

Eksamenskvalitet og innføring av gradert karakterskala ved profesjonsstudiet i medisin

**Rapport fra arbeidsgruppen til styringsgruppen for Oslo 2014,
februar 2016**

Innhold

Bakgrunn og mandat.....	3
Sammendrag av anbefalinger	4
1. Eksamen og kriterier for kvalitet og standardsetting.....	6
1.1 Evaluering av studenter – grunnleggende begreper.....	6
1.2 Kriterier for vurdering av eksamensformer	6
1.3 Standardsetting	6
1.4 Rammene for eksamens- og sensurarbeidet	7
1.5 Kvalitetskriterier i eksamensarbeidet	8
1.6 Kriterier for kvalitetssikring ved standardsetting.....	8
1.7 Arbeidsgruppens anbefalinger:	9
2. Eksamensformer.....	11
2.1 Skriftlige eksamener med essayoppgaver eller kortsvars spørsmål	11
2.2 Databaserte, skriftlige eksamener	11
2.3 Muntlig eksamen.....	12
2.4 Objektiv strukturert klinisk eksamen - OSCE.....	12
2.5 Prosjektoppgaven og andre skriftlige innleveringer	13
2.6. Arbeidsgruppens anbefalinger.....	13
3. Anbefalinger om hvilke eksamener som skal vurderes etter graderte karakterer	14
3.1 Forhold som taler for bruk av bestått/ikke-bestått	14
3.2 Arbeidsgruppens anbefalinger:	14
4. Vitnemål	16
4.1. Hovedkarakter.....	16
4.2. Arbeidsgruppens anbefalinger.....	16
5. Implementering.....	17
5.1 Tidsplan for implementering av graderte karakterer	17
5.2. Arbeidsgruppens anbefalinger.....	17
Kilder	18

Bakgrunn og mandat

Fakultetsstyret vedtok 17. september 2013 å innføre en gradert karakterskala (A-F) i profesjonsstudiet i medisin, men kun fra og med de studentkull som tas opp på studiet fra høsten 2014. Gradert karakterskala skal implementeres etter at det er gjennomført en kvalitetssikring av eksamensformer og gjort en utredning av hvilke emner som skal evalueres med bruk av bokstavkarakterer. Implementeringen skal skje senest i 2017. Det åpnes for at man i ny ordning kan bruke bestått/ikke-bestått for enkelte emner.

Styringsgruppa for Oslo 2014 vedtok i juni 2015 å opprette en arbeidsgruppe med følgende mandat:

- foreslå kriterier for å operasjonalisere kravet om kvalitetssikring fra fakultetsstyret
- gi anbefaling om hvilke eksamener som skal vurderes etter gradert karakterskala
- levere rapport

Arbeidsgruppen skal være ferdig med sitt arbeid innen utgangen av februar 2016.

Arbeidsgruppen har bestått av følgende medlemmer:

- Kristin Wium, administrativ prosjektleder Oslo 2014, gruppens leder
- Kristian Bjørø, professor, KLINMED
- Trond Buanes, professor, KLINMED
- Petter Gjersvik, førsteamanuensis, KLINMED
- Inger Nina Farstad, professor KLINMED
- Jan Frich, professor/faglig prosjektleder Oslo 2014, HELSAM,
- Erik Dissen, IMB
- Per Grøttum, Helsevitenskapelig utdanningscenter
- Liv Bjøntegaard Finess, Studieseksjonen
- Anna Werenskjold, medisinstudent
- Johannes Slørdahl, medisinstudent
- Jakob Seth Fiskå, medisinstudent

Arbeidsgruppen har hatt fem møter i perioden august 2015 til januar 2016.

Sammendrag av anbefalinger

- *Bedre kvalitetssikring kan oppnås ved systematisk organisering av eksamensarbeidet, kontinuitet i eksamenskommissjonene, nitid teamarbeid med utforming av oppgaver - inkludert detaljerte sensurveiledninger, psykometrisk analyse av resultater og tilbakemeldinger til sensorer og oppgaveforfattere. Ved skriftlig eksamen bør foreleserne og de som har laget oppgavene delta i sensurering av besvarelser*
- *Bedre kvalitet på eksamen kan oppnås ved at oppgavene tester sentral kunnskap - inkludert grunnleggende kunnskaper, evne til resonnering og anvendelse av kunnskap/ferdigheter, at eksamen tester bredt fra mange ulike områder av faget (høy utvalgsvaliditet), at antall spørsmål er høyt, at spørsmålene er entydige og har kjente og velkalibrerte egenskaper.*
- *Etter at ekstern sensor og faglærere/eksamenskommissjon er ferdig med å sensurere besvarelsene avholdes et sensurmøte der alle som hovedregel skal være tilstede. Ved store avvik mellom sensors og faglærers karaktersetting bør besvarelsen vurderes av en tredje person. Spesiell grundighet bør utvises med besvarelser der det er tvil om kandidaten har bestått.*
- *Det faglige nivåkravet til bestått opprettholdes, men man bør vurdere å bruke en mindre del av eksamen på å avklare om en student har bestått (sertifiseringsfunksjonen) for å heve kvaliteten på karaktergivingen (A-E). Arbeidsgruppen foreslår at 55 % av maksimal oppnåelig score på eksamen brukes til sertifiseringsoppgaven og 45 % av maksimal oppnåelig score brukes til diskrimineringsfunksjonen. Dette vil kreve at en andel av spørsmålene får en høyere vanskelighetsgrad.*
- *Eksamenskommissjonene må tydeliggjøre og nedfelle hva som kjennetegner studenter på begge sider av strykgrensen og hva som skiller de ulike karakterer (A-E) fra hverandre.*
- *Innføring av graderte karakterer forutsetter at det nødvendige oppgavegrunnlaget er på plass, er stabilt og at de prinsipielle valgene mellom sertifisering og gradering er truffet.*
- *Ved skriftlige eksamener må man tilstrebe at oppgavesettene har de samme egenskapene med hensyn til vanskelighetsgrad og utvalgsvaliditet slik at prestasjonene måles med lik målestokk fra kull til kull. Økt kvalitetssikring kan oppnås ved gjenbruk av spørsmål og gradvis raffinering av oppgaver og oppgavesett, slik fakultetet nå legger opp til ved digitale eksamener.*
- *Ved bruk av reelle pasienter må man vurdere om pasienten er egnet for å delta i en eksamenssituasjon. Oppgaver/kasus bør skriftliggjøres. Vurderings- og skåringsskjemaer som brukes bør gjennomgåes med alle som skal være eksaminator/sensor, og man må utvikle delpunkter som muliggjør kriteriebasert karaktersetting.*

