

Nasjonal delprøve i medisinstudiet - gjennomføring av piloten mars 2017

Rapport 2017

Nettversjon, 8. juni, 2017

Innhold

1. Bakgrunn, mandat og gruppens arbeid	3
2. Hovedkonklusjoner og anbefalinger	4
3. Prosjektgruppens arbeid og etablering av fagkomitéene	5
4. Fagfellevurdering av oppgavene	9
5. Tekniske løsninger	12
6. De administrative sidene	16
7. Oppgaveanalyse og resultater	18
8. Studentenes erfaringer.....	23
9. Videre arbeid.....	25

Nasjonal delprøve: <http://www.med.uio.no/om/prosjekter/nasjonal-delprove-medisin/>

1. Bakgrunn, mandat og gruppens arbeid

Det nasjonale utdanningsmøtet i medisin vedtok 10. oktober 2014 å starte arbeidet med en pilot på en nasjonal delprøve i klinisk resonnement. Det ble nedsatt en nasjonal arbeidsgruppe for å utrede muligheten for en slik prøve, med følgende mandat (27. mars 2015):

«... å utarbeide en skriftlig digital delprøve i klinisk resonnement som bygger på læringsmål som er gjenfinnbare i studieplanen på alle fire læresteder til og med 11. semester. Delprøven gis til 6.års studenter som er inne i sitt 12.semester, tidlig i vårsemesteret. Prøven skal ha et omfang på 2-4 timer og vurderes med godkjent/ikke godkjent».

I en rapport til det nasjonale utdanningsmøtet 20. oktober 2015 presenterte arbeidsgruppen en plan for neste fase i prosjektet og anbefalte at det skulle arbeides videre med en nasjonal delprøve. Det nasjonale utdanningsmøtet sluttet seg til anbefalingen, og det nasjonale dekanmøtet vedtok i desember 2015 at det skulle arbeides mot å gjennomføre en pilot. Arbeidsgruppen/prosjektgruppen har i prosjektperioden (fra april 2015) bestått av:

Jan Frich (UiO), prosjektleder
Linda K. Røine (UiO), sekretær
Hanne-Guro W. Aabelvik (UiO), sekretær

Henrik Schirmer (UIT), vitenskapelig representant
Tobias S. Slørdahl (NTNU), vitenskapelig representant
Anders Bærheim (UiB), vitenskapelig representant

Elin Holm og Kristin Elisa Ruud Hansen (UiT), administrative representanter
Eirik Dalheim (UiB), administrativ representant
Marte Laugen (NTNU), administrativ representant

Studentrepresentanter:

Stine Andersen (UiT)
Oda Lockert (UiO) – fra november 2016
Eivind Valestrand (UiB) – til oktober 2016
Gard A. Skulstad Johanson (UiB) – fra november 2016
Maren Ellingsen (NTNU) – fra november 2016

Eivind Valestrand (UiB) gikk ut av prosjektgruppen etter eget ønske da han ble valgt som ny leder av Norsk medisinstudentforening (Nmf).

Denne rapporten er utarbeidet av prosjektgruppen i fellesskap og omhandler ulike sider ved planleggingen og gjennomføringen av piloten. Hovedkonklusjoner og anbefalinger er samlet innledningsvis. Prosjektet ble presentert på det nasjonale dekanmøtet 2. juni 2017.

Oslo, 8. juni 2017

Jan Frich, leder av arbeidsgruppen/prosjektgruppen

2. Hovedkonklusjoner og anbefalinger

- Nasjonal delprøve i medisin ble gjennomført som planlagt 24. mars 2017. Det faglige og administrative samarbeidet har fungert svært godt. Prøven var frivillig, og totalt 319 (83 %) av 384 studenter i sistesemester ved de fire studiestedene gjennomførte prøven. Oppmøtet varierte fra 73 % til 94 % på tvers av studiesteder.
- Snittskår var 70,7 % korrekte svar. Resultatene var tilnærmet like mellom studiestedene og de små forskjellene er vanskelig å tolke i lys av ulik oppmøteprosent. I tillegg må man vurdere resultater over tid for å kunne si noe om forskjeller. Studentene har dessuten noe ulik erfaring med oppgavetyper som er brukt og digital eksamen. Hovedinntrykket er at nivået er rimelig jevnt på tvers av de fire studiestedene.
- Fasit med begrunnelser ble publisert etter av prøven var over. Studentene kom med mange gode tilbakemeldinger på oppgavene etter prøven var gjennomført, som ble tatt hensyn til i sensuren. Løsningen vi valgte for individuell tilbakemelding synes å ivareta studenters behov for individuell tilbakemelding og begrunnelse.
- Kvalitetssikringen med bruk av fagkomitéer og tre fagfeller (spesialist, lege i spesialisering og allmennlege) har fungert godt og gitt spørsmål av høy kvalitet. Spørsmålene ble jevnt over ble oppfattet som relevante og gode av studentene.
- Av de 319 studentene som gjennomførte prøven, svarte 152 (48 %) på et evalueringsskjema. Studentene anga at enkelte spørsmål hadde en spesialistorientering og etterlyste et bedre digitalt system for eksamensavvikling. Det må legges mer arbeid i en korrekturlesning for å sikre konsistens (bl.a. for referanseverdier) i prøvesettet som helhet. Det synes å være en viss enighet om at en nasjonal delprøve i seg selv er et positivt tiltak (84 %), og 68 % svarer at de ønsker at det skal jobbes mot en felles skriftlig avsluttende eksamen.
- Prosjektgruppen tilrår at prosjektet videreføres med en ny nasjonal delprøve som et obligatorisk arbeidskrav i studienes siste semester, vår 2018. Et nasjonalt sekretariat og prosjektledelse er viktig for å sikre koordinering i prosjektet.
- Høy reliabilitet og validitet kan oppnås ved å øke testtiden og antall spørsmål på prøven. På bakgrunn av studentenes faktiske tidsbruk anbefaler vi at man går ned på 90 sekunder per spørsmål (internasjonal standard) og øker omfanget til en 4-5 timer per prøve. Nye fagområder kan da inkluderes, f.eks. lunge, endokrinologi og infeksjonsmedisin og muskel-skjelett/ortopedi.

