk bedre enn forventet. Imidlertid viste det seg at eksamensformen krevde mer forberedelse enn praktisk/muntlig stasjonseksamen. Det var nødvendig å bruke faglærere til å vurdere besvarelsene ved småstasjonene. Dermed fikk man i praksis en sensur (av eksaminator og sensor) ved den store eksamen, og vurderinger av ulike faglærere ved de seks små stasjonene. Studentene var til dels kritisk til noen av oppgavene ved de små stasjonene, da de mente visse oppgaver like gjerne kunne vært gitt til skriftlig eksamen.

Det var på forhånd bestemt at stryk på den store stasjonen pluss en av de små stasjonene, betød stryk ved eksamen. Svært få studenter klarte "fullt hus" ved småstasjonene. I praksis var dermed stryk på den store stasjonen ensbetydende med stryk ved eksamen.

#### **Konklusjon:**

Denne eksamensformen innebar en betydelig svekket vurdering sammenholdt med tre-stasjonseksamen hvor kandidaten vurderes av tre sett eksaminator/sensor. Forsøket ble evaluert som mislykket, og forlatt til fordel for den etablerte trestasjonseksamen.

Arbeidsutvalget mener OSCE-forsøket slik det ble gjennomført i 2001 er meget forskjellig fra OSCE slik det er tenkt i sin opprinnelige form, med langt flere, og mer likeverdige stasjoner. Forsøket bør derfor ikke utelukke bruk av OSCE-formatet i fremtiden.

### **G) Konsultasjonseksamen Stadium IIIC**

#### **Beskrivelse:**

Av ressurs-hensyn tar eksamen utgangspunkt i ca.5 min (maks. 10 min) av et videoklipp av en konsultasjon som studenten selv har gjennomført og valgt ut. Studenten gir først en kort orientering (2-3 min.) med bakgrunnsinformasjon om pasienten. Videoklippen vises. Studenten gjør rede for resten av konsultasjonen, og begrunner hvorfor han/hun har valgt å vise den delen av opptaket. Det eksamineres deretter hovedsaklig i konsultasjonsprinsipper og -modeller samt kommunikasjon. Eksamenstid er tilsammen 30 min. Studentene får tilbakemelding på hva som var bra/kunne vært gjort bedre.

#### **Fordeler:**

- Vurderer studentenes evner til kommunikasjon og kunnskaper om konsultasjonsprosess
- God validitet
- Gir tilbakemelding før avsluttende muntlig/klinisk eksamen

**Ulemper:**

- Stort formativt potensiale går tapt på grunn av summativ evaluering
- Tar utgangspunkt i kort utdrag av en hel konsultasjon
- Ressurskrevende
- Lav reliabilitet

**Konklusjon:**

En grundig, oppsummerende vurdering av studentenes kommunikasjonsevner og kunnskaper om konsultasjonsprosess etter utplasseringsperioden. Eksamen har forbedringspotensiale hva gjelder en grundig gjennomgang av konsultasjonsprosessen med tanke på formativ evaluering. Dette går tapt da hovedfokus for en eksamen er summativ evaluering. Eksamen har lav reliabilitet.

**H) Muntlig/Kliniske eksamener Stadium IIID****Beskrivelse:**

I dagens ordning er studentene i termin IIID oppe til to eksamener, i ett "stort" fag (definert som indremedisin eller kirurgi) og ett "mindre" fag (allmenmedisin, fysikalsk medisin og rehabilitering, gynekologi / obstetikk, hud og venerea, nevrologi, neurokirurgi, onkologi, ortopedi, pediatri, psykiatri, reumatologi, øre-nese-halssykdommer eller øyesykdommer). Hvilket fag som eksamineres i kunngjøres en uke i forveien, eksaminator blir kunngjort en dag før eksamen. Varighet på muntlig eksamen vil være fra ca 1 til inntil 2 timer. Det forventes at rammen er som ved en sykehusinnleggelse eller ved en poliklinisk undersøkelse ved sykehus med anamnese og nødvendig klinisk undersøkelse. Det er på sin plass at eksaminator innleder med å informere om "settingen" for eksamen. For allmenmedisin, er settingen en konsultasjon i allmenpraksis. Etter anamnese og klinisk undersøkelse vil den etterfølgende eksaminasjon dreie seg om diagnostiske og differensialdiagnostiske overveielser, tolkning av utførte supplerende undersøkelser, forslag til videre utredning og behandling. Det vil også kunne bli eksaminert i forhold til andre problemstillinger i det aktuelle kliniske faget eller relevante momenter.

**Fordeler:**

- Høy validitet
- Tester alle ønskede overordnede læringsmål

**Ulemper:**

- Eksaminasjon i på forhånd annonserte "fag" stemmer dårlig overens med den integrerte studiemodellen og med situasjonen en møter i arbeidslivet, og reduserer både indre og ytre validitet.
- Lav reliabilitet

**Konklusjon:**

Muntlig/praktiske eksamener stadium IIID er gode eksamener på grunn av deres høye validitet, men de har forbedringspotensiale når det gjelder innhold. Det er i uoverenstemmelse med resten av studiet, og tildels med yrket, å opprettholde inndelingen i fag. Eksamnenes svakhet er den lave reliabiliteten.

**3 Alternative vurderingsformer**

Arbeidsutvalget er bedt om å komme med forslag til eventuelle endringer i eksamens- og evalueringsformer. Det er derfor naturlig å redegjøre for de ulike alternativene utvalget har vurdert. Man har valgt å fokusere på aktuelle, gjennomprøvde, tildels standardiserte vurderingsformer som er i vanlig bruk i medisinerutdanning. Utvalget har forsøkt å gjøre rede for spesielle styrker og svakheter ved eksamensformene, med tanke på hvilken rolle de kan få i det helhetlige vurderingsopplegget.

Det er som i forrige kapittel for hver eksamen lagt vekt på:

- hvilke overordnede læringsmål en kan teste
- validitet og reliabilitet
- ressursbruk
- praktiske forhold i forbindelse med gjennomføring av eksamen

## A) *MEQ - Modified Essay Questions*

### Formål:

At teste evnen til klinisk resonnement og problemløsning, gjennom beskrivelse av pasient kasuistik(ker) over tid, sådan som de(n) utvikler sig i det virkelige livet.

### Beskrivelse:

En klinisk problemstilling eller en patients sykehistorie beskrives på noen få linjer. På basis av disse informasjonen, må studenten treffe en rekke valg eller forklare noen forhold. Nye informasjonen kommer til og nye valg må tas eller gamle må korrigeres. Problematikken utvikler sig over tid. Man kan prinsipielt ikke gå tilbake og endre på noe man "har gjort". Man kan heller ikke gå frem før man er ferdig med problematikkene, da man jo ikke "kan se ind i fremtiden".

De spørsmål der stilles kan enten dreie sig om at utvikle hypoteser om årsaker, der er til problemet eller hvilke diagnostiske overvejelser studenten gjør sig. En annen type spørsmål går på hva studenten vil gjøre for at innsamle ytterligere data for at belyse problematikkene og bekrefte eller avkreftede en given diagnose. Dette kan lede videre til spørsmål om problemløsning eller patient "management". Endelig kan man bruke metoden til at spørre om bakvedliggende viten om patofysiologiske eller psykologiske mekanismer, som gir innblik i studentens evne til at kombinere basalfags viten med pasienters helseproblemer. Brukt til eksamen kan den teste mange studenter i en relativ kort periode, for relevante problemstillinger, som ligner på problemstillinger fra det virkelige liv.

Studenten skal gi korte svar og forklaringer på spørsmål der reises i sykehistorien eller problemstillingen. Den kan strække sig over en tidperiode på få minutter (i en konsultasjon) til dele av et langt liv.

Det største problem er at udforme MEQ'en så man unngår at studentene kan læse løsningene på svarene hvis de leser hele oppgaven igjennom før de starter på besvarelse (og det gjør de, med mindre man teknisk forhindrer det). Ideelt set skal kun en del av oppgavene kunne ses og løses ad gangen, og dette er vanskelig med bruk av papir modeller, mens det vil kunne lade sig gjøre hvis man benytter elektroniske besvarelses modeller (computere), som kan sikre, at kun en side kommer frem ad gangen, og at man ikke kan gå tilbake når man først har forladt den.

### Resurser:

Planlegging og testing krever mye tid og involvering av mange fakulty medlemmer for at komme frem til gode spørsmål og svar, og en oppbygning der sikre at studenten ikke kan lese svarene ut av hele oppgaven.

Skoringssystem mht til en enkel og reliabel evaluering er vanskelig og krever både erfaring og forhåndstesting blandt undervisere og sensorer.

Sensur er tidkrevende for eksaminatorer og sensor.

### Konklusjon:

Metoden er serdeles veletnet til testing af studentenes evne til problemløsning og relasjonstenking. Den er veletnet og let at administrere i daglig undervisning, men vanskelig når det dreier sig om en eksamens situasjon, sålenge man ikke har utviklet en computer basert besvarelses model som kan sikre, at man ikke kan se mere end den situasjon man skal besvare. Den er ressurskrevende både i for- og etterkant av eksamen, har tilfredsstillende validitet, men lav reliabilitet.

## B) *Flervalgstester / MCQ - Multiple Choice Questions*

### Formål og beskrivelse:

Flervalgstester er en skriftlig testform der studentene får presentert ferdig formulerte svaralternativer og skal angi hvilke alternativer som har de egenskapene som det spørres etter. Svarene kan gis på ferdigtrykte ark, men datamaskiner i nettverk kan også brukes til å besvare slike oppgaver.

Flervalgstestene har sin styrke i å måle faktakunnskap. Ingen andre vurderingsformer måler kunnskap med samme høye grad av validitet og reliabilitet som en velformulert flervalgstest. Andre menneskelige egenskaper enn kunnskap er vanskelig å vurdere med flervalgstester og denne evalueringsformen er blitt kritisert for å være for endimensjonal til å brukes til viktige eksamener. Egenskapene som ikke så lett lar seg vurdere med en flervalgstest er evne til resonnement, problemløsning, sosial intelligens, samarbeidsevne, kritisk vurderingsevne, evne til muntlig og skriftlig presentasjon, praktiske ferdigheter og kreativitet.

På tross av de erkjente svakhetene ved flervalgstestene, er denne testformen svært mye brukt som evalueringsredskap, også i utdanning av leger. For eksempel er trinn 1 og 2 i The United States Medical Licensing Examination (USMLE) i sin helhet oppbygd som MCQ-eksamener. Disse eksamenene brukes mye i amerikansk medisinstudium og er obligatorisk for utlendinger som ønsker å praktisere medisin i USA. I medisinstudiene i Norge er det en viss bruk av MCQ i Bergen. Siste år var skriftlig øye-eksamen en MCQ. Det brukes også MCQ som en del av en større skriftlig eksamen ved de skriftlige anatomieksamenene og i eksamen i klinisk biokjemi. Ved legeutdanningen i Oslo og Tromsø brukes det ikke MCQ-tester, men det er en diskusjon om MCQ skal innføres på de tidligste stadiene i studiet som et ledd i Kvalitetsreformen. Fagprøven for utenlandske leger som administreres av DMF i Oslo består av flere MCQ-tester.

#### **Ressurser:**

Fra lærer-synspunkt er MCQ-eksamener krevende å designe, men lette å administrere på selve eksamensdagen og i etterkant av eksamen. Det kan benyttes EDB-baserte skåringssystemer som i teorien kan gjøre eksamensresultatet tilgjengelig like etter eksamen.

Det finnes en rekke ulike MCQ-formater og det er mange fallgruver for den som skal lage MCQ-oppgaver (Se Vedlegg). Derfor er det nødvendig at de som skal forfatte oppgaver, får en opplæring i kunsten å lage flervalgstester. Kvalitetssikringen av oppgavene bør tillegges stor vekt og oppgavene bør testes ut på tidligere studenter og andre med kunnskaper på fagfeltet.