- *«Veileder for eksaminatorer og sensorer», «Rettleader til leder og medlemmer av eksamenskommisjoner for modifiserte OSCE eksamener» og «Rettleader for gjennomføring av modifisert OSCE» bør revideres og tilpasses eksamen i ny studieplan. Det må nedfelles prinsipper for hvordan eksamenskommisjonene skal arbeide med å lage oppgavesett og planlegge praktisk gjennomføring av eksamen, hvor rutiner for å utforme oppgaver og oppgavesett, utvikling av skåringsmaler, rutiner for tilbakemelding fra studenter og etterarbeid formuleres. Oslo 2014-prosjektet tildeles ansvar for å arbeide med dette i samarbeid med Studieseksjonen og andre relevante parter.*
- *Det foreslås bruk av bestått/ikke-bestått og graderte karakterer i henhold til tabell 1 (i denne rapporten).*
- *Informasjon om karakterer i de enkelte emner skal fremgå av i vitnemålet.*
- *Det gjøres ikke endringer i opptakskravet til doktorgradsprogrammet etter innføring av graderte karakterer.*
- *Det implementeres graderte karakterer fra studiets start (modul 1) fra og med kull H-16 (høsten 2016); fra og med modul 2 for kull V-16 (høst 2017); fra og med modul 4 for kull H-14 (høst 2017); fra og med modul 3 for kull H-15 (vår 2018) under forutsetning av at kvalitetssikringen anses som innfridd.*

1. Eksamen og kriterier for kvalitet og standardsetting

1.1 Evaluering av studenter – grunnleggende begreper

Evaluering av studenters kompetanse eller læringsutbytte har ulike funksjoner (1, 2):

- Kontroll av om en students kompetanse oppfyller bestemte minstekrav (*sertifisering*)
- Tilbakemelding til studenter om læringsbehov eller deres faktiske kompetansenivå eller læringsutbytte (blant annet som *gradering*)
- Tilbakemelding til undervisningsmiljøene om studenters kompetansenivå eller læringsutbytte
- Motivasjon for læring
- Informasjon til fremtidige arbeidsgivere

Det er vanlig å skille mellom *formativ* og *summativ* evaluering (2). Formativ evaluering har til hensikt å fremme læring, mens summativ evaluering har til hensikt å avklare om en student har oppnådd et bestemt mål, kompetansenivå eller læringsutbytte.

1.2 Kriterier for vurdering av eksamensformer

I litteraturen er det formulert ulike kriterier for å vurdere eksamensformer innen medisin (4). Kort oppsummert kan vi si at eksamener bør være:

- valide (tester det de er ment å teste)
- reliable (pålitelige)
- relevante
- egnet til å gi studentene læringseffekter (har en formativ effekt)
- praktisk gjennomførbare med hensyn til bruk av tid og ressurser
- akseptabel for involverte parter (inkludert eventuelle pasienter)

Vurderinger og beslutninger rundt bruken av ulike eksamensformer i et studium vil innebære avveininger mellom de ulike punktene over. For eksempel vil en eksamen med høy validitet kunne ha lav reliabilitet og vice versa. Videre kan det være ønskelig med en omfattende eksamen av hensyn til reliabiliteten, samtidig som man må ta i betraktning tidsbruk og praktisk gjennomførbarhet.

I og med at alle evalueringsmetoder har styrker og svakheter bør eksamensformen som benyttes tilpasses det aktuelle emnet. I løpet av et medisinsk studium bør man i lys av dette ha en portefølje av ulike eksamensformer som tester ulik kompetanse slik som faktakunnskap, kliniske ferdigheter, teoretiske ferdigheter og klinisk resonnering.