3. Prosjektgruppens arbeid og etablering av fagkomitéene

Arbeidsgruppen pekte i sin rapport fra 2015¹ (1) på at det på et overordnet nivå var rimelig godt samsvar i læringsmål på tvers av studiene, men at detaljeringsgraden var svært forskjellig fra studiested til studiested. En ordning med nasjonale fagkomitéer ble løsningen på spørsmålet om forankring av spørsmål i lokale studieplaner. Fagkomitéene var dessuten viktige i utarbeidelsen av spørsmål. Tre fagkomitéer som dekket seks medisinske/kirurgiske fagområder var blitt foreslått opprettet av arbeidsgruppen:

- Gastroenterologi/gastrokirurgi
- Kardiologi/thoraxkirurgi
- Nefrologi/urologi

Fagkomitéene

Prosjektgruppen laget en oversikt over en ønsket sammensetning av komitéene og etablerte også en paraklinisk referansegruppe (patologi, mikrobiologi, farmakologi og klinisk biokjemi). I perioden mars og april 2016 ble representanter fra de enkelte fakultetene oppnevnt til de nasjonale fagkomitéene. Hvert av fakultetene oppnevnte 1 kirurg, 1 indremedisiner og 1 allmenntilleggsmedisiner (med unntak av UiO) til de tre fagkomitéene, samt 1 representant til den parakliniske referansegruppen (UiO ble bedt om å oppnevne to medlemmer til denne).

Gastroenterologi og gastrokirurgi

Kristian Bjøro, UiO (leder)
Trond Buanes, UiO
Arne Wibe, NTNU
Arne Sandvik, NTNU
Irene Hetlevik, NTNU
Jan Gunnar Hatlebakk, UiB
Anne Karliczek, UiB
Jon Florholmen, UiT
Øivind Irtun, UiT

Kardiologi og thoraxkirurgi

Henrik Schirmer, UiT (leder)
Sverre Kjeldsen, UiO
Steinar Solberg, UiO
Stein Samstad, NTNU
Alexander Wahba, NTNU
Peter Schuster, UiB
Rune Haaverstad, UiB
Anders Bærheim, UiB
Kristian Bartnes, UiT

Nefrologi og urologi

Knut Ivar Aasarød, NTNU (leder)
Aud Høieggen, UiO
Stig Müller, UiO
Carl-Jørgen Arum, NTNU
Hans Peter Marti, UiB
David Thomas Davidsson, UiB
Vinjar Fønnebø, UiT
Marit Solbu, UiT
Erling Aarsæther, UiT

Paraklinisk referansegruppe

Inger Nina Farstad, UiO
Hege Thoresen, UiO
Jan Egil Afset, NTNU
Rune Ulvik, UiB

Det ble arrangert et oppstartsseminar på Gardemoen 18. mai 2016. Her var alle komitémedlemmene samlet, med godt oppmøte fra alle fagområder og institusjoner.

¹ Nasjonal delprøve i medisinstudiet: rapport 2015. http://www.med.uio.no/om/prosjekter/nasjonaldelprøve-medisin/prosjektdokument/nasjonaldelprøve_rapport_uttildning_mote_1910151.0.pdf

Hensikten med samlingen var å etablere en felles plattform for komitéenes arbeid gjennom å gjøre alle kjent med prosjektets mandat og målsetninger. Vi ønsket også å gi en opplæring i det å arbeide med *flervalgsoppgaver med ett beste svar*, som var den oppgavetypen som var blitt valgt. Det ble utpekt ledere for de enkelte komitéene:

- Kristian Bjørø (UiO, gastroenterologi og -kirurgi)
- Henrik Schirmer (UiT, kardiologi og thoraxkirurgi)
- Knut Ivar Aasarød (NTNU, fagkomitéleder nefrologi og urologi).

Sistnevnte komité valgte å dele seg i to undergrupper som jobbet hver for seg. Stig Müller (UiO) ledet arbeidet innen urologi.

Parakliniske fag

Representantene fra de parakliniske fagene ble plassert i hver av de tre fagkomitéene for å sikre at de ble involvert i de kontinuerlige arbeidet. Tanken var at de da lettere skulle kunne lage spørsmål som knyttet seg til resten av oppgavene, i tillegg til at de skulle ha en rolle på tvers av de tre komitéene.

Arbeidsform og etablering av database

Utover høsten 2016 møttes komitéene hver for seg på Gardemoen (1-2 møter). I tillegg samarbeidet de via mail. Prosjektledelsen hadde egne oppfølgingsmøter med lederne av fagkomitéene i løpet av høsten. I oktober 2016 forelå utkast til oppgaver fra samtlige komitéer. Kvalitetssikring av oppgavene ble gjort ved bruk av ekstern fagfelleevaluering, med involvering av prosjektledelsen, i perioden november-desember 2016. De endelige oppgavene ble ferdigstilt i januar 2017.

Parallelt pågikk arbeidet med å etablere oppgavedatabasen. NTNU utlyste en stilling for drift og etablering av databasen for nasjonal delprøve. Stillingen kom på plass med oppstart høsten 2016. Målsettingen i pilotfasen var å etablere struktur og funksjonalitet for en database som skulle gi fagkomitéene et digitalt verktøy for samarbeid om oppgavene. Til piloten var det ikke mulig å få på plass verktøyet tidsnok til at komitéene kunne ta det i bruk i arbeidet sitt, og de arbeidet derfor med bruk av Word-dokumenter. Databasen ble brukt fullt ut av sekretariatet i forbindelse med innlegging av endelige oppgaver. En nærmere beskrivelse av den tekniske løsningen gis siden i rapporten.

Prosjektgruppens arbeide

Gjennom 2016 var prosjektgruppens opptatt av klargjøring av oppgaver til prøven og en teknisk løsning for å distribuere den. Samarbeidet gikk via felles samlinger på Gardemoen og jevnlig telefonmøter. I tillegg ble det holdt løpende kontakt og møter knyttet opp mot de ulike arbeidsoppgavene. Et utvalg av prosjektgruppens medlemmer forfattet en artikkel om nasjonal delprøve som skulle tjene som en felles informasjonsressurs². Informasjon ble dessuten fortløpende lagt ut på prosjektets nettside³. Det ble, med Tobias S. Slørdahl som primus motor, utviklet en egen veileder for komitéene⁴.