En kan også se for seg at det opprettes en database med flervalgsspørsmål knyttet opp mot de ulike terminene. Disse kan brukes i undervisning og "self-assessment", der studentene kan teste seg selv i utvalgte spørsmålssett tilgjengelig på intranett, for eksempel knyttet opp mot læringsmålene for terminen eller sammen med resyméer på Levende Timeplan. Dette vil også særlig i starten av studiet gi studentene nødvendig trening i å besvare flervalgsspørsmål.

På grunn av den enkle skåringsprosedyren vil MCQ-eksamener totalt sett være klart ressursbesparende i forhold til de integrerte skriftlige eksamenene vi har ved medisinstudiet i dag.

#### **Diskusjon om fordeler og ulemper med MCQ i forhold til MEQ og dagens skriftlige eksamensformer ved DMF:**

Forskning som sammenlikner MCQ og MEQ viser nokså entydig en svært høy korrelasjon mellom testskår som studenter oppnår på MCQ og MEQ innenfor et kunnskapsområde. Faktisk er korrelasjonen omtrent så høy som man kan oppnå med den reliabiliteten som de to testformene har. Imidlertid er reliabiliteten klart høyere for MCQ-tester enn for MEQ. Det er altså mindre variasjon i det testresultatet som en student oppnår ved gjentatt MCQ enn ved MEQ. MCQ har i tillegg den åpenbare fordel at retteprosessen er langt mindre ressurskrevende enn for MEQ.

Det aktuelle bruksområdet for flervalgstester ved DMF er å la dem gå inn som full eller delvis erstatning for de skriftlige integrerte årseksamenene. Dersom vi fortsatt skal beholde dagens skriftlige eksamensformer, må vi kunne identifisere egenskaper ved denne vurderingsformen som oppveier de påviste fordelene ved MCQ. Man kan tenke seg flere slike egenskaper, her formulert som påstander:

1. Dagens skriftlige eksamensformer er bedre til å måle tenkning på høyere nivå, som for eksempel problemløsning og resonerende evne.
2. Dagens skriftlige eksamensformer måler evne til skriftlig presentasjon, noe MCQ ikke gjør.
3. Dagens skriftlige eksamensformer gir bedre tilbakemelding til læreren om hvor skoen trykker for studentene, det er lettere å identifisere områder hvor den pedagogiske innsatsen bør forbedres.
4. Dagens skriftlige eksamensformer er bedre til å identifisere svake felter og misforståelser hos den enkelte studenten og er derfor et bedre utgangspunkt for videre dialog mellom lærer og student.

Vi vil nå diskutere de enkelte av disse påstandene.

1. Problemløsning.

Tilhengerne av MCQ hevder at slike tester kan konstrueres på en måte som gjør dem like godt egnet som MEQ/ dagens skriftlige eksamensformer til å teste evne til problemløsning og resonnement. Det stiller imidlertid visse krav til testformatet og det er nødvendig at testforfatterne får en teoretisk innføring i hvordan gode MCQ-tester skal lages.

2. Skriftlig presentasjon.

Denne innvendingen mot MCQ-tester kan like gjerne snus rundt: Den uforholdsmessige fordelene ved tradisjonelle skriftlige eksamensformer for studenter med god evne til skriftlig presentasjon, reduseres med MCQ. Gode språkevner er naturligvis viktig for en lege, men det kan settes spørsmålsteget ved om denne evnen i for stor grad påvirker resultatet ved en tradisjonell skriftlig eksamen. Noen studier har funnet at gutter/menn i gjennomsnitt gjør det bedre på MCQ enn på MEQ og man har ment at dette kanskje skyldes at jenter/kvinner i gjennomsnitt har bedre språkevner. Vår konklusjon på dette punktet er at evne til skriftlig framstillingsevne bør testes i studiet, men at dette kan ivaretas bedre ved andre vurderingsformer, for eksempel ved skriftlige prosjektoppgaver som kan inngå som en del av en mappeevaluering.

### 3. Tilbakemelding til lærerne.

Denne påstanden er ikke riktig. Ved en MCQ vil det være mulig å identifisere områder der studentene gjør det dårlig. Det er snarere lettere enn ved dagens skriftlige eksamensformer, fordi testformatet tillater spørsmål innenfor flere emner i løpet av tiden som er til rådighet.

### 4. Tilbakemelding til den enkelte studenten.

Denne påstanden er nok riktig. Ved MEQ (og tildels dagens skriftlige eksamensformer) er det lettere å følge tankegangen hos en student som har svart feil, enn det vil være ved MCQ. Vår erfaring er imidlertid at skriftlig eksamen i dag i liten grad brukes som utgangspunkt for individuell tilbakemelding til den enkelte studenten. Unntaket her er de få som stryker, men også overfor disse brukes det etter vår erfaring relativt beskjedne ressurser på individuell veiledning. Vårt forslag er at noe av den tiden som frigjøres for eksaminator i etterkant av eksamen ved at den tidkrevende retteprosessen faller bort ved MCQ, brukes til mer forpliktende individuell oppfølging av studenter som har levert svake besvarelser.

### Konklusjon:

Eksamensformen har alene noe lav validitet, men høy reliabilitet. Den krever mye ressurser i forkant av eksamen, men lite etterarbeid. Dersom vi velger å ta i bruk flervalgstester i medisinstudiet i Trondheim, er det viktig at andre egenskaper og ferdigheter som er sentrale for vordende leger enn ren faktakunnskap, testes ved en parallell bruk av andre vurderingsformer. Flervalgstester bør aldri bli den eneste vurderingsformen. Brukt sammen med tester av andre typer, kan flervalgstester bidra til at den totale "evalueringspakken" blir bedre enn om vi velger å måle faktakunnskap med tester som har lavere validitet/reliabilitet på kunnskapsområdet, selv om disse testene måler andre egenskaper på en bedre måte.

## C) OSCE - Objective Structured Clinical Examination

### Formål:

Metode til at teste studentenes kliniske ferdigheter og viten innenfor en lang række områder, og gjøre dette på en relativ objektiv og valid måte. Metoden er egnert til at teste mange studenter på relativt kort tid.

### Beskrivelse:

Studentene går fra en stasjon til en anden. Antallet av stasjoner og tiden på hver kan variere. Flere stasjoner øker reliabiliteten. En anvendt model som gir tilfredsstillende reliabilitet er ca. 20 stasjoner med 5-10 min. på hver. Stasjonene kan være observerte av en lærer ("eksaminator stasjoner") og studenten skal utføre en handling som bedømmes av læreren. På en annen type stasjoner skal studenten svare på spørsmål, tolke pasient data eller nedskrive relevante fund i et præparat eller på en pasient ("skrive stasjoner"). Skuespillere (simulerte pasienter) kan med fordel brukes frem for rigtige pasienter, som ikke kan holde til at møte for mange studenter på en dag. Video avspilninger og anatomiske modeller kan også brukes med stor fordel. Studentene lager besvarelser på fortrykt papir, som er let å summere (evt. data behandle) og dermed hurtigt at point give. Ved eksaminator stasjonene skal der utarbejdes check-lister som eksaminator skal krysse av på, avhengig af den observasjon han gjør.

### Resurser:

Følgende resurser er nødvendige for at gjennomføre en OSCE for et større antal studenter:

1. Plads.
2. Oppbakning til utvikling af stasjoner fra lærerkollegiet.
3. Deltakelse av mange undervisere.
4. Trening av observatørene.

5. Testing av utstyr.
6. Trening og "betaling" af simulerte pasienter.
7. Administrativt personale til koordinering, organisering og gjennomførelse.

**Praktiske forhold:**

1. Innsamling av tilstrekkelig mange test for at give en god vurdering av kliniske ferdigheter er kritisk.
2. Logistikken med mange studenter under avviklingen av en OSCE er en utfordring.
3. Planleggingen er arbeidskrevende, mens avviklingen og evalueringen er arbeidsbesparende.
4. Jo mere rutine, jo enklere er det at gjennomføre.

**Konklusjon:**

OSCE er en reliabel og relativt valid måte at teste mange studenters kliniske viten og ferdigheter på. Den er meget ressurskrevende å planlegge og gjennomføre, særlig de første gangene. En innførelse av en OSCE krever råd og veiledning av noen med stor erfaring i gjennomførelse av OSCE for mange studenter (mer end 100), så den første OSCE innholdsmessig og avviklingsmessig bliver en succes.

**D) Klinisk Avslutningseksamen IIID**

**Formål:**

Med utgangspunkt i en pasients aktuelle problem og med referanse til samlede læringsmål for studiet, forventes å teste bredt kandidatens evne til klinisk resonnement, viten, ferdigheter og holdninger ved å:

1. oppta en systematisk eller problemorientert anamnese
2. utføre en fullstendig eller målrettet klinisk undersøkelse
3. drøfte parakliniske parametre
4. drøfte diagnostiske / differensialdiagnostiske overveielser
5. foreslå videre diagnostisk utredning
6. gjøre greie for basalfaglige aspekter knyttet til pasienten
7. drøfte terapeutiske valg og alternativer
8. vise evne til å kommunisere med pasient og kolleger muntlig og skriftlig
9. medinnbringe pasienten i beslutninger om diagnostik, behandling og fremtidig egenomsorg, herunder kontrol og oppfølging

**Beskrivelse:**

Arbeidsgruppen anser at Klinisk Avslutningseksamen kan bestå av tre tradisjonelle "long-case" eksaminasjoner:

1. omfattende med ramme som en sykehusinnleggelse – 2 timer
2. med ramme som en poliklinisk konsultasjon – 1 time
3. med ramme som en allmenmedisinsk konsultasjon – 1 time

Varigheten på hver del er ment å dekke anamneseopptak, undersøkelse av pasient og påfølgende eksaminasjon.

Tradisjonell inndeling i store vs mindre fag kan kanskje i dagens situasjon virke noe arbitrær og basert på tidligere praksis. Vi er vitne til en økende grenspesialisering innenfor de store disiplinene indremedisin og kirurgi. Utbyggingen av det nye St. Olavs Hospital med sentermodellen delvis på tvers av eksisterende spesialiteter er også et element som kan tilsa at det i større grad vurderes innholdet i eksamen enn at det å priori defineres som indremedisin eller kirurgi. Et alternativ vil være å definere/ navngi eksamen etter graden av omfang / tidsmessig forbruk. Innholdet i den lengste eksaminasjonen kan forventes å være som skissert i Formålsbeskrivelsen ovenfor. Det virker naturlig at settingen for denne eksaminasjonen er en sykehusinnleggelse og at en relativt stor del (anslagsvis 45-60 min) går med til anamnese og undersøkelse. Med slike premisser burde det ikke være reelle motforestillinger mot at en slik eksaminasjon kunne gjøres innen enkelte fag som tidligere er definert som mindre fag, f.eks en pasient med cerebrovaskulær sykdom eksaminert av en nevrolog, eller en pasient med arthritt eksamineres av rheumatolog. Siden en hovedintensjon med Klinisk Avslutningseksamen er å teste evnen til klinisk resonnement, er det ikke naturlig at det signaliseres på forhånd hvilket fagfelt kandidaten kommer opp i.

Det synes ønskelig å kontinuere ordningen med at tidligere såkalte "mindre fag" gis anledning til å eksaminere muntlig også til avslutningseksamen. Som tidligere forutsettes da dette å være i mindre omfang enn den førstnevnte "store kliniske eksamen", f.eks. i settingen en poliklinisk konsultasjon.