1.3 Standardsetting

Prosessen med å etablere standarder for karakterer ved en eksamen, inkludert skillet mellom bestått/ikke-bestått, kalles standardsetting. Metoder for standardsetting kategoriseres som enten *relative* (normbaserte) eller *absolutte* (kriteriebaserte). Normbaserte metoder tar utgangspunkt i prestasjonene til en definert gruppe og setter en grensepoengsum for bestått basert på denne gruppens prestasjon. Beståttgrensen og graderte karaktergrenser vil da

varierte fra kull til kull. Ved kriteriebaserte metoder etableres grensen for bestått før eksamen administreres, og studentene blir vurdert på grunnlag av kvaliteten på besvarelsen de leverer opp mot forventet læringsutbytte (8).

Arbeidsgruppen har lagt til grunn at karaktergivingen i Norge skal være kriteriebasert (jf. Kunnskapsdepartementet presiseringer i skriv 14.12.2011). Vurderingsuttrykket ved eksamen skal følge europeisk standard og kan enten være bestått/ikke-bestått eller følge en gradert skala med fem trinn fra A til E for bestått og F for ikke-bestått.

De generelle, kvalitative beskrivelsene for graderte karakterer på bokmål er som følger:

<i>Symbol</i>	<i>Betegnelse</i>	<i>Generell, ikke fagspesifikk beskrivelse av vurderingskriterier</i>
A	Fremragende	Fremragende prestasjon som klart utmerker seg. Kandidaten viser svært god vurderingsevne og stor grad av selvstendighet.
B	Meget god	Meget god prestasjon. Kandidaten viser meget god vurderingsevne og selvstendighet.
C	God	Jevnt god prestasjon som er tilfredsstillende på de fleste områder. Kandidaten viser god vurderingsevne og selvstendighet på de viktigste områdene.
D	Brukbar	En akseptabel prestasjon med noen vesentlige mangler. Kandidaten viser en viss grad av vurderingsevne og selvstendighet.
E	Tilstrekkelig	Prestasjonen tilfredsstiller minimumskravene, men heller ikke mer. Kandidaten viser liten vurderingsevne og selvstendighet.
F	Ikke bestått	Prestasjon som ikke tilfredsstiller de faglige minimumskravene. Kandidaten viser både manglende vurderingsevne og selvstendighet.

Av flere ulike grunner er det vanskelig å uttrykke kvalitetsbaserte standarder i detaljerte verbale beskrivelser, kategorier eller lister. Det eksisterer ikke en uavhengig betydning av ord og karakterbeskrivelser. Det er derfor helt avgjørende at fagpersoner lager og sensurerer eksamen, at det er etablert en felles forståelse for konseptet standardsetting og hvordan standarder settes, og at det eksisterer et system for kompetansespredning til alle involverte (10).

1.4 Rammene for eksamens- og sensurarbeidet

Eksamenskommisjoner med en stabil sammensetning er viktig for kvalitetssikringen av eksamen. Ved profesjonsstudiet i medisin oppnevnes eksamenskommisjonene for 3 år av gangen i alle emner (moduler). Det foreligger i dag retningslinjer og veiledere for eksamenskommisjonens arbeid for OSCE og klinisk muntlig eksamener. Arbeidsgruppen mener disse bør revideres og tilpasses ny studieplan, og at det i tillegg bør lages retningslinjer for kvalitetssikring av eksamenskommisjonenes arbeid generelt.

Utarbeidelse av sensorveiledninger, og bruk av eksterne sensorer på alle eksamener, gir gode muligheter for kvalitetssikring. Det er viktig at eksterne sensorer får oppgavene i forkant og

at vektingen av de ulike oppgavene fremgår av sensorveiledningen. Økt bruk av skåringsmaler der dette er hensiktsmessig anbefales, og også en simulering og samkjøring av skåringspraksis i forkant av eksamen der dette er mulig.

De fleste eksamener krever mye administrativ støtte for å kunne gjennomføres, og tydelige rollebeskrivelser for gjennomføring av eksamen er viktig. Eksamen må planlegges godt i alle ledd, og det er viktig at kommunikasjonen mellom de ulike stedene fungerer optimalt.

Alle skriftlige og stasjonsbaserte eksamener gjennomleses av 1-2 studenter på høyere kull enn de som skal avlegge eksamenen. Studentens oppgave er først og fremst å vurdere om oppgavene er klart og entydig formulert, og om de virker rimelige i forhold til semesterets undervisning og læringsmål.

1.5 Kvalitetskriterier i eksamensarbeidet

Utforming av eksamensoppgaver og kasus danner grunnlaget for mange av de andre prosessene i eksamensarbeidet, og det er ekstra viktig at vi retter oppmerksomhet mot dette arbeidet for å opprettholde og forbedre våre eksamener. Læringsutbyttebeskrivelsene som er laget i alle emner skal tjene flere formål, men i en eksamenskontekst skal de definere innholdet i emnet og sette rammene for eksamen uten å bli for detaljstyrende. Det er viktig å unngå at læringsutbyttene blir enkeltkomponenter frakoblet helheten (9).