² Valestrand EA, Slørdahl TS, Bærheim A, Schirmer H, Andersen S, Aabelvik H-G, Frich J. Mot en nasjonal delprøve i medisinstudiet. Tidsskr Nor Legeforen 2016; 136: 390-1. Lenke:

<http://tidsskriftet.no/2016/03/kommentar-og-debatt/mot-en-nasjonal-delprove-i-medisinstudiet>

³ Nasjonal delprøve i medisin: <http://www.med.uio.no/om/prosjekter/nasjonal-delprove-medisin/>

⁴ Nasjonal delprøve i medisin: veileder (2016). <http://www.med.uio.no/om/prosjekter/nasjonal-delprove-medisin/aktuelle-saker/veileder-for-nasjonal-delprove.html>

Informasjon og studentinvolvering

I forbindelse med oppstart av pilotarbeidet ble det holdt et informasjonsmøte med studenter og tillitsvalgte ved alle studiestedene 7. mars 2016. Vi benyttet et online møterom (Adobe Connect) til å holde presentasjon med tale og presentasjon. Her kunne studentene i tillegg stille spørsmål skriftlig og muntlig. Tema for møtet var her nasjonal delprøve generelt. Målsetningen var å informere om prosjektet og arbeidet med en pilot, og å bidra til involvering av alle lærestedene ved å gi mulighet til å stille spørsmål og gi prosjektet innspill.

Studentrepresentanter fra samtlige læresteder ble også invitert med på oppstartsseminaret 18. mai 2016. Prosjektgruppen så videre behov for å sikre gode kommunikasjonskanaler inn mot studentene ved alle læresteder, gjennom å inkludere representanter fra også UiO og NTNU i arbeidsgruppens kjerneaktiviteter. Utdanningsmøtet ga tilslutning til dette i 3. november 2016.

Til studentgruppen som skulle være med i piloten ble det sendt ut et infoskriv i oktober 2016. En planlagt pilottesting av teknologien i november 2016 ble for ambisiøs og prosjektet måtte endre planen og teste i mindre skala. Nasjonal delprøve ble av ulike grunner ikke innført som et formelt obligatorisk arbeidskrav ved noe studiested. Vi la derfor stor vekt på læringsverdien av prøven overfor studentene og på betydningen av individuell tilbakemelding. Alle kull ble gitt informasjon og et sett med treningsoppgaver sendt ut primo mars 2017, slik at studentene på forhånd kunne gjøre seg kjent med systemet de skulle bruke. I tillegg fikk de da se eksempler på oppgavetyperne de ville møte under prøven. Informasjonsmøtene hadde godt oppmøte hvor et flertall av studentene var til stede. Vi valgte også å ta initiativ til en sak i *Dagens Medisin*, for å skape oppmerksomhet om prøven og dens formål⁵.

Praktiske lokale utfordringer

Arbeidet med piloten har pågått i et landskap preget av endringene som skjer på institusjonene i forbindelse med den generelle innføringen av digital eksamen. Blant annet har samtlige institusjoner bygget om eksamenslokalene sine for å få en bedre tilpasning til behovene knyttet til digitale eksamener. Lokalene ved NTNU og UiT var ikke klare til 24. mars 2017. Alternative lokaler måtte derfor finnes og klareres med tanke på tilstrekkelig strøm- og nettverkstilgang. Ved UiB ble eksamenslokalet klart, men en del ombygging pågikk fremdeles i nærheten av lokalet som skulle brukes, og det var lenge en fare for at det kunne være støy under prøvens gjennomføring. Disse tingene var en utfordring for lokale administrativt ansvarlige. Alle institusjonene har imidlertid gjort en stor innsats for å finne gode løsninger som gjorde det mulig å få gjennomført piloten.

Administrativt samarbeid

På det administrative området har det vært et godt samarbeid i planleggingsprosessen frem mot gjennomføringen. Samarbeidet har foregått via løpende kontakt på telefon og e-post, og man har hyppig benyttet online møterom med Adobe Connect og telefonmøter. Det å ha jevnlig felles møter har vært svært nyttig og konstruktiv med tanke på å identifisere og diskutere likheter og ulikheter i lokale rutiner som i en nasjonal sammenheng må samkjøres. I tillegg har det bidratt til at tanker om muligheter, risiko eller ideer har kunnet bli delt mellom alle institusjonene og sånn sett bidratt til et godt resultat.

⁵ Engen ØB. Nå får medisinstudentene for første gang en nasjonal delprøve. *Dagens Medisin* 17.2. 2017: <https://www.dagensmedisin.no/artikler/2017/02/17/arrangerer-nasjonal-prove-for-forste-gang/>

Vurderinger:

- Fagkomitéene virker å være tilstrekkelig store, med en sammensetning som sikrer tilstrekkelig lokal forankring ved de fire studiestedene.
- Erfaringen tilsier at fagkomitéene trenger minst 3-4 mnd. arbeid for å utvikle oppgavene.
- Det er positivt at representanter for parakliniske fag er involvert i arbeidet. I praksis ble det slik at medlemmene i den parakliniske referansegruppen ble fullverdige medlemmer i en enkelt komité. Det var lite behov for konsultasjon på tvers av komitéene med hensyn til parakliniske problemstillinger. Vi tilrår at en fagperson med paraklinisk bakgrunn er med i fremtidige fagkomitéer.
- Allmennmedisinsk representasjon i komitéene er viktig og man kan i fremtiden vurdere å supplere med en ekstra allmennmedisiner i hver komite.
- Det er viktig at det finnes et nasjonalt sekretariat/prosjektledelse som kan bidra til koordinering og tjene som en ressurs for fagkomiteene.
- Det er viktig at det er enkeltpersoner lokalt som har et dedikert ansvar for gjennomføringen. Felles møter og samarbeid mellom sekretariat og de lokale representantene er en god erfaring som bør videreføres.