Studentene skal også kunne håndtere en allmenmedisinsk konsultasjon. Denne vil teste deres evne til å tenke i sammenheng og integrere alle de kliniske og samfunnsmedisinske fag til en pasientsentrert problemløsning. Det kan derfor være fornuftig å inkludere allmenmedisin i Klinisk Avslutningseksamen.

En tradisjonell "long-case" eksamen ligger tett opp mot en virkelig yrkessituasjon, og har svært høy validitet. Ulempen med en slik eksamen er at bedømmelsen av hver enkelt kandidat helt og holdent beror på eksaminators og sensors skjønn. Dette gir denne eksamensformen lav reliabilitet.

Reliabiliteten økes ved:

- å observere studenten under anamneseopptak og undersøkelse
- å øke antallet pasienter
- bedømmelse av flere uavhengige eksaminatorer og sensorer
- strukturert eksaminasjon og vurdering

Dette søkes oppfylt ved å benytte tre eksaminasjoner med uavhengig sensor og eksaminator for hver eksaminasjon. Sensor og eksaminator skal observere studentens interaksjon med pasienten, hvis mulig via enveisspeil. Bruk av Objective Structured Examination Record (OSLER– Se Vedlegg) som hjelpemiddel til eksaminasjon og vurdering bør vurderes.

Et problem ved vurderingen av kandidatene ved Klinisk Avslutningseksamen er at det ikke er praktisk gjennomførbart at sensorer og eksaminatorer fra de ulike eksaminasjonene kan møtes og avgjøre eksamensresultatet i plenum, slik det praktiseres ved muntlig/praktisk tre-stasjonseksamen. Tilfeller der det er tvil om vurderingen ved flere av eksaminasjonene er vanskelige å fange opp med dagens system. Dette svekker kontrollfunksjonen ved eksamen.

En mulig løsning er å innføre en ny graderingsskala for de enkelte eksaminasjonene; bestått, "borderline" og ikke bestått. Eksamen skal fortsatt vurderes med bestått/ikke bestått. Dersom en student får "ikke bestått" på en eksaminasjon eller "borderline" på to eksaminasjoner, er eksamen ikke bestått.

Det er vanskelig å avgjøre når studentene skal meddeles resultatet fra de enkelte eksaminasjonene. Hvilken praksis som vil påføre studentene minst stress underveis er ikke åpenbart.

#### **Ressurser:**

Avvikling av muntlig eksamen er ressurskrevende, men det krever ikke mye forberedelse og etterarbeid. Det foreligger en begrensning i tilgang og rekruttering av pasienter. Denne kan avhjelpest ved at en tar inn pasienter fra flere fag enn indremedisin og kirurgi.

#### **Konklusjon:**

Arbeidsgruppen anser Klinisk Avslutningseksamen for å være en optimal måte å vurdere om kandidatene har de nødvendige kunnskaper, ferdigheter og holdninger for å utøve legeyrket. En avsluttende eksamen slikt den er skissert ivaretar alle de viktige elementer i den forestående jobb som turnuslege: evner til å kunne ivareta en udifferensiert konsultasjon i allmennpraksis, en fullstendig pasientundersøkelse av en innlagt pasient og en målrettet poliklinisk konsultasjon. Det er disse oppgaver en kandidat i medisin må kunne mestre etter embedseksamen og for trygt å kunne fungere i det norske helsevesen. Eksamen har således den høyest tenkelige validitet. Den lave reliabilitet kan økes ved bruk av standardiserte objektive kriterier for vurdering.

## **E) Mappedevaluering**

*Hva er en mappe?*

En mappe i utdanningsammenheng er en samling av faglig arbeid gjort over en lengre tidsperiode. Den kan inneholde arbeid gjort i ulike arbeidsformer, individuelt eller i gruppesammenheng. Den rommer gjerne flere typer arbeid som er relevant å vurdere for utdanningen, og ikke nødvendigvis bare tekst og bilder.

Mappearbeid kan innlemmes i et studium i forskjellig grad, og grovt kan man skille mellom en enkel og en kompleks modell. I en enkel modell samles arbeidet i mappen og evalueres ved avslutning av studiet eller det stadium av studiet hvor mappearbeidet foregår. I en kompleks modell kan man for eksempel legge inn milepæler med innlevering av arbeid underveis, tilbakemelding fra kontrollelementet og ytterligere finpuss i ettertid, med en endelig evaluering ved slutten.

*Hensikten med en mappe?*

En mappe kan være så mangt, avhengig av den gitte hensikt. Grovt kan det deles inn i læringsmapper, vurderingsmapper og kompetansemapper. Man kan også kombinere overnevnte funksjoner i én mappe.

*Læringsmappen* er først og fremst et redskap for å stimulere god læring.

Her gis studenten frihet til å definere egne mål, innenfor utdanningsinstitusjonens arbeidskrav for å få ta eksamen. Mappen er altså ikke laget med tanke på formell vurdering.

*Vurderingsmappen* er tenkt brukt som et grunnlag for formell vurdering med eller uten karakterer. Mappen skal brukes til summativ vurdering og kan brukes til sammenligning mellom studenter. Det settes mer spesifikke krav til innhold.

*Kompetansemappen* er en samling arbeider som skal dokumentere en students kompetanse.

Hensikten med en mappe er:

- å stimulere til bedre læring ved å gi studenten økt innflytelse over studiesituasjonen
- å gi et bedre grunnlag for evaluering ved at arbeid over et lengre tidsrom presenteres
- å stimulere studenten til mer faglig refleksjon
- å gi stimulere økt selvrefleksjon over eget arbeid og egen faglig utvikling
- å dokumentere studentens kompetanse under og etter utdanning

### **Bruk av mappeevaluering i medisinutdanning:**

En rekke universiteter i utlandet har i varierende grad tatt i bruk mappeevaluering ved utdanning og spesialisering av leger. Denne gruppen har tatt utgangspunkt i legeutdanning og sett på universitetet i Dundee som eksempel. En grov oversikt over Dundees løsning følger nedenfor.

#### *Mappeevaluering ved medisinstudiet i Dundee*

Dundee har definert 12 sentrale læringsresultater ("outcomes") som parametre for vurdering av kandidaten. Noe lignende våre "tre akser" er disse fordelt på tre konsentriske sirkler med pkt. 1 i kjernen, pkt. 2 rundt og pkt. 3 ytterst:

1. "Hva legen er i stand til å gjøre"
  - Kliniske ferdigheter
  - Praktiske prosedyrer
  - Undersøkelse av pasient
  - Håndtering av pasient
  - Helsefremmende og -forebyggende arbeid
  - Kommunikasjonsevne
  - Informasjonshåndtering og -innhenting
2. "Hvordan legen tilnærmer seg problemet"
  - Forståelse av basal, klinisk og sosial vitenskap
  - Passende holdninger, etisk forståelse og juridisk ansvar
  - Passende fattning av beslutninger, klinisk resonnement og bedømmelse
3. "Legen som profesjonell"
  - Legens rolle i helsevesenet
  - Åpenhet for personlig utvikling

Disse læringsresultatene danner rammeverket for læring, undervisning og evaluering av studentene. De er også sentrale i mappen, der studentene skal reflektere over resultatene enkeltvis og gi en skriftlig vurdering av sin egen prestasjon, og levere denne selvverurderingen som mappebidrag.

Arbeidet med mappen foregår i 4. og 5. studieår, altså de to siste år. Studentene mottar muntlig og skriftlig informasjon om mappearbeidet og -eksamen på forhånd. Grovt skissert foregår eksaminasjonen slik:

Fase 1 - Slutten av 4. studieår:

- Studentene har avsluttet hoveddelen av klinikken
- Eksaminasjon av kunnskap og problemløsningsevner med EMI (extended matching items) og CRQ (constructed response questions)
- Eksaminasjon av kliniske ferdigheter med OSCE (objective structured clinical examination)

- Studentene må ha bestått fase 1 før eksamen fase 2

#### Fase 2 - Slutten av 5. studieår:

- Mappeevaluering (=embetseksamen)
- To eksaminatorer leser mappen og finner i samråd sterke og svake punkter som blir tatt opp under muntlig eksaminasjon
- Muntlig eksaminasjon, 40 minutter, konsentrert om mappen
- Individuell karaktersetting, drøfting og konsensusavgjørelse om bestått eller videre evaluering
- Videre evaluering (=ikke bestått) fører til OSCE-eksamen og ny muntlig prøve etter seks uker, med forbedring av svake punkter i mappa
- Hvert læringsresultat bedømmes med karakterer A-D (bestått) og E-G (ikke bestått)

#### Foreløpige konklusjoner fra Dundee om verdien av mappeevaluering:

- Mappeevaluering er et verdifullt bidrag til repertoaret av evalueringsmetoder som til sammen skal gi vurdering av en kompleks utdanning
- Mappebidragene som går på faglig refleksjon og egenvurdering vil gi økt grunnlag for å vurdere legens profesjonalitet (refleksjonsevne er sentralt element, kognitive og humane kvaliteter er to viktige komponenter)
- Mappeevaluering har et potensiale som lærings- og evalueringsform for nye læringsstrategier (som f eks "personal development planning" i Storbritannia).
- Mappeevaluering har høy formativ verdi, den gir studenten god tilbakemelding på egen fremgang og godt grunnlag for å forbedre læringsstrategier underveis
- Mappeevaluering har moderat summativ verdi, den gir foreløpig ikke sikkert grunnlag for karaktersetting og sammenligning mellom studenter. Potensialet er ikke fullt utforsket ennå, studier pågår fortsatt
- Studenters usikkerhet ved innføring av nye eksamensformer må tas hensyn til og søkes minimert ved god planlegging, informasjon og stor stabstilgjengelighet.
- Evalueringsformen må være gjenstand for systematisk revisjon og eventuell forbedring basert på studenters, vitenskaplige ansattes og administrativt personales erfaringer med ordningen.

#### Bruk av mappeevaluering ved medisinstudiet i Trondheim:

Dette utvalgets mandat hva angår mappeevaluering går ut på å foreslå innhold til mappen. Utvalget ser store muligheter for å innføre mappearbeid som del av studiemodellen i Trondheim. I påvente av flere erfaringer fra andre utdanningsinstitusjoner anbefaler utvalget å vektlegge læringselementet i mappeevaluering, men med kontrollpunkter underveis som sikrer at mappearbeidet ivaretas av den enkelte student.

En innføring av multiple choice questions (MCQ) som eksamensform ved alle skriftlige eksamener vil gi behov for trening i skriftlig fremstilling, både for hovedoppgaven i stadium III og for arbeidslivet. Ved innføring av obligatoriske innleveringer som blir å betrakte som mappebidrag, vil vi kunne ivareta dette behovet.

Andre egenskaper vil også være vanskeligere å evaluere med MCQ sammenlignet med dagens skriftlige eksamensform. Dette gjelder særlig resonnement, klinisk tankegang og holdninger. Dette kan bedre ivaretas ved bruk av innleveringer.