Det sentrale validitetskriteriet er at eksamensspørsmålene samsvarer med læringsutbytter både i faglig bredde, dybde og betydning. Det er viktig at det etableres en omforent forståelse av dette. Bruk av fordelingsnormer (blue-printing) for å utforme spørsmål som er valide med hensyn til læringsutbyttebeskrivelsen og et gitt faglig nivå, kan være nyttig. Spørsmålenes vanskelighetsgrad (klassifisering) bør vurderes i lys av eksamensresultatene, men man bør bestrebe seg på å lage så gode og treffsikre oppgaver at dette blir en ekstra kvalitetssjekk av om man har truffet riktig nivå heller enn en rekalkibrering som tenderer mot en normbasert sensurering.

Reliabilitet handler om å få stabile vurderinger av klassifikasjonsprosessen eller av sensurarbeidet. Et sentralt virkemiddel er brede fagfellevurderinger og objektivt anvendbare kriterier. Vurderinger av en eksamens egenskaper må skje i brede faglige fora og presise sensurveiledninger må utvikles.

1.6 Kriterier for kvalitetssikring ved standardsetting

Sertifisering og gradering konkurrerer på sett og vis om plassen i design og bedømmelse av en eksamen. Vår gamle ordning med bestått/ikke-bestått vektlegger sertifisering, og alle spørsmålene kan i prinsippet vies til denne oppgaven. Det har ingen hensikt å skille dyktige og mindre dyktige studenter fra hverandre, bare å finne ut om de er kvalifiserte til å praktisere medisin. Kravet til bestått hos oss har til nå vært 65 % av maksimalt oppnåelig score, hvilket indikerer sertifiseringsoppgavens store plass (en stor del av spørsmålene/oppgavene brukes til sertifiseringsformål).

Det er en sammenheng mellom antall spørsmål/eksamensomfang og påliteligheten i karakterfastsettelsen, og ved implementering av graderte karakterer er det nødvendig å ha eksamenssett som spenner ut hele karakterspekteret. En økt pålitelighet i fastsettelsen av de

graderte karakterene kan oppnås ved å bruke en større del av eksamen til diskrimineringsfunksjonen og en mindre del av eksamen til å ivareta sertifiseringsoppgaven f.eks. i størrelsesforholdet 55 % til sertifiseringsoppgaven og 45 % til diskrimineringsoppgaven. Dette vil kreve at eksamenens vanskelighetsgrad økes for at man ikke skal slippe igjennom studenter som er ukvalifiserte etter dagens norm. Det faglige nivået som kreves for å bestå må samsvare med det som praktiseres i dag, men man vil bruke en noe mindre del av eksamen på å avklare om studenten oppfyller minstekravet til faglig nivå. En justering av vanskelighetsgrad er ikke triviell, og den kan best gjennomføres dersom man kan bruke utprøvde, men gjenbrukbare spørsmål til å simulere ulike eksamenskonstellasjoner. Dette taler for bred utvikling av gjenbruksbanker. Et annet tiltak for å bedre reliabiliteten til de graderte karakterene vil være å lage eksamener med et størst mulig antall spørsmål, altså å utvide eksamensomfanget.

Noen terskler er mer avgjørende enn andre og eksamenskommissjonene bør tydeliggjøre og nedfelle hva som kjennetegner studenter på begge sider av strykgrensen. Videre bør man formulere på forhånd hva som skiller ulike karakterer (A-E) fra hverandre.

Det har vært et diskusjonstema blant studentene at de gjerne skulle hatt samme bestått-grense ved alle de fire studiestedene. Man kan forstå dette ut fra et ønske om å bli bedømt på like vilkår, men hvilken bestått-grense man velger sier mer om en eksamens oppbygning enn om vanskelighetsgrad. Nasjonal eksamen er sannsynligvis den eneste måte å etablere en bedømmelse som muliggjør sammenlikning av prestasjoner på tvers av studiesteder.

Innføring av graderte karakterer forutsetter at det nødvendige oppgavegrunnlaget er på plass, er stabilt og at de prinsipielle valgene mellom sertifisering og gradering er truffet.

1.7 Arbeidsgruppens anbefalinger:

- *Bedre kvalitetssikring kan oppnås ved systematisk organisering av eksamensarbeidet, kontinuitet i eksamenskommissjonene, nitid teamarbeid med utforming av oppgaver - inkludert detaljerte sensurveiledninger, psykometrisk analyse av resultater og tilbakemeldinger til sensorer og oppgaveforfattere. Ved skriftlig eksamen bør foreleserne og de som har laget oppgavene delta i sensurering av besvarelser*
- *Bedre kvalitet på eksamen kan oppnås ved at oppgavene tester sentral kunnskap - inkludert grunnleggende kunnskaper, evne til resonnering og anvendelse av kunnskap/ferdigheter, at eksamen tester bredt fra mange ulike områder av faget (høy utvalgsvaliditet), at antall spørsmål er høyt, at spørsmålene er entydige og har kjente og velkalibrerte egenskaper.*
- *Etter at ekstern sensor og faglærere/eksamenskommissjon er ferdig med å sensurere besvarelsene avholdes et sensurmøte der alle som hovedregel skal være tilstede. Ved store avvik mellom sensors og faglærers karaktersetting bør besvarelsen vurderes av en tredje person. Spesiell grundighet bør utvises med besvarelser der det er tvil om kandidaten har bestått.*