4. Fagfellevurdering av oppgavene

En viktig del av kvalitetssikringen og utarbeidelsen av oppgaver til nasjonal delprøve foregikk gjennom *ekstern fagfellevurdering*. Det var bred enighet om at fagfellene måtte representere både nyutdannede og erfarne leger, og sammen ville de kunne gi tilbakemelding på vanskelighetsnivå, kvaliteten på oppgaver og fasit. Gjennomleserne skulle være uavhengige av universitetene og heller ikke være tilknyttet undervisning eller praksis på noen av de 4 steder. I arbeidsgruppens møte den 14. september 2016 ble ulike rekrutteringsstrategier diskutert, inkludert bruk av fagmedisinske foreninger. Prosjektet besluttet i prosjektfasen å rekruttere fra lokale nettverk. Gjennom en slik rekrutteringsstrategi kom samtlige fagfeller på plass og var klare til gjennomlesning når oppgavene forelå, noe som hindret en forsinkelse i videre prosess.

Hvem var fagfellene?

Det ble besluttet å ta i bruk totalt ni eksterne fagfeller, tre for hvert oppgavesett. Det var ønskelig at hver fagkomité fikk oppgavene vurdert av en spesialist i allmenntidmedisin, en lege i spesialisering (LIS) og en ferdig spesialist. Arbeidet ble honorert med 10 timer (ved bruk av standard takster) og representantene skulle som hovedregel lese igjennom alle oppgavene til en fagkomité. Med unntak av en komité (kardiologi og thoraxkirurgi) hvor LIS-lege og spesialist kun leste gjennom hvert sitt område, leste samtlige allmenntidmedisinere, LIS-leger og spesialister gjennom hele oppgavesettet til en komité. Fagfellene fikk to ukers frist og tilbakemeldingene ble for det meste oversendt sekretariatet i tråd med denne fristen, men lengre overskridelser fra enkeltpersoner medførte at bearbeidningen av fagfellevurderingen i sekretariatet ble noe forsinket.

Vurderingsskjema

For å sikre en mest mulig enhetlig og strukturert tilbakemelding ble det utarbeidet et Excel-skjema til utfylling, samt at fagfellene ble bedt om å kommentere på oppgavene direkte i en Word-fil. Skjemaet i Excel besto av 5 kolonner med følgende innhold:

1	2	3	4	5
Hvordan vurderer du oppgavens relevans for hva en nyutdannet norsk lege bør kunne? 1) Essensielt 2) Viktig 3) Akseptabelt 4) Irrelevant	Tester oppgaven klinisk resonnement? JA eller NEI?	Er det beste svaralternativ uten tvil det beste? JA eller NEI? Angi evt. alternativ.	Hvordan vil du vurdere begrunnelsen for det riktige svaret? 1) God 2) Akseptabel 3) Dårlig	Hva er din totalvurdering av oppgaven? 1) Kan gis på eksamen 2) Bør revideres før eksamen 3) Fjernes fra eksamen

Alle oppgaver måtte vurderes ut ifra denne malen, og i kombinasjon med utdypende kommentarer i Word-filen var dette en hensiktsmessig struktur. Der gjennomleser kun hadde markert i Excel-filen f.eks. i kolonne 3, 4 og 5 var det mer utfordrende å bruke dette som grunnlag for en god tilbakemelding. Jevnt over fikk man allikevel nok kommentarer i Word-formatet til å være matnyttig for fagkomité. Det anbefales at det ikke bare brukes Excel-fil i tilbakemeldingen, men heller benyttes en kombinasjon av strukturert tilbakemelding og kommentarer. Tilbakemeldingene fra fagfellevurderingen ble bearbeidet og strukturert ytterligere i sekretariatet og ble gjennomgått av prosjektleder før de ble oversendt fagkomité.

Oppgaver med behov for gjennomsyn og eventuell endring ble markert i et Excel-skjema. Det ble også laget et tilbakemeldingsnotat til hver komité hvor det ble gitt en generell oppsummering av tilbakemeldinger og en samlet oversikt over kommentarer til den enkelte oppgave. En nærmere beskrivelse av resultatet av fagfelle vurderingen følger her:

- *Kardiologi og thoraxkirurgi*: Totalt 51 oppgaver, hvorav 11 i sekvens (4 sekvensielle oppgaver), ble sendt til fagfelle vurdering. I Excel-arket var det 14 oppgaver hvor en eller flere av fagfellene påpekte behov for justeringer. Det ble også meldt inn et behov for mindre justeringer i 11 oppgaver og det ble foreslått at tre oppgaver utgikk.
- *Nefrologi og urologi*: Totalt 42 oppgaver, hvorav 7 sekvensielle, ble sendt til fagfelle vurdering. I Excel-arket var det 25 oppgaver hvor en eller flere av fagfellene påpekte behov for justeringer. Det ble også meldt inn behov for mindre justeringer i 20 oppgaver og det ble foreslått at 4 oppgaver utgikk.
- *Gastroenterologi og gastrokirurgi*: Totalt 45 oppgaver, hvorav 9 i sekvens (3 sekvensielle oppgaver), ble sendt til fagfelle vurdering. I Excel-arket var det 7 oppgaver hvor en eller flere av fagfellene påpekte behov for mindre justeringer.

Fagkomitéene fikk ca. 2 uker på å revidere oppgavesettet, men tidslinjen kunne med fordel ha vært noe lenger. I fremtiden kan man planlegge for noe bedre tid til prosessen rundt ferdigstilling av endelig prøvesett. Resultatet av siste runde i komité var følgende endringer i oppgavesettet:

- *Kardiologi og thoraxkirurgi*: Etter tilbakemelding fra fagfelle vurdering foreslo komité at en 1 oppgave ble strøket og 10 oppgaver ble foreslått fjernet i en prioritert rekkefølge ved avgrensning til 40 oppgaver. Fagkomité endret oppgavetekst eller tydeliggjorde fasit på 8 oppgaver og gjorde en tydeligere begrunnelse av fasit på oppgaver hvor fasit var korrekt, men hvor fagfelle vurderingen hadde innvendinger/spørsmål.
- *Nefrologi og urologi*: Etter tilbakemelding fra fagfelle vurdering foreslo komité at 2 oppgaver ble strøket. Fagkomité endret oppgavetekst eller tydeliggjorde fasit i 12 oppgaver. Fagfelle vurderingen vurderte 10 av oppgavene som for vanskelige eller med feil svar, men komité redegjorde i sin revisjon for riktig svar og sjekket ut at tema blir undervist på gitt nivå.
- *Gastroenterologi og gastrokirurgi*: Etter tilbakemelding fra fagfelle vurdering vurderte komité fortsatt samtlige oppgaver som aktuelle for bruk i prøven. Komité har ved siste revisjon utdypet enkelte begrunnelser og rettet opp oppgavetekst der det var behov for dette.