Svakheterne ved MCQ kan i noen grad kompenseres for ved bruk av mappeevaluering. Av disse kan nevnes:

- Tankegangen som sier at det ene er rett og det andre er galt
- Tilgjengeligheten av mange alternative løsninger ved en problemstilling, hvorav den ene er helt riktig
- Mulighetene for å gjette blindt og ha 20-25% sjanse for å gi rett svar

Utvalget har under sine møter funnet noen forslag til mappebidrag, men ser at potensialet er stort også for elementer som ikke nevnes her. Forslag til mappeinnhold:

#### Stadium I:

- Rapporter fra kurs
- Besvarelser til praktiske prøver (i f eks morfologi og makroanatomi)
- Erfaringer fra en allmennpraksis-konsultasjon
- Elementer fra Lege/Pasient-kurs-permen

**Stadium II:**

- Kasuistikker
- Journaler
- Fordypningsoppgave, prosjektoppgave el. innenfor selvvalgt eller tildelt felt

**Stadium III:**

- Hovedoppgave og eksaminator/sensors kommentarer
- Ferdighetslister fra praksis
- Kasuistikker
- Journaler
- Videoopptak fra konsultasjon i allmennpraksis
- Samfunnsmedisinoppgave
- Artikkelvurdering
- Case-diskusjoner

Andre elementer som kan vurderes inkludert, men må sees i sammenheng med vektlegging av mappeevalueringen ift andre eksamener og summativ versus formativ vektlegging:

- Skriftlig evaluering av mappebidragene – både av studenten selv, lærere og evt. medstudenter
- Studentens skriftlige betraktninger om egen prestasjon innen gitte læringsresultater
- Formell vurdering gjort av veiledere under praksisperiodene

Innføring av mappeevaluering vil ikke nødvendigvis kreve mye mer av kontrollressurser. En fornuftig fordeling mellom "self assessment", "peer assessment" og "teacher assessment" kan hindre at "få får mye" merarbeid. DMF har et bare delvis utnyttet potensiale hva angår motiverte og kvalifiserte studenter med og uten læringsassistentkurs. Som ved andre institutter ved NTNU kan studenter gjøre mye av grovarbeidet og dermed skjerme de relativt få lærerressurser blant de vitenskapelig ansatte.

**Konklusjon:**

Utvalget ser mulighetene for å benytte mappeevaluering som læringsredskap. Etterhvert vil fakultetet sitte med egne og andres erfaringer med mappeevaluering for i neste omgang å vurdere å benytte dette til summativ vurdering av studenter. En eventuell debatt om å la mappeevaluering erstatte oppsummerende eksamener, slik kvalitetsreformen anbefaler, ligger derfor flere år fram i tid. Planlegging av en eventuell innføring i Trondheim er en stor oppgave som vil kreve mye arbeid og bør gjøres av en egen arbeidsgruppe. Arbeidsgrunnlaget for en slik gruppe bør avklare blant annet følgende spørsmål:

- Hva er viktigst av kontrollfunksjonen og læringsfunksjonen, og hvordan skal den innbyrdes balansen være? Skal mappen kun være en dokumentasjon av kompetanse?
- Hvordan skal mappen vektet i forhold til andre eksamener?
- Evalueringsbehovet: Hvilke kvaliteter skal dekkes mappeevaluering?

Følgende problemområder bør være belyst og vurdert av arbeidsgruppas oppdragsgiver:

- Studenters usikkerhet omkring mappeinnholdet versus for sterke føringer (standardisering) av mappeinnholdet
- Meningsfullt mappearbeid versus papirjakt
- Personlig preg på mappen (autensitet, validitet) versus minstekrav av kriterier for å kunne evaluere arbeidet riktig (samsvar mellom eksaminatorer, samsvar i bedømmelsen av alle kandidatene, reliabilitet)
- Ressursbruk (eksaminatortid) og fakultetets begrensninger
- Bruk av eksterne sensorer (forutsetninger om god kjennskap til fakultetets sentrale målsetninger og pedagogiske filosofi, fare for bias)
- Skal det settes karakterer, og skal disse settes underveis eller til slutt?

**4 Forslag til endringer i eksamensordning**

En nærmere redegjørelse og begrunnelse for de enkelte deler av arbeidsutvalgets forslag til ny eksamens- og evalueringsordning i studiet.

**A) Skriftlige integrerte flervalgstester**

Arbeidsutvalget går inn for å erstatte skriftlig eksamen med flervalgstester (MCQ).

**Fordeler:**

- Redusert ressursbruk, særlig i etterkant av eksamen
- Mulighet for å forkorte sensurtiden
- Mulighet for å korte noe ned på eksaminasjonstid
- Økt indre validitet på grunn av mulighet for å teste langt flere fagområder
- Økt reliabilitet
- Økt forutsigbarhet på grunn av standardiserte spørsmålsformer

**Ulemper:**

- Krevende å teste resonnement og problemløsning
- Tester ikke skriftlig framstilling og formuleringsevne

Skriftlig eksamen er ikke velegnet for å teste skriftlig framstilling og formuleringsevne. Det er diskutabelt om språkferdigheter bør spille inn på resultatet av en kunnskapstest.

**Begrunnelse:**

Arbeidsutvalget ser for seg at skriftlig eksamens funksjon i hovedsak skal være å teste studentenes kunnskapsnivå og deres evne til å anvende sine kunnskaper teoretisk. Det er ønskelig å øke reliabiliteten ved skriftlig eksamen. Vurdering av skriftlige kommunikasjonsferdigheter, bruk av fagterminologi, resonnement, problemløsningsevner mm. kan med fordel ivaretas ved parallell bruk av mappeevaluering.

For at bruk av flervalgstester skal bli vellykket bør følgende punkter oppfylles:

- Det må utarbeides føringer for hvilke spørsmålsformuleringer og hvilket skåringssystem som skal benyttes. (se vedlegg: Veiledning for forfattere av flervalgstester)
- Studentene må på forhånd informeres om hvilke systemer som vil bli benyttet.
- Det er nødvendig at testforfattere kurses i å forfatte flervalgstester.
- Studentene bør ha tilgang til øvingstester i løpet av året for å venne seg til spørsmålsformatet.

**B) Bruk av PBL ved skriftlig eksamen Stadium I**

Arbeidsgruppen går inn for å avvikle ordningen med PBL-oppgave i forkant av skriftlig eksamen Stadium I, da PBL-metoden ikke bør benyttes til summativ evaluering. Ordningen har ikke vist seg å tilføre eksamen vesentlige forbedringer. Det vil i tillegg redusere ressursbruken ved planlegging og gjennomføring av eksamen. Ønsker man å vurdere studentenes evne til å utarbeide og oppfylle læringsmål kan dette gjøres best som en del av mappeevalueringen, og ikke i eksamenssammenheng.

**C) OSCE Stadium I**

Arbeidsgruppen går inn for å erstatte muntlig/praktisk eksamen stadium I A/B med en integrert OSCE-eksamen ved slutten av stadium I.

**Fordeler:**

- Mulighet for å teste bredere utvalg av praktiske ferdigheter
- Mulighet for å teste praktiske ferdigheter også fra stadium I C/D
- Økt reliabilitet

**Ulemper:**

- Noe økt ressursbruk
- Stor eksamen med mye stoff

**Begrunnelse:**

Arbeidsgruppen mener muntlig/praktisk eksamen for stadium I i hovedsak bør teste de mange praktiske ferdigheter studentene skal tilegne seg i løpet av studiets to første år. Unntak er kommunikasjonsferdigheter, som testes bedre ved en egen grundigere eksamen som i dag. Eksaminasjon av teorikunnskaper kan med fordel i hovedsak utføres ved skriftlig eksamen.

Reliabiliteten ved muntlig/praktisk eksamen bør økes. Det er behov for å testet de praktiske kunnskaper studentene tilegner seg i løpet av andre studieår på en bedre måte enn i dag.

En større avsluttende eksamen som dette krever at studentene får tilbud om god og systematisk evaluering av sine praktiske ferdigheter underveis, og gjerne repetisjonsundervisning. Dette fungerer i dag tilfredsstillende når det gjelder kliniske ferdigheter (i form av Lege-pasientkurset), men kan med fordel gjøres mer systematisk, for eksempel ved mer gruppeundervisning, innen andre fagområder som for eksempel morfologi.

#### **D) Kommunikasjonseksamen I C**

Arbeidsgruppen går inn for at dagens "Double Jump"-eksamen i lege-pasientkommunikasjon, informasjonssøking og klinisk undersøkelse stadium I C/D erstattes med en eksamen i lege-pasientkommunikasjon stadium IC (desember). Studenter som ikke består eksamen bør få tilbud om ytterligere oppfølging og veiledning fram mot kontinuasjonseksamen.

##### **Fordeler:**

- Økt fokus på kommunikasjon ved å overføre testing av undersøkelsesteknikk til OSCE stadium I
- Tidligere evaluering og vurdering av studentene gir tid til forbedring før kliniske terminer
- Fordeling av ressursbruk ved å flytte eksamen til desember

##### **Ulemper:**

- Flytting av eksaminasjonen fra juni til desember kan skape problemer i samarbeidet med dramastudiet.

##### **Begrunnelse:**

Arbeidsgruppen ønsker å beholde en god evaluering av studentenes kommunikasjonsferdigheter i stadium I. Det bør legges vekt på å gi den enkelte student en grundig tilbakemelding etter eksamen. Dagens "Double Jump" kan med fordel endres slik at eksaminasjon av undersøkelsesteknikk overlates til OSCE stadium I, og vurdering av læringsmål/informasjonsøk utgår. Informasjonsøk kan som nevnt best vurderes i mappeevalueringen. Fortsatt bruk av standardiserte, objektive vurderingskriterier anbefales.

#### **E) OSCE / Muntlig/Praktisk eksamen Stadium II**

Arbeidsutvalget inn for å vurdere innføring av OSCE opp mot fortsatt bruk av 3-stasjons muntlig/praktisk eksamen når det foreligger ytterligere erfaringer med bruk av OSCE, og evt. nye, mer egnede lokaler står klare.

##### **Fordeler:**

- Økt reliabilitet
- Mulighet for å teste langt flere praktiske, kliniske og parakliniske ferdigheter

##### **Ulemper:**

- Sannsynlig økt ressursbruk
- Ytterligere redusert mulighet for å teste klinisk resonnement muntlig

Vurdering av klinisk resonnement, anamneseopptak, holdninger og dypere forståelse er tenkt ivaretatt ved bruk av mappeevaluering.

Det er på nåværende tidspunkt av praktiske årsaker svært vanskelig å gjennomføre en OSCE-eksamen. Arbeidsutvalget går derfor inn for å utsette en eventuell innføring av OSCE til nye lokaler er tilgjengelige, og erfaringene fra gjennomføring av en eventuell OSCE for stadium I foreligger.

##### **Begrunnelse:**

Arbeidsutvalget ønsker at muntlig/praktisk eksamen stadium II skal teste et bredere utvalg av praktiske, kliniske og parakliniske ferdigheter enn ved dagens tre-stasjons eksamen. Det er ønskelig å øke reliabiliteten fra dagens eksamensordning. Klinisk resonnement, strukturert anamneseopptak og dypere forståelse kan med fordel vurderes ved parallell bruk av mappeevaluering. Arbeidsutvalget mener 3-stasjons muntlig/praktisk eksamen inntil videre er et tilfredsstillende kompromiss for praktisk

eksaminasjon av de fag som er undervist i terminen. Det bør vurderes å utarbeide standardiserte, objektive kriterier for vurdering for å øke reliabiliteten.

### **F) Konsultasjonsevaluering IIIC**

Arbeidsutvalget går inn for å erstatte muntlig/praktisk "video"eksamen IIIC med en formell evaluering med utgangspunkt i og gjennomgang av et videoopptak av en konsultasjon og mappa fra studentens utplasseringsperiode. Studenten bør levere en egevaluering av konsultasjonen sammen med videoen. Hovedvekten bør ligge på kommunikasjonsferdigheter og konsultasjonsprosess. Det bør utarbeides føringer for hvordan møtet skal foregå, og hvilke temaer som skal drøftes. Evalueringen skal være formativ.

Evalueringen er tenkt å inngå som en del av mappeevalueringen. Denne evalueringen er også tenkt å fange opp de studentene som trenger tilbud om ytterligere oppfølging av kommunikasjonsferdigheter fram mot avsluttende eksamen i klinisk medisin IIID.