- *Det faglige nivåkravet til bestått opprettholdes, men man bør vurdere å bruke en mindre del av eksamen på å avklare om en student har bestått (sertifiseringsfunksjonen) for å heve kvaliteten på karaktergivningen (A-E). Arbeidsgruppen foreslår at 55 % av maksimal oppnåelig score på eksamen brukes til sertifiseringsoppgaven og 45 % av maksimal oppnåelig score brukes til diskrimineringsfunksjonen. Dette vil kreve at en andel av spørsmålene får en høyere vanskelighetsgrad.*
- *Eksamenskommisjonene må tydeliggjøre og nedfelle hva som kjennetegner studenter på begge sider av strykgrensen og hva som skiller de ulike karakterer (A-E) fra hverandre.*
- *Innføring av graderte karakterer forutsetter at det nødvendige oppgavegrunlaget er på plass, er stabilt og at de prinsipielle valgene mellom sertifisering og gradering er truffet.*

2. Eksamensformer

I profesjonsstudiet i medisin er ulike eksamensformer i bruk:

- Skriftlige eksamener med essayoppgaver eller kortsvarspørsmål
- Databaserte, skriftlige eksamener
- Muntlige og kliniske eksamener
- Objektiv, strukturert klinisk eksamen (OSCE)
- Prosjektoppgave og andre skriftlige innleveringer

Eksamensformer vil i varierende grad være egnet til å teste ulike nivåer av kompetanse (2, 3). Når man skal vurdere om en eksamensform eller en kombinasjon av ulike eksamensformer er egnet, må man ta i betraktning hvilken kompetanse man ønsker å evaluere. Man må også vurdere styrker og svakheter ved de ulike eksamensformatene.

2.1 *Skriftlige eksamener med essayoppgaver eller kortsvarspørsmål*

Skriftlig eksamen med essayoppgaver eller kortsvarspørsmål er den klassiske eksamen med relativt omfattende og åpne oppgaver. Fordeler med denne eksamensformen er at man kan teste både faktakunnskap og resonneringsevne - også innenfor komplekse problemstillinger uten entydige svar. Videre testes kandidatens evne til å drøfte, organisere tanker logisk, og formuleringsevne. Eksamensformen kan således ha høy validitet og relevans.

Ulemper kan være at oppgavene ofte er så omfattende at det bare er rom for et fåtall oppgaver i en eksamen, hvilket kan gi et problem med utvalgsvaliditeten. Videre vil sensor måtte basere seg på en skjønnsmessig vurdering ved poenggiving, noe som kan føre til lav reliabilitet, og den åpne og omfattende formen kan gjøre det vanskelig å lage presise sensurveiledninger. Reliabiliteten kan bedres om man klarer å lage strukturerte essay, og etablerer klare retningslinjer for poenggiving og karaktersetting (2).

2.2 *Databaserte, skriftlige eksamener*

Databaserte eksamener gis flere steder som samlinger av disjunkte multiple-choice-spørsmål av typen «single best answer». Ved vårt fakultet er de databaserte oppgavene typisk brukt til å gjenskape den kliniske utrednings- og beslutningsprosessen fra den første konsultasjon med pasienten, gjennom undersøkelser og prøvesvar frem til endelig diagnose og behandling. I preklinikken tillempes dette med andre resonneringsformer. Oppgavene er således bygget opp i sekvenser. Kandidatene leverer inn en deloppgave, får et fasitsvar, og fortsetter deretter med neste deloppgave. Ved en slik fremgangsmåte unngås følgefeil. Databaserte eksamener kan også inkorporere fritekstsvaer (miniessay) og ulike typer flervalgsspørsmål.

Fordelen med databaserte, skriftlige eksamener er at man kan få god faglig validitet, god utvalgsvaliditet, og høy reliabilitet. Ulempen er at det kreves mye erfaring for å skrive gode spørsmål på et høyt kognitivt nivå og for å få til et stabilt nivå i vanskelighetsgrad. Opplæring og øvelse er nødvendig.

I relasjon til digital eksamen bygger fakultetet nå opp egne «item writing»-kurs for å bidra til at oppgavesettene er valide og tester det de er ment å teste. Når sensureringen er gjort, vil systematiske analyser av resultatene gi verdifull informasjon om spørsmålenes egenskaper - både reliabilitet og validitet – og om sensurens kvalitet gjennom statistiske analyser som

sammenfattes under begrepet psykometri. Fakultetet har i samarbeid med UiOs Centre for Educational Measurement (CEMO) bygd opp en slik analytisk kapasitet.

2.3 Muntlig eksamen

Muntlige eksamener skiller seg fra skriftlige ved at man responderer muntlig på spørsmål/oppgaver foran eksaminator og sensor. I den nye studieplanen vil det gjennomføres en muntlig eksamen i anatomi hvor kandidaten blir forelagt anatomiske mikro- og makropreparater, og blir eksaminert i disse. I tillegg skal kandidaten gjennomføre funksjons- eller overflateanatomiske undersøkelser.

Ved klinisk muntlig eksamen undersøker kandidaten en pasient med anamneseopptak og somatisk undersøkelse og kommer med forslag til diagnose, videre undersøkelser og behandling. I tillegg kan kandidatene eksamineres i teoretisk kunnskap. Eksamen skal illudere en klinisk konsultasjon.