Tilbakemeldingene fra fagfelle vurderingen har øket kvaliteten på oppgavene, enten ved justering av spørsmål eller tydeliggjøring av fasit. Et slikt eksempel er følgende spørsmål:

En 45 år gammel kvinne søker fastlegen for anfall med magesmerter. Smertene sitter øverst i abdomen og stråler bak i ryggen. Anfallene varer fra en halv til to timer og er takvise og kraftige. Hun føler seg frisk ellers. Blodprøver (Hb, CRP, lever- galleprøver) er alle normale. Hvilken første undersøkelse er mest adekvat hos denne kvinnen?

A X 1 Ultralyd abdomen

Dette er beste alternativ da anamnesen tyder på at hun har gallekolikk, ultralyd er sensitivt og spesifikt for diagnosen av galleblærekonkrementer.

B 1 EKG

Intet coronart preg over disse plagene.

C 0 Gastroskopi

Ingen ting som minner om dyspepsi i denne sykehistorien.

D 0 CT abdomen

Ukomplisert galleblærestein er mest sannsynlig og da er ultralyd best

Ved fagfelle vurderingen ble alternativ A vurdert til ikke å være klart beste svaralternativ, men komitéen opprettholdt svaralternativ A som det klart beste. På sensurmøtet ble det imidlertid besluttet å gi uttelling også for svaralternativ B. Totalt sett er det allikevel få slike eksempler, noe som tyder på at innspillene fra fagfelle vurderingen enten førte til endring i spørsmålstilling eller utdypning i fasit, eller at de i noen tilfeller ble vurdert som lite relevante.

Samtlige spørsmål med fasit er publisert på nett⁶.

Vurderinger:

- Fagfelle vurderingen bidro til kvalitetssikring og forbedring av en rekke oppgaver og den bidro til tydeliggjøring av fasiter og begrunnelser.
- Det bør minst være tre fagfelle vurderere på et sett av oppgaver fra en fagkomité.

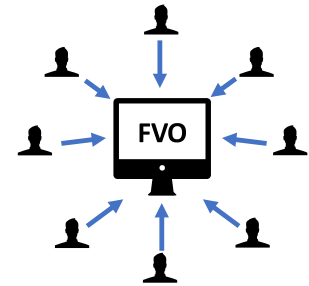
⁶ Fasit: <http://www.med.uio.no/om/prosjekter/nasjonal-delprove-medisin/informasjon-til-studenter/fasit/>

5. Tekniske løsninger

Oppgavedatabase

Det overordnede målet for den nasjonale databasen er å utvikle en felles elektronisk plattform hvor alle oppgaver til en nasjonal delprøve samles. De overordnede kravene til en nasjonal oppgavedatabase er at:

- Undervisere som er medlemmer av en fagkomité skal ha tilgang og mulighet til å legge inn oppgaver og kunne evaluere sine oppgaver i etterkant av en prøve
- Oppgavene må kunne eksporteres til de ulike digitale eksamensverktøyene som benyttes ved de ulike studiestedene.
- Systemet må støtte kravet om fagfelleevaluering av alle oppgaver.



Utviklingen av basen har vært ledet og utført ved NTNU. Utviklingen startet i september 2016. I desember var den første versjonen av databasen klar. I pilotfasen har de to første kravene blitt prioritert, da dette dreier seg om grunnleggende funksjonalitet som er nødvendig for å kunne legge inn og kvalitetssikre flervalgsoppgavene fra de ulike fagområdene, inkludert fasitsvar og begrunnelser for svaralternativene.

Det har vært avgjørende å få på plass funksjonalitet for endringshåndtering, oppbygning av prøvesett og produksjon av fasit. En offentlig tilgjengelig fasit med begrunnelser for hvorfor det riktige svaret er riktig ble anbefalt i rapporten fra 2015, og har vært et grunnleggende premiss i arbeidet. Videre har det vært sentralt å få på plass funksjonalitet for analyse av resultater (såkalt "psykometri"). Nasjonal delprøve består i tillegg til frittstående spørsmål av sekvensielle spørsmål (inntil 25 %). Databasen støtter ikke innlegging av spørsmål i sekvenser, noe som da måtte håndteres manuelt. Det bidro til noe merarbeid i forbindelse med klargjøring og oppsett av prøven samt endelig fasit og innebærer ekstra feilkilder.

Oppgavene som skulle inngå i nasjonal delprøve var allerede sendt ut på fagfelleevaluering da databasen ble tilgjengelig for bruk. Derfor ble den i denne runden kun testet ut av sekretariatet i forbindelse med innleggingen av de endelige oppgavene. Imidlertid var det også positivt, da erfaring med verktøyet gjør det enklere å tilrettelegge for at fremtidige komitéer og eksterne fagfeller kan ta det i bruk. I tillegg vil en første versjon ha noen barnesykdommer som vi enkelt fikk tatt tak i.

Løsning for prøvens gjennomføring

For å kunne gi en nasjonal delprøve med sekvensielle oppgaver ble det besluttet å bruke verktøyet som benyttes i Oslo på nasjonal delprøve, Questionmark Perception (QP). Dette systemet støtter sekvensielle oppgaver. Oppgavene ble dermed importert inn i QP og selve gjennomføringen ble gjort ved UiO. En utfordring var at UiO har en eksamensløsning basert på stasjonære maskiner mens de andre bruker laptopen på medbrakte maskiner. Studentene ved UiB, UiT og NTNU brukte dermed QP via sikker nettleser. Det ble vurdert som uhensiktsmessig i denne runden å involvere lokalt IT-personell i systemet. All monitorering foregikk derfor ved UiO, i tett dialog med det lokale gjennomføringsapparatet. Dette fungerte godt for prøven som sådan, men det opplevdes lokalt som uvant og uhensiktsmessig ikke å ha en tettere rolle i den tekniske delen av gjennomføringen. Det ble ikke registrert tekniske

problemer med de stasjonære maskinene i Oslo. Det ble registrert noen problemer underveis hos de som brukte egne maskiner med sikker nettleser. Hendelsene ble loggført og tatt tak i på stedet når de oppsto, slik at man var helt sikre på at ingen data gikk tapt (jf. vanlige rutiner for eksamensgjennomføring).