#### **Fordeler:**

- Mulighet for grundigere og bedre evaluering enn ved dagens eksamen
- Formativ evaluering
- Redusert ressursbruk, blant annet ikke krav til ekstern sensor og egen kontinuasjon
- Fortsatt mulighet til å fange opp studenter som trenger tilbud om ytterligere oppfølging

#### **Ulemper:**

- Mister kontrollfunksjonen ved en eksamen

Kontrollfunksjonen blir ivarettatt ved Klinisk Avslutningseksamen, særlig ved utvidet eksaminasjon i allmenmedisin.

#### **Begrunnelse:**

Arbeidsutvalget ønsker i likhet med studentene å beholde en god og grundig evaluering av studentenes kunnskaper og ferdigheter i forbindelse med den allmenmedisinske konsultasjon, og muligheten til å fange opp de studenter som trenger tilbud om ekstra oppfølging av dette. Denne evalueringen styrkes ved å benytte video av en hel konsultasjon, framfor dagens kompromiss med et fem minutters utdrag. Arbeidsutvalget mener det ikke er nødvendig å gjøre dette i form av en eksamen, da studentene på dette stadiet har best nytte av en formativ evaluering. Et formativt utgangspunkt vil påvirke studentens valg av konsultasjonsvideo, og hans/hennes framstilling av sine evner. Denne ordningen blir også langt enklere og mindre ressurskrevende å arrangere enn en eksamen.

### **G) Klinisk Avslutningseksamen IIID**

Arbeidsutvalget går inn for å endre muntlig/kliniske eksamener i stadium IIID til en tredelt Klinisk Avslutningseksamen. Det kan med fordel utarbeides standardiserte, objektive kriterier for vurdering, for eksempel ved bruk av OSLER. (Se Vedlegg)

#### **Fordeler:**

- Økt validitet
- Styrket vurdering av eksaminasjon i allmenmedisin
- Økt reliabilitet
- Mulig fordeling av ressurskrav på flere fagmiljøer

#### **Ulemper:**

- Økte ressurskrav til det allmenmedisinske fagmiljø

Dette kan i noen grad kompenseres for ved å erstatte konsultasjonseksamen IIIC med en formell evaluering.

#### **Begrunnelse:**

Arbeidsutvalget mener muntlig/klinisk eksamen er en svært god eksamensform som bør beholdes som avsluttende eksamen, men det er nødvendig å øke reliabiliteten. De skisserte endringer vil øke reliabiliteten, validiteten og gjøre eksamen mer i overenstemmelse med resten av studiet.

### **H) Mappeevaluering**

Arbeidsutvalget går inn for at det etableres en langsgående mappeevaluering gjennom hele studiet. Denne skal brukes parallelt med de eksisterende/foreslåtte vurderingsformer. En mappeevaluering vil kunne bli et nyttig læringsredskap, og et svært verdifullt tilskudd til den totale evalueringss"pakken" i studiet. Mappeevalueringen kan særlig brukes til å vurdere overordnede læringsmål som ikke egner seg til vurdering ved eksamen.

Det er nødvendig å oppnevne et eget utvalg som kan se nærmere på utformingen og tilpasningen av mappeevaluering til studiemodellen. Det er mulighet for å prøve ut mappeevalueringer i deler av studiet, fortrinnsvis i utplasseringsperiodene og i forbindelse med lege-pasientkurset. Se eget kapittel om Mappeevaluering 3 D).

## **5 Forslag til endring av eksamensavvikling og kontinuasjonensordninger**

### **A) Integrering av forskerlinja**

Innføring av forskerlinjen gjør det nødvendig å arrangere eksamen for stadium II i overgangen fra høst- til vårtermin. Det synes mest fornuftig å arrangere eksamen etter at undervisningen i terminen er fullført, altså i slutten av januar.

Arbeidsutvalget går inn for å beholde ordningen med integrert eksamen etter hvert studieår. Med hensyn på kapasitet og ressursbruk er det ikke ønskelig å innføre avsluttende eksamen etter hver termin.

### **B) Endringsforslag til avvikling av kontinuasjonseksamen**

#### **Stadium I:**

Tidspunkt for kontinuasjonseksamen i stadium I forblir uendret.

#### **Stadium II:**

Ut i fra ressursmessige hensyn er det ønskelig at avviklingen av kontinuasjonseksamen for stadium II slås sammen med ordinær eksamen for forskerlinja. De som stryker til ordinær eksamen (vår) i stadium II fremstiller seg til kontinuasjonseksamen i januar. Kontinuasjonseksamen for forskerlinjestudentene (evt. andre studenter som av ulike grunner har sitt første forsøk i desember) blir ved neste ordinære eksamensavvikling i juni.

#### **Stadium III:**

De studenter som ved oppsummerende evaluering etter utplassering i allmenpraksis stadium IIIC vurderes for svake i kommunikasjonsferdigheter skal tilbys ekstra oppfølging og evaluering fram mot muntlig eksamen i klinisk medisin.

Dagens kontinuasjonensordning etter stadium IIID er lite tilfredsstillende. Det er svært lite sannsynlig at studenter som stryker til embedseksamen klarer å tilegne seg manglende kunnskaper og ferdigheter i løpet av tre uker. Dagens ordning foreslås erstattet av følgende ordning:

- Kontinuasjonseksamen for skriftlig eksamen IIID avvikles i august 2
- Kontinuasjon av muntlig eksamen i klinisk medisin avvikles i november. Alle deler av eksamen kontinueres. Studenter skal tilbys ekstra oppfølging og veiledning i perioden fra sensur til kontinuasjonseksamen. Denne ordningen lar seg kombinere med "venteordningen" for turnustjenesten, slik at studenter som må kontinuere kan starte opp turnustjenesten i januar.

## **6 Utplassering**

Arbeidsutvalget er bedt om å foreslå hvordan studentene kan evalueres og vurderes i utplasseringsperiodene. Dette bør sees i sammenheng med ferdighetsgruppas innstilling.



**A) Lokalsykehus**

Evaluerings og vurdering av studentenes utvikling i utplasseringsperioden er en utfordring fordi studentene er spredt og befinner seg langt unna fakultetet. Vurderingen slik den fungerer i dag består i at studentene leverer inn en "ferdighetsliste" der de kvitterer for hvilke undersøkelser og prosedyrer de har utført i løpet av utplasseringsperioden, i tillegg til at utplasseringsstedet "godkjenner" oppholdet. Tilbakemeldingen fra studentene tyder på at mange savner et opplegg for evaluering i denne perioden. Det er også fra fakultetes side et ønske om en bedre vurdering av studentene etter utplasseringen.

**Mappeevaluering:**

Arbeidsgruppen mener at mappeevaluering er den best egnede evalueringsmetoden i denne perioden. Det konkrete opplegget for mappeevalueringen under utplassering i lokalsykehus bør sees i sammenheng med "ferdighetsgruppas" instilling. Studentens egnevaluering bør utgjøre et betydelig innslag. Det er naturlig å benytte legene ved utplasseringsstedet til evaluering/vurdering av mappebidrag, for eksempel journalnotater og case-diskusjoner. Utplasseringsperioden er godt egnet til å prøve ut mappeevaluering som metode før den evt. innføres som en del av det helhetlige evalueringssopplegget ved studiet.

**Oppfølging underveis:**

Det bør i forbindelse med mappeevalueringen og i sammenheng med "ferdighetsgruppas" instilling utarbeides rutiner for mer veiledning og med direkte tilbakemelding på studentenes praktiske ferdigheter. En måte å oppnå dette kan være å omarbeide ferdighetslista slik at det kommer tydeligere fram at en del av prosedyrene skal skje under veiledning, for eksempel ved at den aktuelle veileder signerer i lista etter å ha observert studenten og gitt tilbakemelding.

Det bør etableres en form for tutorordning for studentene ved utplasseringsstedet. Det er naturlig at legene ved utplasseringsstedet fungerer som tutorer. Det er et ønske fra studentene å få mer kontakt med erfarne leger under utplasseringsperioden.

Tutorordningen kan tenkes etablert på flere måter, for eksempel at hver av studentene får en lege på utplasseringsstedet som sin personlige tutor, med jevnlig møter, f.eks. 30 min-1 time en gang i uka eller annenhver uke. Eventuelt kan studentene på et utplasseringssted fungere som en tutorgruppe som samles jevnlig sammen med gruppens tutor. Hvis dette blir "praksislærers" oppgave kan en også benytte dette forum til fortløpende evaluering av opplegget på utplasseringsstedet.

**B) Allmenpraksis**

I stadium IIIC er studentene utplassert ved et allmenlegekontor i en periode på seks uker. Evaluering skjer underveis av legen(e) på stedet, løpende og i form av en midtveis-evaluering. I tillegg tar en utgangspunkt i en video av en konsultasjon fra perioden som brukes ved muntlig/praktisk eksamen IIIC. Det er ønskelig å utvide den fortløpende evalueringen under utplasseringsperioden i stedet for en avsluttende eksamen, i tråd med kvalitetsreformen. Studentene mener det er nyttig med en avsluttende/oppsummerende evaluering fra en "ekstern" lærer på fakultetet i tillegg til tilbakemeldingen fra legen ved utplasseringsstedet.

**Mappeevaluering:**

Som for utplasseringen i lokalsykehus (se over), med et par merknader angående innholdet i mappa: Det eksisterer allerede en "perm" til studentbruk for utplasseringsperioden. Denne kan være et utgangspunkt for utarbeidelse av mappeevalueringen. Studentene skal også i løpet av perioden utarbeide en samfunnsmedisinsk oppgave, som er en naturlig del av en eventuell mappeevaluering. Det er i tillegg ønskelig at det stilles krav til økt bruk av videoopptak av konsultasjoner, og jevnlig evaluering av disse. Det er naturlig at legen på utplasseringsstedet vurderer disse videoene, i tillegg til studentens egnevaluering. Dette forutsetter at fakultetet utstyret utplasseringsstedene med videoutstyr. Det er ønskelig å oppnå 5 konsultasjoner på video med gjennomgang og evaluering fra student og lærer i løpet av perioden.

**Oppsummerende/avsluttende evaluering:**

Se egen beskrivelse av Konsultasjonsevaluering IIIC i kapittel 4 F).

## 7 Personlig oppfølging og veiledning

Kvalitetsreformen (st.meld. nr. 27) går inn for økt personlig oppfølging og veiledning av studentene. Arbeidsutvalget er bedt om å gjøre rede for hvordan studentene kan få mer jevnlig, systematisk tilbakemelding enn i dag, og spesielt se på hvordan PBL-veilederne kan brukes i denne sammenheng. Arbeidsgruppen har valgt å fokusere ikke bare på tilbakemelding på faglige prestasjoner, men også på andre kvaliteter en har bruk for i legeyrket.

### A) PBL-veilederne

#### Bakgrunn:

Arbeidsgruppen er bedt om å vurdere om PBL-veilederne kan gi mer jevnlig tilbakemelding til den enkelte student. Per i dag er PBL-veilederens rolle å veilede studentene i gruppeprosess på PBL-møtene. Det er etablert et system med et "evaluerings-PBL-møte" i midten av terminen, der medstudenter og veileder evaluerer hver enkelt student, med vekt på hvordan en fungerer og hvilken rolle en har i gruppa. Ideen bak å benytte PBL-veilederne til mer utstrakt evaluering og vurdering av studenten har utgangspunkt i at veilederne har mye og jevnlig kontakt med studentene i løpet av terminen.

En kan tenke seg to prinsipielt ulike utvidede funksjoner der PBL-veilederne benytter sin observasjon av studenten:

- Å gi tilbakemelding om studentens læringsprosess, arbeidsmetoder og faglige nivå, med tanke på å hjelpe studenten til å forbedre og effektivisere disse (altså en ren formativ evaluering).
- Å utarbeide en vurdering av studenten med tanke på funksjon og rolle i gruppa, personlig læringsprosess og arbeidsmetoder, og faglig nivå. Vurderingen inngår som en del av mappeevalueringen (altså som en del av en summativ evaluering).