Muntlig eksamen kan også gjennomføres mer strukturert med ulike stasjoner, og det som skiller en slik stasjonseksamen fra en objektiv strukturert klinisk eksamen (OSCE) er at studentene ikke får like spørsmål. På den måten opprettholder man en strikt struktur, men får mer fleksibilitet i spørsmålene som stilles slik at man kan få et bredere vurderingsgrunnlag.

Muntlig eksamen kan ha høy validitet og relevans, men er sårbar for eksaminatorer som ikke fungerer optimalt. Bruk av reelle pasienter ved kliniske eksamener kan skape forskjellige utfordringer. Det er ved denne eksamensformen viktig å ha dyktige sensorer som har den nødvendige oversikt og evne til å vurdere prestasjonene på en god måte. Reliabiliteten kan også påvirkes av hvilke oppgaver/case som blir brukt, nervøsitet hos kandidaten osv. Noe av disse faktorene kan reduseres ved å øke antall case, øke testlengden og ha en mer strukturert presentasjon av pasientens sykehistorie eller den aktuelle problemstillingen (3).

2.4 Objektiv strukturert klinisk eksamen - OSCE

Ved en objektiv strukturert klinisk eksamen (OSCE) er oppgavene standardisert slik at alle kandidatene får de samme oppgavene. Vanligvis avholdes OSCE i form av en stasjonseksamen der kandidatene roterer mellom flere stasjoner hvor de presenteres for de samme skriftlige, muntlige eller praktiske oppgaver av ulike slag med utgangspunkt i simulerte pasienter, dukker eller ulike typer materiale, men en slik eksamen kan også gjennomføres i form av én lengre stasjon.

OSCE tester medisinske kunnskaper og kliniske ferdigheter slik som klinisk vurderingsevne, undersøkelsesmetodikk, evne til resonnering, samt problemløsning. Videre testes kommunikasjonsferdigheter inklusive atferd overfor pasienten, og språklig formidlings- og oppfattelsesevne. Kandidatenes kompetanse om utredning og behandling av sykdom testes også. Eksamensformen ble opprinnelig utviklet i England og USA for å skape mer objektive og reliable kliniske eksamener. Det benyttes predefinerte, objektive kriterier for hva som skal vurderes og hvordan. Eksamensformen kan gi både økt utvalgsvaliditet og reliabilitet, og potensielt også faglig validitet.

Reliabiliteten av eksamen sett under ett skal representere et estimat av korrelasjonen mellom en konstruert tilstand (eksamen) og en hypotetisk/ideell eksaminering satt sammen av hele porteføljen av kliniske ferdigheter som studentene skal kunne, men struktur og objektivitet alene garanterer ikke suksess. Gjennomgang av forskningslitteraturen som er gjort på denne typen eksamen viser at reliabiliteten til stasjoner som tester praktiske kliniske ferdigheter stort sett er god, men stasjoner som tester kommunikasjon og pasienthåndtering viser en mer vekslende reliabilitet (5).

2.5 Prosjektoppgaven og andre skriftlige innleveringer

Alle studenter skal i ny studieplan levere et selvstendig skriftlig arbeid i løpet av studiet, vektet til 20 studiepoeng. Det er satt av totalt 13 uker til arbeid med prosjektoppgaven. Hensikten er å stimulere til vitenskapelig arbeide og forskning og gi studentene mulighet til å få spesialisert innsikt i en avgrenset del av et fagområde. Studentene kan i andre sammenhenger levere inn skriftlige oppgaver individuelt eller i gruppe.

Slike eksamener tester evne til refleksjon, resonneringsevne og skriftlig fremstillingsevne. Det er viktig at det er etablert presise krav som sensuren kan vurdere oppgaven opp mot.

2.6. Arbeidsgruppens anbefalinger

- *Ved skriftlige eksamener må man tilstrebe at oppgavesettene har de samme egenskapene med hensyn til vanskelighetsgrad og utvalgsvaliditet slik at prestasjonene måles med lik målestokk fra kull til kull. Økt kvalitetssikring kan oppnås ved gjenbruk av spørsmål og gradvis raffinering av oppgaver og oppgavesett, slik fakultetet nå legger opp til ved digitale eksamener.*
- *Ved bruk av reelle pasienter må man vurdere om pasienten er egnet for å delta i en eksamenssituasjon. Oppgaver/kasus bør skriftliggjøres. Vurderings- og skåringsskjemaer som brukes bør gjennomgå med alle som skal være eksaminator/sensor, og man må utvikle delpunkter som muliggjør kriteriebasert karaktersetting.*
- *«Veileder for eksaminatorer og sensorer», «Rettleder til leder og medlemmer av eksamenskommissjoner for modifiserte OSCE eksamener» og «Rettleder for gjennomføring av modifisert OSCE» bør revideres og tilpasses eksamen i ny studieplan. Det må nedfelles prinsipper for hvordan eksamenskommissjonene skal arbeide med å lage oppgavesett og planlegge praktisk gjennomføring av eksamen, hvor rutiner for å utforme oppgaver og oppgavesett, utvikling av skåringsmaler, rutiner for tilbakemelding fra studenter og etterarbeid formuleres. Oslo 2014-prosjektet tildeles ansvar for å arbeide med dette i samarbeid med Studieseksjonen og andre relevante parter.*

3. Anbefalinger om hvilke eksamener som skal vurderes etter graderte karakterer

Det ligger noen begrensninger i arbeidsgruppens anbefaling om hvilke eksamener som skal vurderes etter gradert karakterskala all den tid det gjenstår mer arbeid i modulgruppene.