Fra studentene kom det tilbakemeldinger som viser at de opplevde det uvant å bruke QP som er et annet system enn de er vant til. Samtidig erfarte vi at QPs sikre nettleser som ble benyttet ved UiT, NTNU og ved UiB ikke fungerte hensiktsmessig i forhold til kriteriene for sikkerhet og brukervennlighet knyttet til en eksamensgjennomføring, kriterier det er naturlig at en nasjonal delprøve legger seg tett opp til. Etter prøven ble resultatene eksportert tilbake til NTNU for videre analyse.

Fasit

Databasen genererer på en enkel måte fasit som viser begrunnelser for hva som er rett svar. I basen er det mulig å legge inn begrunnelser også for de alternativene som ikke er poenggivende. Alle begrunnelser som legges inn på et spørsmål, vises i fasiten. Fasit ble lagt ut på nett umiddelbart etter prøven, slik at studentene kunne gå igjennom oppgavene og utforme sine tilbakemeldinger. Ny fasit ble publisert på prosjektets nettside etter sensurmøtet, 7. april. Her har systemet som er utviklet ved NTNU markert hvilke oppgaver som har blitt justert eller tatt ut. Et psykometrinotat forfattet av sensurkomiteen⁷ og gir en nærmere forklaring av endringene som er gjort, i tillegg til at sensurkomiteen her gir generelle betraktninger om prøvens reliabilitet og validitet og besvarer studentenes innspill som ble sendt inn i etterkant av prøven.

Egen besvarelse

Løsningen som ble brukt i piloten har ikke en arkiv-funksjon hvor studentens besvarelse lagres og er tilgjengelig for studenten etter prøven. Istedenfor ble besvarelsene sendt ut via et e-post-program som sendte rett besvarelse til rett student, basert på kandidatnummer. Det er en fordel om en fremtidig løsning for digital eksamen inneholder en arkiv-funksjon som en del av løsningen, slik at studentene umiddelbart kan ha tilgang til både fasit og egen besvarelse.

Tilbakemeldingsskjema

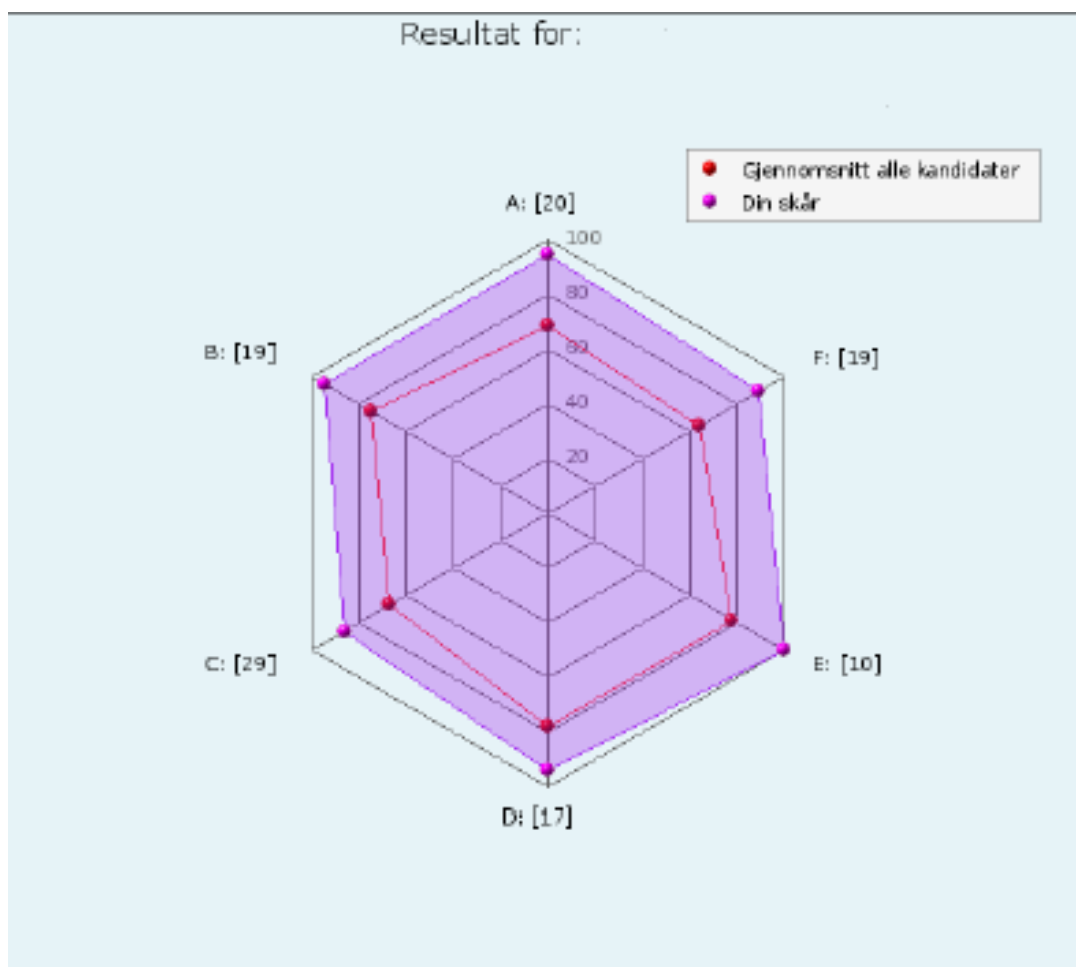
I rapporten fra 2015 ble det lagt vekt på prøvens formative egenskaper og nasjonal delprøve som en fremtidig læringsarena. Det ble lagt vekt på at det i den tekniske løsningen skulle inngå et opplegg for produksjon av en skriftlig tilbakemelding som kunne eksporteres ut fra det digitale verktøyet. Som et svar på dette ble det utviklet en rapport som genereres fra FVO-basen hos NTNU, og som kunne legges på nett. Tilbakemeldingen inneholdt informasjon om:

- Egen skår i forhold til nasjonalt gjennomsnitt på prøven
- Oversikt over delskår per fagområde i forhold til nasjonalt snitt oppgitt i prosent samt informasjon om hvilken prosentil man tilhørte blant kandidatene på eget fakultet
- Resultatene ble også fremstilt i et spider-diagram:

⁷ Psykometri og ny fasit: <http://www.med.uio.no/om/prosjekter/nasjonaldelprøve-medisin/informasjon-til-studenter/psykometri-og-ny-fasit/>

Tabell: Fordeling av oppgavene på fagområder

ID	Fagområde	Antall oppg.	Nasjonalt snitt (%)	Prosentil på eget fakultet
A	Gastroenterologi	20	89	95
B	Gastrokirurgi	19	75	95
C	Kardiologi	29	67	95
D	Nefrologi	17	78	95
E	Thoraxkirurgi	10	78	95
F	Urologi	19	64	95



Rapporten med tilbakemeldingene ble lagt ut på nett fire dager etter sensurmøtet, 7. april. Der kunne studentene finne sitt kandidatnummer og laste ned sin individuelle tilbakemelding. Tilbakemeldingene var tilgjengelige til 19. april. Etter denne datoen kunne skjemaene mottas via studiekonsulentene. Det viste seg at kobling mellom kandidatnummer og navn ved en inkurie ble kjent for noen studenter. Med denne hendelsen ble skjemaene umiddelbart fjernet. En anbefaling er at det i utviklingsplanen legges inn arbeid med en løsning for å gi individuell tilgang til disse tilbakemeldingene. Tilbakemeldingene må ikke kunne benyttes som en alternativ karakterutskrift.

Vurderinger

- Dersom sekvensielle oppgaver skal inngå bør basen ha en viss støtte for oppsett av dette i prøven og i fasit for å redusere risiko for eventuelle feilkilder som ligger i det som må ha manuell håndtering.
- I en fremtidig løsning er det en fordel med arkivfunksjon for besvarelse som del av eksamensløsningen og at tilbakemelding til studentene går direkte til hver enkelt, samtidig som tilbakemeldingene skal kunne benyttes som en alternativ karakterutskrift.
- Løsningen vi har valgt for tilbakemelding, sammen med publisering, av fast synes å ivareta den enkelte students behov for individuell tilbakemelding/begrunnelse.
- Vi anbefaler en nyutdannet lege fra hvert studiested engasjeres til en siste gjennomlesing av prøven slik den foreligger i eksamenssystemet som skal brukes, med henblikk på å oppdage feil i prøven i siste instans før gjennomføring.

6. De administrative sidene

Plassering av en ny nasjonal delprøve

Rammene for testen gitt av mandat var tidlig i 12. studiemester. Prøven ble plassert med god klaring til avsluttende eksamen og god klaring i forhold til forrige semesters konter. En runde med alle institusjonene viste at fredag 24. mars kl. 9-13 var mulig for alle å få til. Tidspunktet ble lagt inn i de lokale timeplanene.

Obligatorisk arbeidskrav

Prøven som et obligatorisk arbeidskrav ble ikke lagt inn i emnebeskrivelsene. Imidlertid ble det opprettet en obligatorisk aktivitet i FS ved alle studiesteder, slik at det var mulig å gi studentene kandidatnumre til prøven. Kandidatnummer var viktig med tanke på at studentene i etterkant ville få tilsendt riktig besvarelse og kunne finne riktig tilbakemeldingsskjema, da det i pilotfasen ikke var en løsning hvor dette kunne samles for hver enkelt.

Ved gyldig forfall til en obligatorisk prøve, for eksempel på grunn av dokumentert sykdom, må det arrangeres en utsatt prøve noen uker senere slik at studentene det gjelder får oppfylt kravene for å fremstille seg til avsluttende eksamen. Å arrangere en utsatt prøve innebærer at fagkomitéene må klargjøre ett oppgavesett til den ordinære prøven og ett oppgavesett til den utsatte prøven. Det vil ikke være nok tid til å lage et nytt oppgavesett i perioden mellom den ordinære og utsatte prøven. Administrativt må det klargjøres og avvikles en ekstra delprøve. De kan enten arrangeres samtidig, evt. over to runder (fordrer da flere oppgavesett – sånn sett faller kanskje et evt. forslag om to konter).

Gjennomføring

Totalt 319 (83 %) av 384 kandidater gjennomførte prøven. Oppmøtet fordelte seg slik på institusjonene (i prosent):

Institusjon	Oppmøte
UIB	80 %
UIT	88 %
UIO	94 %
NTNU	73 %
Totalt oppmøte	83 %

Innloggingsark med kandidatinformasjon ble laget ved UiO og innloggingsark for kandidater ved institusjonen ble sendt kontaktpersonene for hvert enkelt sted. Selve prøven ble gjennomført med utgangspunkt i rutine for eksamensgjennomføring. Studentene ble bedt om å møte 30 minutter før oppstart. De ble registrert inn på kandidatlistene og vanlige rutiner for eksamensavvikling ble brukt. Ved oppstart fikk studentene utdelt sitt innloggingsark, eller tildelt plass hvor innloggingsarket lå klart. Da bruk av innloggingsark ikke inngår i de lokale rutine ved lærestedene ble det litt ekstra arbeid med å distribuere dem ved oppstarten.

Prøven ble startet opp kl. 9.00 Ved NTNU ble oppstart forskjøvet med 30 minutter pga. forsinket oppmøte fra eksamensvaktene. Underveis i prøven holdt administrative kontinuerlig kontakt via en chat og et online møterom, i tillegg til telefonkontakt og SMS. Alle avvik eller

spørsmål ble rapportert inn via chatten og ble dokumentert i et loggskjema. Ved hvert sted var en faglærer tilgjengelig for å svare på spørsmål.

Prosess

Selv om piloten var en prøve ble de lokale eksamensapparatene brukt til gjennomføringen, dvs. at lokale eksamenskontorer stilte med eksamensvakter og IT-vakter underveis. De lokale administrative representantene koordinerte dette apparatet, mens sekretariatet hadde en koordinerende rolle ut i håndteringen av prøven ut mot alle institusjonene. Det er behov for begge disse rollene i forhold til å sikre samkjøring og enhet i gjennomføringen av en nasjonal prøve.