Det er imidlertid motsetninger ved de to funksjonene. PBL er en læringsmetode, og PBL-møtene er en læringssituasjon. Det kan i utgangspunktet virke som en god idé å vurdere hvordan studenten opptrer i en læringssituasjon. Problemet er følgende: I en lærings-/formativ situasjon, (eks. PBL-møtet) ønsker studenten å avdekke mangler i sin egen kunnskap og forståelse for å kunne utbedre disse. I en summativ vurderingssituasjon vil studenten framstille seg og opptre på en annen måte. Han vil her framheve sine sterke sider, og så godt som mulig dekke over sine mangler og svakheter. Dette er også studentens soleklare rett. En kan derfor ikke blande sammen formative og summativ evalueringer. (Lauvås/Jakobsen). Det er uhenksom å blande vurdering med PBL-metoden, fordi dette vil ødelegge PBL-metodens funksjon. Det er derimot ikke problematisk å utvide PBL-veilederens formative evaluering av studentene.

#### Metode:

Arbeidsgruppen har sett for seg en ordning der veilederen skal i tillegg til "evaluerings-PBL-møtet" ha rent formative "personalsamtaler" med hver enkelt student på gruppa, varighet ca. 15 min. I samtalen tar en opp på forhånd definerte eller aktuelle spørsmål/problemstillinger som angår studentens læring, arbeid og funksjon.

Dersom dette skal fungere som en utvidelse av den eksisterende evalueringen i PBL-systemet bør en gjennomføre dette to ganger i semesteret, etter ca.  $\frac{1}{4}$  og  $\frac{3}{4}$  av terminen.

#### Ressurser:

Hver runde vil belaste veileder med to timer, tilsvarende ett PBL-møte. En kan diskutere om dette skal foregå i timene man vanligvis har PBL-møter, og altså bruke ytterligere to PBL-møter i terminen til evaluering, eller om en skal sette av tid utenom PBL-møtene og påføre PBL-veilederne tilsammen ca. fire timer ekstra veiledning i semesteret. Dette vil i så fall utgjøre ca. 25 veiledningstimer per kull per termin.

#### Fordeler:

- Studentene kan få individuell og systematisk veiledning og tilbakemelding fra en veileder som observerer dem jevnlig gjennom hele semesteret. Dette vil særlig styrke den formative evalueringen av faglig/teoretisk nivå og arbeidsmetoder. En slik ordning vil oppfylle noe av intensjonene i kvalitetsreformen.

#### Ulemper:

- En slik ordning vil være ressurskrevende. Det vil øke kravene som følger med PBL-veilederjobben. Det kan gjøre det mer krevende og mindre attraktivt å være PBL-veileder, og redusere rekrutteringen til jobben.
- Det eksisterer allerede et omfattende system for evaluering i forbindelse med PBL. Både studenter og veiledere synes det er et problem at det brukes så mye tid på dette, og oppfatter situasjonen slik den er som utstrakt "over-evaluering", med mye papirarbeid uten særlig gevinst. En ytterligere utvidelse av evalueringen kan derfor bli dårlig mottatt, og tilsvarende dårlig fulgt opp både av studenter og veiledere.

#### **Konklusjon:**

PBL-veilederene kan ikke under noen omstendighet benyttes til summativ vurdering. PBL-arbeidet må ikke benyttes eller blandes inn i vurderingssammenheng.

En systematisk utvidelse av PBL-veilederenes oppgave må vurderes nøye. Den økte ressursbruken må veies opp mot en eventuell gevinst. Utvidelsen må oppfattes som nyttig og verdifull av både studenter og veileder. En ordning slik beskrevet har stort potensiale, men må sees i sammenheng med det omfattende evalueringssystemet som allerede er knyttet opp mot PBL-ordningen. Arbeidsutvalget anbefaler at evalueringssopplegget knyttet opp mot PBL revideres i sin helhet, med en utvidelse av PBL-veilederenes evalueringssoppgaver som en mulig utvidelse.

### **B) Mentor-ordning**

#### **Bakgrunn:**

Når studentene starter på studiet, er der mange der ikke har noe forhåndskjennskap hverken til studentmiljøet, byen eller til studiet selv. Ikke alle medisinstudenter har familie der er leger eller jobber i helsesektoren. For disse studenter kan starten på medisinstudiet bli noe av et "kultursjokk". Flere steder (Oslo, Odense, London) har man derfor overveiet hvordan man best mulig kan integrere studentene i studiemiljøet og gi dem en opplevelse av, at være med i et medisinsk fagmiljø. Noen steder holder man introduksjonsdager for de studerende, hvor eldre studenter og yngre lærere prøver at inn dra dem i studiet og informere dem om de kommende studieår. Studentene kommer herved til at kjenne hverandre bedre, og det bliver lettere at etablere grupper og få dem til at fungere sammen.

#### **Metode:**

Der er vanligvis ringe kontakt mellom studentene på tvers av årgangene. For at fremme denne kontakt har man prøvet at utnevne mentorer blant lærene. Disse mentorer er undervisere ansatt ved DMF, som en del av deres akademiske forpliktelser, også skal påta sig at være mentor for en gruppe på 6 studenter, en student fra hvert kull (evt. 2 fra noen kull, for at få gruppe opp på 8-10, som er nødvendig for en god gruppeprosess).

Legen=læreren fungere da som fasilitator for gruppe prosessen. To gange pr. halvår invitere han studentene til et møte - enten på sit kontor, på avdelingen, i praksis eller hjemme. Målet er at få studentene til at utveksle erfaringer på tvers av semestrene, således at de eldre studerende beretter og hjelper de yngre studerende. De yngre studerende får sat ord på deres forventninger, høre om de "rare undervisere" på neste semester osv. Mentor kan forberede de eldste studerende på turnus tiden, spesialist utdanningen og livet etter embetseksamen. En velfungerende gruppe, kan bli en "basisgruppe", dvs en gruppe hvor studentene kan hente hjelp når de trenger det av studiemessige eller personlige grunde ("faddere" for hverandre).

#### **Fordele:**

1. Mentoren kan være simulator (rollemodel) for studentene og ofte også en personlig veileder, når en student får det vanskelig under studiet.
2. Bibringe studenten spennende opplysninger om medisinenes allsidighet, muligheter og begrensninger.
3. Fungere som en gjennomgående person i hele studiet som studentene kan forholde sig til og spørre til råds (en funksjon som kan fortsette ut over studiet, fordi man bliver personlig knyttet til hinannen gjennom 6-7 års kontakt).
4. Fakultets medlemmer får en annen rolle overfor studentene enn blot den undervisningsmessige, de bliver interessert i studentene som personer på en hel ny måte.
5. Lærerne får et nyt syn på studiet, de enkelte fag, inkl. deres eget og på kollegenes undervisning og omdømme blant studentene.

6. Der er minst et fakultetsmedlem som kjenner den enkelte student personlig. Han/hun kan være en verdifull resurs for DMF, hvis man opplever at studenten har problemer.

#### Ulemper:

1. Utskiftningen blant lærene gjør at ikke alle grupper fungerer bra.
2. Ikke alle studenter og lærere bryter sig om denne forpliktende kontakt, og de vet ikke hva de ska stille opp med hinannen.
3. Hvert år kommer der en ny student, og en "gammel" reiser fra gruppen. Kan gi "uro" som forstyrre gruppe prosessen.
4. Krever et par ildsjel for at få det til at fungere. Mentorene må samles i blant, og de skal få en opplevelse av at dette er bra for studentene og dem selv.
5. Uklar rolle kan forvirre studentene og gjøre denne utrygg. Mentoren må ikke bedrive summativ evaluering, men være en "trusted adviser and helper".

#### Konklusjon:

Arbeidsutvalget mener en mentor-ordning kan bli et verdifullt tilskudd til medisinstudiet i Trondheim. For at ordningen skal fungere best mulig for alle parter skal følgende være oppfylt :

- Frivillighet for studentene og lærerne. Lærerne kan godt i stillingsbeskrivelse få at vite, at det er en funksjon man forventer de påtar sig på linje med andre oppgave som undervisning, eksamen, forskning og deltakelse i curriculum utvikling.
- Der skal være et program, som kan være inspirasjon og veiledning til hvordan mentor-student samlingene kan foregå. Hjelp til de lærere som ikke kan finne ut av det selv.
- Det bør kun være læger der er mentorer. Studentene ønsker at høre om legeyrket og livet som lege, og det kan "ikke leger" ikke fortelle dem.

## 8 Karakterer

Det har i det siste vært en debatt innad på fakultetet angående gjeninnføring av karakterer ved studiet. Bakgrunnen for saken er implementeringen av kvalitetsreformen og revisjonen av studiet. Både kvalitetsreformen og NTNU åpner for å beholde bestått/ikke bestått, eller innføre en karakterskala A-F. Fakultetet står derfor fritt til å velge om man ønsker å gjeninnføre karakterer eller ikke. Arbeidsutvalget har fått i oppdrag å gi råd om innføring av karakterer. Da man ikke har lyktes i å komme til enighet innad i utvalget har man valgt å gi en grundig innføring i saken, uten å ta stilling i saken.

### A) Bakgrunn

Fra St.meld. 27 (kvalitetsreformen) punkt 5-3-6 Karakterer:

"I Universitetsrådets innstilling *Mål med mening* (rapport 2000) foreslås det at man i tillegg til bestått/ikke bestått innfører en karakterskala med langt færre trinn enn dagens skalaer, fortrinnsvis en skala med fem bestått-trinn A - E og F for stryk. [...] Departementet går inn for at alle fag der det skal gis karakterer, bygger på denne skalaen. Dette er også i tråd med ECTS (European Credit Transfer System)."

"ECTS Grading Scale" er bygget opp på følgende måte:

ECTS Grade	Definition	Percentage of successful students
A	excellent - outstanding performance with only minor errors	10%
B	very good - above the average standard with some errors	25%
C	good - generally sound work with a number of notable errors	30%
D	satisfactory - fair but with significant shortcomings	25%
E	sufficient - performance meets the minimum criteria	10%
FX	fail - some more work required before the credit can be awarded	-
F	fail - considerable further work required	-

Med utgangspunkt i kvalitetsreformens anbefalinger har Kollegiet vedtatt en felles karakterskala for NTNU.

Fra Kollegiets møte 28.02. 2001, K-sak 15/01 Endringer i eksamens- og gradsforskrifter:

"Eksamen/prøven bedømmes med bokstavkarakterer etter en skala fra A til F eller bestått/ikke bestått. Fullført/ikke fullført kan brukes der det ikke kreves vurdering. A er beste karakter og E er dårligste ståkarakter. Karaktertrinnene gis følgende betegnelse og generell beskrivelse:

Symbol	Betegnelse	Generell, kvalitativ beskrivelse av vurderingskriterier
A	Fremragende	Høyt kunnskapsnivå. God analytisk evne. Kan bruke kunnskapen selvstendig.
B	Meget god	God oversikt over kunnskapsfeltet. Kan bruke kunnskapen selvstendig.
C	God	Kan gjøre greie for de viktigste elementene i fagfeltet. Kan til en viss grad bruke kunnskapen selvstendig.
D	Brukbar	Oversikt over de viktigste kunnskapselementene mangler. Kan ikke bruke kunnskapen selvstendig.
E	Tilstrekkelig	Tilfredsstiller minimumskravene, men ikke mer. Kan ikke bruke kunnskapen selvstendig.
F	Ikke bestått	Mangler både detaljkunnskap og oversikt.