Det er også kommet signaler fra dekanatet om at det på sikt jobbes for å få på plass en eller flere felles avsluttende eksamener for alle de fire medisinstudiene i Norge. Dersom det blir en realitet, vil premissene for våre anbefalinger om avsluttende eksamen kunne endres, men det har arbeidsgruppen ikke tatt hensyn til i denne rapporten.

3.1 Forhold som taler for bruk av bestått/ikke-bestått

Evaluering av studenters kompetanse eller læringsutbytte har ulike funksjoner. Ved eksamener som primært har en sertifiseringsfunksjon, vil man kunne hevde at det ikke vil være hensiktsmessig å bruke en gradert karakterskala. Arbeidsgruppen har vurdert dette slik at:

- små eksamener som eksamen i propedeutikkblokken i modul 2, hvor studentene kun har hatt få uker med klinisk undervisning, bør vurderes med bestått/ikke-bestått
- klinisk muntlig eksamen i modul 3, som er viktig for sertifisering av egnethet for et fremtidig yrke som lege, bør vurderes med bestått/ikke-bestått
- prosjektoppgaven som primært skal dokumentere grunnleggende vitenskapelig kompetanse, og gi studentene mulighet til å få spesialisert innsikt i en avgrenset del av et fagområde, bør vurderes med bestått/ikke-bestått

3.2 Arbeidsgruppens anbefalinger:

Tabell 1: Oversikt over arbeidsgruppens anbefalinger	
Modul 1:	Anbefaling
<ul style="list-style-type: none">• Blokk 1: Skriftlig eksamen i statistikk, helse, biologi og samfunn (essay eller kortsvarspørsmål, utregningsoppgaver og bruk av formelverk)	A-F
<ul style="list-style-type: none">• Blokk 2: Skriftlig eksamen i cellebiologi (essay eller kortsvarspørsmål, oppsett av reaksjonsligninger)	A-F
<ul style="list-style-type: none">• Blokk 3: Skriftlig, digital eksamen i blod/immunologi/mikrobiologi/thorax	A-F
Modul 2:	
<ul style="list-style-type: none">• Muntlig eksamen i anatomi	A-F
<ul style="list-style-type: none">• Skriftlig, digital eksamen i fysiologi, ernæring og medisinsk biokjemi	A-F
<ul style="list-style-type: none">• Strukturert muntlig eksamen (5 stasjoner) i propedeutikk,	Bestått/ikke-

førstehjelp, atferdsfag og etikk	bestått
Modul 3:	
<ul style="list-style-type: none"> Skriftlig digital eksamen I (indremedisin, kirurgi, anesthesiologi, farmasi, radiologi og klinisk biokjemi) 	A-F
<ul style="list-style-type: none"> Skriftlig digital eksamen II (patologi, hud- og veneriske sykdommer, immunologi, mikrobiologi og ernæring) 	A-F
<ul style="list-style-type: none"> Klinisk muntlig eksamen/OSCE (kliniske fag, hud- og veneriske sykdommer, atferdsfag) 	Bestått/ikke-bestått
Modul 4:	
<ul style="list-style-type: none"> Skriftlig, digital eksamen i nevrologi, øre-nese-hals, øyesykdommer, allmennmedisin, parakliniske fag 	A-F
<ul style="list-style-type: none"> Klinisk muntlig eksamen (50 % nevrologi, 25 % øye, 25 % ØNH) 	A-F
Modul 5:	
<ul style="list-style-type: none"> Skriftlig eksamen i psykiatriske fag, rettsmedisin, allmennmedisin, samfunnsmedisin og medisinsk etikk. 	A-F
<ul style="list-style-type: none"> Skriftlig digital eksamen i farmakologi 	A-F
<ul style="list-style-type: none"> Klinisk muntlig eksamen (4/6 psykiatri, 1/6 allmennmedisin, 1/6 samfunnsmedisin) 	A-F
Modul 6:	
<ul style="list-style-type: none"> Skriftlig, digital eksamen for alle fag i modulen (gynekologi/obstetrikk, pediatri, patologi, samfunnsmedisin, allmennmedisin, medisinsk etikk, farmakologi, atferdsfag, klinisk biokjemi, genetikk, psykiatri) 	A-F
<ul style="list-style-type: none"> OSCE (gynekologi/obstetrikk, pediatri, patologi) 	A-F
Modul 7:	
<ul style="list-style-type: none"> Skriftlig, digital eksamen med noen short-essay-spørsmål i samfunnsmedisin, allmennmedisin, KLoK og medisinsk etikk 	A-F
Modul 8:	
<ul style="list-style-type: none"> Skriftlig, digital eksamen 	A-F
<ul style="list-style-type: none"> OSCE 	A-F
<ul style="list-style-type: none"> Mini-CEX 	A-F
Prosjektoppgaven	Bestått/ikke-bestått

4. Vitnemål

4.1. Hovedkarakter

Vitnemålet studentene får etter endt utdanning, vil i ny studieplan bestå av en samlet karakter for hvert emne (emne = modul). De ulike karakterene som er oppnådd ved de ulike deleksamenene vil være synlige på vitnemålet, men det vil bli én emnekarakter som er den tellende.

Karakterene på vitnemålet og på diploma supplement vil være en blanding av bestått/ikke-bestått og graderte karakterer. Det vil da ikke fremkomme en hovedkarakter på vitnemålet for hele graden, men det slås fast at NN har oppnådd graden candidatus/candidata medicinae på grunnlag av studieprogram: medisin med en normert studietid på 6 år. Informasjon om karakterer i de enkelte emner skal likevel fremgå av vitnemålet.

Som hovedregel bør bestått-medisinsk-studium gi rett til opptak på doktorgradsprogrammet. Dagens praksis er at alle som har fullført medisnutdanningen i Norge, er kvalifisert for opptak, mens det for andre er et krav om B-snitt eller høyere. Når vi innfører graderte karakterer ved vårt studium, vil det være to studiesteder som har graderte karakterer (Bergen og Oslo) og to som ikke har graderte karakterer (Tromsø og Trondheim) i Norge. Dette kan få implikasjoner dersom doktorprogrammet endrer opptakskriteriene for våre studenter.

4.2. Arbeidsgruppens anbefalinger

- *Informasjon om karakterer i de enkelte emner skal fremgå av vitnemålet.*
- *Det gjøres ikke endringer i opptakskravet til doktorgradsprogrammet etter innføring av graderte karakterer.*

5. Implementering

I fakultetsstyrets vedtak ligger det et krav om gjennomført kvalitetssikring av eksamensformer før implementering. Det må gjøres en helhetlig vurdering av eksamen opp mot kvalitetskriterier formulert tidligere i dette dokumentet, og arbeidsgruppen har pekt på tiltak under de ulike eksamensformene som bør være på plass før graderte karakterer innføres.

En overordnet vurdering bør gjøres av studiedekan (programleder) i samråd med programrådet, basert på dokumentasjon fra ulike instanser herunder eksamenskommissjoner, studieseksjonens rapporter og vurderinger fra Helsevitenskapelig utdanningscenter ved førstegangs implementering, og som en løpende kvalitetssikring av de ulike eksamener.

5.1 Tidsplan for implementering av graderte karakterer

Arbeidsgruppen ser verdien av å vinne erfaring med hvordan eksamen fungerer i nye moduler før det implementeres graderte karakterer. Videre er det et premiss at graderte karakterer skal implementeres kun fra og med de studentkull som tas opp på studiet fra høsten 2014, og altså ikke overgangskull som blir faset over i ny studieplan etter 6.semester og inn i modul 4. Arbeidsgruppen anbefaler på det grunnlaget at det implementeres graderte karakterer fra studiets start (modul 1) fra og med kull H-16 (høsten 2016) under forutsetning av at kvalitetssikringen anses som innfridd. Tabellen under viser anbefalt implementeringsplan for kull i ny studieplan.

Kull	M1	M2	M3	M4	M5	M6	M7	M8
H-14				X	X	X	X	X
V-15				X	X	X	X	X
H-15			X	X	X	X	X	X
V-16		X	X	X	X	X	X	X
H-16	X	X	X	X	X	X	X	X
V-17	X	X	X	X	X	X	X	X

5.2. Arbeidsgruppens anbefalinger

- *Det implementeres graderte karakterer fra studiets start (modul 1) fra og med kull H-16 (høsten 2016); fra og med modul 2 for kull V-16 (høst 2017); fra og med modul 4 for kull H-14 (høst 2017); fra og med modul 3 for kull H-15 (vår 2018) under forutsetning av at kvalitetssikringen anses som innfridd.*

Kilder

1. Frich JC, Røttingen J-A. Evaluering av studenter i grunnutdanningen. Tidsskr Nor Lægeforen 1995; 115: 237-8.
2. Epstein RM. Assessment in medical education. New Engl J Med 2007; 356: 387-96.
3. Wass V, der Vleuten CV, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001; 357: 945-9.
4. Norcini J et al. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 conference. Med Teach 2011; 33: 206-14.
5. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Med Educ 2011; 45: 1181-9.
6. Lubarsky S, Charlin B, Cook DA, Chalk C, van der Vleuten CPM. Script concordance testing: a review of published validity evidence. Med Educ 2011; 45: 329-38.
7. Donabedian, A. Evaluating the Quality of Medical Care. The Milbank Quarterly, Vol. 83, No. 4, 2005 (pp. 691-729)
8. Norcini JJ. Setting standards on educational tests. Med Educ 2003; 37: 464-9.
9. Sadler DR. Interpretations of criteria-based assessment and grading in higher education. Assessment & Evaluation in Higher Education 2010; 30: 175-94.
10. Sadler DR. Grade integrity and the representation of academic achievement. Studies in Higher Education 2009; 34: 807-26.