### B) Påstander og argumenter

Dette er et forsøk på å samle hovedargumentene i debatten omkring gjeninnføring av karakterer ved medisinstudiet i Trondheim. De er ikke å regne som bekreftede "sannheter". Påstander og argumenter er i hovedsak satt sammen fra allmøtereferater, MSUs utredning, debatter i Æsculap og Nocebo, diskusjoner i arbeidsgruppen og boka "Exit eksamen – Eller?" (Lauvås/Jakobsen)

#### Arbeidsmarked og ansettelsesprosess

- vil gi økt konkurransegrunnlag i et eventuelt vanskelig arbeidsmarked
- vil tilføre ansettelsesprosessen et objektivt element
- mangel på dokumentasjon av eget kunnskaps- og ferdighetsnivå i form av karakterer vil være en ulempe ved arbeidssøking
- vil belønne "flinke" studenter som oppnår gode eksamensresultater
- vil føre til overfokusering på eksamensprestasjoner på bekostning av mindre målbare egenskaper og ferdigheter

#### Dokumentasjon av kunnskap og kompetanse

- få brede eksamener vil gi få og uspesifikke karakterer
- lav "ytre validitet" ved en del eksamensformer vil gi dårlig bilde av kompetanse
- lav reliabilitet ved en del eksamensformer vil gi dårlig grunnlag for karaktersetting

#### Læring og lesing

- vil stimulere studentene til å lese mer
- vil føre til mer eksamensrettet, resultatfokusert lesing, og snevre inn mangfoldet av kunnskaper studentene tilegner seg – mer vekt på oppgaveløsning og "institusjonens interne forventninger"
- vil bryte med Trondheimsmodellens grunntanker om motivasjonsstyrt læring og om frihet til fordypning
- vil øke stressnivået i forkant av eksamen
- stimulerer til "overfladisk læring" og "memory dumping", og hemmer utvikling av "dypere forståelse/høyere kognitive egenskaper"
- vil føre til prioritering av de (overordnede) læringsmål som testes ved karaktergivende eksamener (dersom bare eksamener med høy reliabilitet skal vurderes med karakter)

#### Internasjonale studier og arbeid

- utenlandske universiteter krever i noen tilfeller å vite studentens karakterer før de godkjenner et utvekslingsopphold
- eksamenskarakter er en forutsetning for å få adgang til det internasjonale arbeidsmarked

#### Evaluering og tilbakemelding

- vil gi systematisk og differensiert vurdering av eksamensprestasjon
- dagens poengsystem gir en god differensiert tilbakemelding på eksamensprestasjon

#### Sosiale konsekvenser

- vil for noen stimulere konkurransementalitet og statusjag
- vil redusere motivasjon for studentpolitisk og –sosialt engasjement

#### Forventede krav fra studentene

- økt forutsigbarhet – langt tydeligere formulert forventet kunnskapsomfang og –nivå

### C) **Muligheter og krav ved innføring av karakterer**

#### Klage og forbedring av karakter:

En må regne med at flere studenter vil klage på karaktervurderingen ved skriftlig eksamen hvert år. Dette vil med dagens skriftlige eksamener medføre økt ressursbruk i sommerferien.

Studieforskriftene ved NTNU fastslår at dersom en student har bestått et emne men er misfornøyd med karakteren, kan han ta opp emnet en gang. Dersom karakteren er satt sammen av flere delvurderinger, må alle vurderinger tas opp. Beste resultat teller. En student kan ta opp igjen ett emne i året.

En må regne med at flere studenter vil forsøke å forbedre eksamenskarakterene. Dette vil føre til økt ressursbruk ved eksamen på grunn av flere kandidater.

Det må avklares om og hvordan studiet skal inndeles i emner, og hva karakterene skal omfatte. Skal hver eksamen ha egen karakter, eller hvert studieår? Det er ikke klart hvordan en skal kombinere dagens integrerte studium med universitetets forskrifter om emneinndeling, karakterer og eksamener. Dersom man skal gi en hovedkarakter for graden cand.med. må hver eksamen/eksamensemne vektas.

#### Reliabilitet og sensur:

Dagens skriftlige og muntlige eksamener har lav reliabilitet. Den er dog tilstrekkelig til å vurdere bestått/ikke bestått. Eksamensbesvarelser som ikke umiddelbart lar seg plassere vurderes grundigere. Ved innføring av karakterer må reliabiliteten ved eksamen økes. I tillegg vil det særlig ved skriftlig eksamen kreve grundigere vurdering av eksamensbesvarelser som ligger godt innenfor grensen av bestått. Dette vil medføre økt ressursbruk ved sensur.

### D) **Ulike standpunkter**

Studentene har på allmøtet 02.12.02 med overveldende flertall fremmet følgende uttalelse: *"Allmøtet ønsker å beholde dagens vurderingsordning, der eksamen vurderes med "Bestått / Ikke bestått", samt at studenten på anmodning kan få oppgitt sin poengsum. Allmøtet vil oppfordre Arbeidsgruppen for Eksamens- og Evalueringsformer til å arbeide med å finne alternative metoder for å gi studentene en mer kontinuerlig tilbakemelding og evaluering av kunnskaper, ferdigheter og holdninger."*

I debatten forut for avstemningen la studentene størst vekt på de mulige konsekvensene karakterer kan få for studiemiljø, læring og læringsmotivasjon – altså hva som kan skje under studiet.

Faglærerene i arbeidsutvalget, som sammen utgjør flertallet, er positivt innstilt til innføring av karakterer. De legger størst vekt på karakterenes betydning i forbindelse med arbeidssøking nasjonalt og særlig internasjonalt i de første årene etter endt utdanning.

## **Kilder**

Knox J.D.E.: "How to Use Modified Essay Questions" Medical Teacher, 1980; 2, 1: 20-24

"Evaluation Methods: A resource Manual" McMaster University, Program for Educational Development, 1987.

Harden R.M. and Gleeson F.A.: "Assesment of clinical competence using objective structured clinical examination (OSCE)" Medical Education, 1979, 13, 41-45.

Smee, S.: "Skill based assessment" BMJ, 2003, 326, 703-706

Haladyna , TM: "Developing and validating multiple-choice test items" (Lawrence Erlbaum, 1994)

Holsgrove og Elzubeir: "Imprecise terms in UK medical multiple-choice questions: what examiners think they mean." Medical education 1998, 32, 343-350.

Wass V, McGibbon D og Van der Vleuten C: "Composite undergraduate clinical examinations: how should the components be combines to maximaze reliability?" Medical education 2001, 35, 326-330.

Lauvås, P. og Jakobsen, A.: "Exit eksamen – eller? Former for summativ evaluering i høgre utdanning" (Cappellen 2002)

KUF: "Gjør din plikt – Krev din rett. Kvalitetsreform av høyere utdanning" (St. meld. nr. 27, 2000-2001)

Hoel, TL.: "Mapper som lærings- og vurderingform" (HF-fakultetet 22.11.2002)

Protokoll fra Kollegiets møte 28.02. 2001,

## **Vedlegg**

Sammenligning av ulike eksamensformer – en oversikt

En minimanual for forfattere av flervalgstester. (Børset, M)

Objective Structured Long Examination Record (OSLER)

Objective Structured Long Examination Record (OSLER)  
-Summary of satisfactory performance criteria

## En minimanual for forfattere av flervalgstester.

Ved Magne Børset

Dersom en flervalgstest (MCQ) skal fungere optimalt og ha den høye validiteten og reliabiliteten som denne eksamensformen kan ha, er det avgjørende at testen er konstruert med kyndighet. Eksaminatorer uten erfaring i å lage MCQ-spørsmål må derfor få en innføring i temaet før de får ansvar for å sette sammen nye eksamenssett. Denne minimanualen er tenkt som en startpakke for å øke bevisstheten om hva MCQ-tester er og hvilke fallgruver som må unngås dersom en test skal fungere.

### Ulike MCQ-formater.

Denne gjennomgangen baserer seg i hovedsak på "Developing and validating multiple-choice test items" av Thomas M. Haladyna (Lawrence Erlbaum, 1994), heretter referert til som "Haladyna".

Standard MCQ-format består av en stamme etterfulgt av ett korrekt og ett eller flere gale svar. Stammen kan være et spørsmål eller et utsagn som de ulike svaralternativene relaterer seg til. En variant er at svaralternativene kan ha gradert sannhetsinnhold (altså ingen direkte gale svar) og at studenten skal angi det beste eller det mest korrekte svaret.

Eksempel: Hvilket medikament er mest effektivt mot sterke smerter?  
a) paracetamol, b) naproksen, c) morfin, d) ketobemidon

Et annet format kan kalles sant/usant-formatet og består av en rekke utsagn som alle skal markeres som sanne eller usanne av testdeltakerne. Her er det altså ingen stamme som innleder hver testenhets.

Eksempel:

- a) paracetamol er et blodtrykksmedikament,
- b) naproksen er et antiinflammatorisk medikament
- c) morfin virker smertestillende
- d) ketobemidon er ikke registret til bruk i Norge

Sant/usant-formatet skal angivelig være mindre reliabelt enn standardformatet for det er vanskelig å sikre en lik fordeling av sanne og gale påstander. Utsagnene vil måtte formes som påstander av typen "er" og "er ikke" og også mellom disse to påstandstypene bør det være en lik fordeling av gale og riktige utsagn. Noen studenter har større tilbøyelighet enn andre til å angi svar som sanne stilt overfor utsagn de ikke vet sannhetsgehalten av. Disse studentene vil skåre høyere enn slumpen skulle tilsi dersom det er en overvekt av sanne utsagn i testen. Disse svakhetene ved dette formatet har gjort at Haladyna ikke anbefaler dette formatet i sin opprinnelige form.

Et bedre format i følge Haladyna er multippel-sant/usannt-formatet. Her har man igjen en stamme som innleder hver testenhets og som består av et spørsmål eller en påstand. Stammen følges av en rekke alternativer og *alle* skal markeres som sanne eller usanne i forhold til stammen.

Eksempel: *Byene nedenfor er hovedsteder. Sett A ved sant svar og B ved usant svar (stamme)*

- a) Trondheim ..., b) Stockholm..., c) København..., d) Manchester...,
- d) Paris....



I slekt med det foregående formatet og med standardformatet er multippel-responsformatet. Det atskiller seg fra standard MCQ-formatet bare ved at det i dette tilfellet kan være flere enn ett riktig svaralternativ. Her skal studentene altså ikke angi om alternativene er sanne eller usanne i forhold til stammen, men velge ett eller flere riktige alternativer.

Eksempel: *Hvilke(n) av byene nedenfor er hovedsteder?*

*a) Trondheim, b) Stockholm, c) København, d) Manchester, e) Paris.*

Forskjellen mellom disse formatene kan synes små, men Haladyna anbefaler ikke multippel-responsformatet, såvidt jeg skjønner fordi det ikke er så gjennomprøvd og følgelig ikke har egenskaper som er så godt dokumentert som standardformatet og multippel-sant/usant-formatet.

Match-formatet er en variant av standardformatet der svaralternativene listes opp først og skal pares med stamme-utsagnene som listes etterpå.

Eksempel: *For hvert utsagn som listes opp nedenfor (1-4) skal du velge ett svaralternativ (a-d):*

*a) digitoksin, b) naproksen, c) penicillin, d) fenytoin.*

*1: er et ikke-steroid antiinflammatorisk middel,*

*2: brukes som hjertemedisin,*

*3: er et antibakterielt middel,*

*4: forebygger epileptiske anfall.*

Match-formatet er økonomisk på den måten at mye kunnskap kan testes på kort tid og liten plass. Alle de gale svaralternativene er rett og slett utelatt.

### **Noen gode råd for testforfattere.**

#### **1) Unngå utenomsnakk!**

En tests reliabilitet og validitet øker med antall delspørsmål. Lange innledninger som ikke er helt nødvendige for å oppfatte spørsmålet (stammen) bør derfor unngås. De begrenser nemlig bare den tiden som studentene kan bruke til selve oppgaveløsningen og dermed også antall mulige delspørsmål.

#### **2) Unngå lurespørsmål!**

En testforfatter kan være fristet til å legge inn små feller for studentene, men det er som regel ikke evnen å styre unna slike feller man egentlig ønsker å teste. Testenhetene bør utformes slik at de er entydige og lette å besvare for studenter som har den nødvendige kunnskapen.

#### **3) Unngå å gi ledetråder!**

I testenheter med flere ledd kan man uforvarende komme til å gi ledetråder som avslører riktig svar i et annet ledd av oppgaven.

Eksempel:

I) Hvilket av disse cytokinene stimulerer til en T-hjelper-1-respons?

a) IL-6, b) IL-10, c) IL-4, d) IFN $\gamma$ .

II) Hvilket av disse cytokinene stimulerer til en T-hjelper-2-respons?

a) IL-4, b) IFN $\gamma$ , c) IL-2, d) IL-12

En student som vet riktig svar på del I) (IFN $\gamma$ ), vil eliminere dette som mulig svar på del II). Siden IL-4 er det eneste cytokinet ved siden av IFN $\gamma$  som opptrer i begge delspørsmålene, vil studenten vite at IL-4 i hvert fall ikke er riktig svar på del I). Det er derfor mer smart å velge IL-4 som svar på del II) enn de andre alternativene (hvilket også er riktig).

4) Unngå testenheter der svaret er basert på holdninger eller meninger!

Eksempel:

Hva er den viktigste funksjonen til pankreas?

Hva er det største helseproblemet blant innvandrere?

Selv om svaret på slike spørsmål kan synes åpenbart for testforfatteren, vil det ofte kunne bli gjenstand for begrunnet diskusjon.

5) Dersom stammen i standard MCQ-format utformes som et ikke-fullført utsagn, må det utelatte leddet stå til slutt!

Eksempel:

I) Surhetsgraden i ventrikkelinholdet vil påvirke blodkonsentrasjonen av:

a) insulin, b) adrenalin, c) gastrin, d) albumin

II) Konsentrasjonen av ----- i blodet vil bli påvirket av surhetsgraden i ventrikkelinholdet.

a) insulin, b) adrenalin, c) gastrin, d) albumin

Det er mye lettere å lese eksempel I) enn eksempel II), selv om innholdet er det samme.

6) Unngå negasjoner i stammen til en testenheter!

Eksempel:

Hvilken av disse aminosyrene er ikke basisk?

a) lysin, b) histidin, c) valin, d) arginin

I farten kan det være lett å overse ord av typen "ikke" eller "unntatt". Hvis man likevel velger å bruke "ikke" etc., bør slike ord markeres ved understreking eller feite typer.

7) Velg gode distraktorer! (distraktor = betegnelse på gale svar)

En distraktor som er så usannsynlig at den unngås også av testdeltakere uten kunnskaper på området, er verdiløs. Kunsten er altså å finne sannsynlige distraktorer og det kan ofte være vanskelig å finne mange gode alternativer. Det later til å være litt uenighet om hvilket antall distraktorer som er optimalt, men Haladyna hevder at det er bedre å begrense antallet valgmuligheter til 3 eller t.o.m. 2, i stedet for å finne på svar som er lette å gjennomskue.

8) Alle valgmulighetene bør ha samme gruppetilhørighet!

Eksempel:

Hvilket hormon er det mannlige kjønnshormonet?

a) østrogen, b) kortisol, c) insulin, d) testosteron

I dette eksempelet er distraktoren insulin det eneste hormonet som ikke er et steroid. Det kan oppfattes som et tegn på at insulin er et galt svar.

9) Velg distraktorer basert på vanlige feil og misforståelser hos studentene!

Eksempel:

Dolor, calor, rubor, tumor og functio laesa er et tegn på (ett riktig svar):

a) infeksjon, b) kontusjon, c) inflammasjon

Tidlig i studiet blander ofte studentene begrepene a) og c) ovenfor.

10) Unngå spøkefulle distraktorer!

Humor er bra, men på en eksamen virker humor i distraktorene bare forstyrrende.

### **Muligheten for å teste høyere tankeprosesser med MCQ**

En bekymring hos skeptikerne til MCQ har vært at slike tester i liten grad skulle egne seg til å teste avanserte tankeprosesser som resonnement og problemløsning. Disse synspunktene har i følge Haladyna, bare begrenset gyldighet. Det er mulig å teste andre egenskaper enn triviell faktakunnskap med MCQ, men det krever at testene designes med dette for øye.

For å nærme seg dette viktige problemet opererer Haladyna med fire innholdskategorier i enhver læringsprosess: 1) fakta, 2) begrep, 3) prinsipper, 4) prosedyrer. (1, 2 og 4 bør være intuitivt forståelige, 3 (prinsipper) er utsagn om forhold, som regel mellom to eller flere begrep, for eksempel: "Når du trener en muskel, vil den vokse"). Alt som læres kan altså inndeles i disse kategoriene.

Disse fire innholdskategoriene kan man behandle med fem ulike mentale operasjoner: 1) gjenkalling, 2) definering, 3) predikering, 4) evaluering og 5) problemløsning. Teoretisk kan man da tenke seg 20 (4x5) ulike kombinasjoner av kunnskapsinnhold og mentale operasjoner, men ikke alle er like aktuelle.

MCQ egner seg særlig godt til å gjenkalle fakta, men også mange av de andre 19 mulige kognitive prosessene lar seg teste på denne måten.

Eksempel:

Hvilken av situasjonene nedenfor gir en sann beskrivelse av forholdet mellom et vevs form og endrete funksjonelle krav?

- a) Plateepitelet i huden fortykkes på steder med stor slitasje.
- b) Cardiac output er størst i diastolen.
- c) Cardiac output er størst i systolen.
- d) Redusert kreatinin clearance fører til en forlengelse av nedad- og oppadstigende del av Henles sløyfe.

Stammen er her en definisjon av et prinsipp gitt i valgmulighet a). Distraktorene er enten usanne [b) og d)] eller sanne, men irrelevante i forhold til definisjonen [c)].

Eksempel:

Hva er den mest sannsynlige følgen av langvarig alkoholisme hos en mann?

- a) nyresykdom
- b) leversykdom
- c) hjertesykdom
- d) kreft

Den mentale operasjonen i dette eksemplet er prediksjon på grunnlag av et prinsipp: alkoholisme fører til leversykdom.

Eksempel:

Hva er den mest effektive behandlingen mot en halsinfeksjon med betahemolytiske streptokokker?

- a) amoxicillin 250 mg x 3 p.o. i 7 dager,
- b) trimetoprim 150 mg x 2 p.o. i 7 – 10 dager,
- c) fenoksymetylpencillin 660 mg p.o. 1+1+2 i 10 dager,
- d) Observasjon og eventuelt sykehusinnleggelse ved alvorlige allmennsymptomer.

I dette eksempelet må studentene evaluere fire ulike prosedyrer.

Haladyna ser på problemløsning som den mest krevende av de fem ulike mentale operasjonene. Begrepet problemløsning er ikke entydig definert, men er ofte en sammensatt prosess som involverer en serie av flere av de andre mentale operasjonene. Derfor er det svært vanskelig å skrive gode MCQ-oppgaver som tester problemløsning. Den beste måten å gjøre det på er antakelig å bruke sammensatte testenheter (test item sets).

Eksempel:

En 40 år gammel mann har arvelig hemokromatose. Tilstanden skyldes som regel en punktmutasjon i et bestemt gen og man må være homozygot for forandringen for å få klinisk sykdom. Prevalensen av heterozygote bærere i befolkningen er ca. 10%

- 1) Hva slags arvegang er dette et eksempel på?
  - a) dominant
  - b) recessiv
  - c) kjønnsbundet

Pasienten er gift og har fire barn. Han har én bror. Gå ut fra at pasientens foreldre er heterozygote bærere av sykdomsgenet.

- 2) Hva er sannsynligheten for at pasientens bror har sykdommen?
  - a) 25%
  - b) 50%
  - c) 75%
  - d) 100%
- 3) Hva er sannsynligheten for hvert av barna for at de har eller vil få sykdommen?
  - a) Samme sannsynlighet som befolkningen ellers.
  - b) 5%
  - c) 25%
  - d) 50%
  - e) 75%

(riktig svar er understreket)

Dette eksemplet illustrerer for øvrig et problem ved slike sammensatte testenheter, nemlig at ett ledd kan gi ledetråder for svaret i et annet ledd. Ved at svaralternativene under ledd 2) i gjennomsnitt har en høyere tallverdi enn alternativene i ledd 3), vil nok mange som ikke har kunnskaper til å velge riktig svar, anta at broren har høyere sannsynlighet for å få sykdom enn barna. Derfor vil de velge riktig svar hyppigere enn det slumpen skulle tilsi. For å unngå dette burde jeg gitt den samme rekken av svaralternativer under begge ledd av oppgaven, for eksempel:

- a) Samme sannsynlighet som befolkningen ellers.
- b) 5%
- c) 25%
- d) 50%
- e) 75%
- f) 100%

### **Kvalitetskontroll av MCQ-tester**

En formell kvalitetskontroll av utkast til MCQ-tester er viktig.

Eksamenskommissjonen bør bruke mer tid i forkant av eksamen enn det som har vært vanlig ved MEQ-eksamener. Ordningen med å la eldre studenter gjennomgå eksamensutkastet er også viktig.

Kontrollen bør skje på flere plan:

- 1) Språkvask: Teksten bør naturligvis være grammatisk korrekt, men like viktig er det å vurdere om noen av begrepene som brukes kan være tvetydige eller mangetydige. Man må spesielt være oppmerksom på bruken av upresise uttrykk som "ofte", "sjelden", "kan", "vanligvis", "(u)vanlig", "assosiert med", "karakteristisk" og hyppig". Slike ord har ingen allment gyldig definisjon og det viser seg at ulike personer oppfatter innholdet av slike ord forbausende forskjellig (Holsgrove og Elzubeir: Medical education 1998, 32, 343-350). Man bør derfor unngå at slike ord settes i posisjoner der de kan få betydning for de valg som testdeltakerne gjør.
- 2) Er prinsippene for god oppbygging av testenheter for MCQ-tester fulgt? De rådene som er gitt ovenfor (se Noen gode råd for testforfattere) kan brukes som sjekkpunkter.
- 3) Svarkontroll: En helt sentral kvalitet ved en MCQ-test er at de forutsatt korrekte svarene og *bare* disse, skal være sant riktige. Det er ikke tilstrekkelig at testforfatteren har gjort dette så godt han eller hun kan. Andre personer som forutsettes å kjenne fagområdet, må inn for å kontrollere testenheter på dette punktet.
- 4) Innholdskontroll: Som ved andre eksamener, er det viktig at det som testes hører til innenfor læringsmålene.
- 5) Det bør også være en kontroll i etterkant av eksamen. Fakultetet bør ta sikte på å etablere en database med MCQ-spørsmål slik at man slipper å lage hele eksamen fra grunnen hver gang. Før nye testenheter slippes inn i databasen, bør man vurdere om de fungerte tilfredsstillende. Testenheter som ikke hadde noen diskriminativ funksjon, bør utelates. Diskriminatorer som ingen av eksaminandene valgte, bør vurderes utelatt eller erstattet. En mer krevende type av etterkontroll er å la en gruppe studenter forklare sine valg for om mulig å identifisere feilvalg som ikke skyldtes manglende kunnskap, men misforståelse og tvetydigheter.