Lokaler

Ved samtlige studiesteder har lokaler til avvikling av digital eksamen vært et sentralt tema de siste årene. Ved UiO ble høsten 2016 brukt å komme på plass i et nytt eksamenslokale, herunder etablering av nye rutiner for eksamensavvikling. Dette arbeidet fortsetter i 2017. Med helt ferske eksamensvakter under gjennomføringen 24. mars 2017 ble det en del forvirring rundt opplegget. Ved UiB, NTNU og UiT har eksamenslokalene vært under ombygging i perioden hvor nasjonal delprøve skulle avvikles. Det var en stor utfordring å finne egnede rom og kvalitetssikre at rommene hadde strøm- og nettverkskapasitet til kandidatene.

Sikkerhet

I pilotarbeidet har fagkomiteene arbeidet seg i mellom med oppgavene. Utsendinger av oppgaver og annet konfidensielt materiale har etter vurdering blitt sendt enten via e-post eller sikker filsending. Et godt verktøy for samarbeid om eksamensoppgaver på tvers av institusjonene finnes per tid ikke. Vår vurdering er at bruk av FVO-basen er den løsningen som vil bidra til best sikkerhet i oppgavearbeidet

Vurderinger:

- Det var et godt samarbeid om det administrative arbeidet ved gjennomføringen av nasjonal delprøve.
- En nasjonal koordinatorfunksjon vil være nødvendig for å gjennomføre av fremtidige prøver.

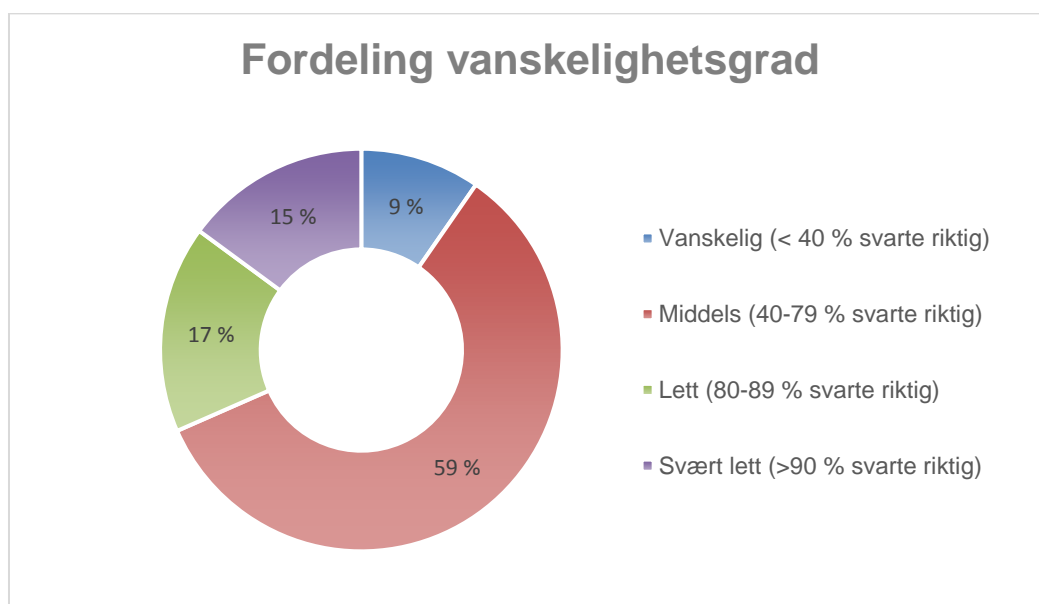
7. Oppgaveanalyse og resultater

Oppgaveanalyse og tilbakemeldinger

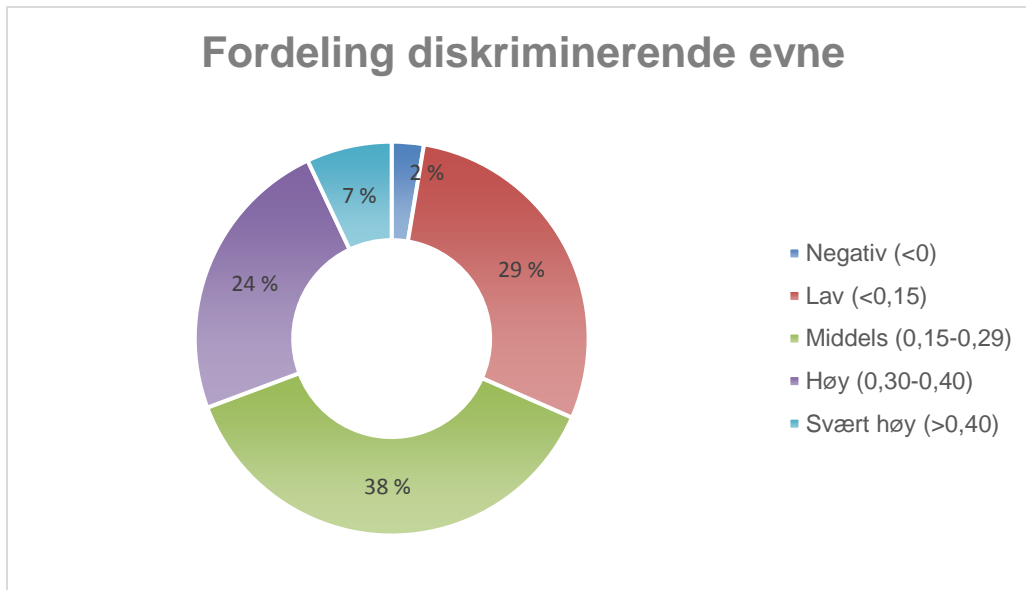
120 oppgaver fordelte seg på fagene som følger: gastroenterologi 20, gastrokirurgi 20, kardiologi 30, thoraxkirurgi 10, nefrologi 20 og urologi 20. 120 oppgaver var derfor gjenstand for oppgaveanalyse i etterkant av eksamen. For å si noe om prøvens reliabilitet (pålitelighet) har vi benyttet oss av Kuder-Richardson formula 20 (KR20). Dette er en test som sier noe om intern konsistens på prøven, og er et indirekte mål for prøvens reliabilitet. Desto høyere tallet er mellom 0 og 1, jo bedre er prøvens reliabilitet. For nasjonal delprøve i medisin 2017 var KR20 reliabiliteten 0,81. For en prøve med kun 120 spørsmål er dette en god reliabilitet, spesielt når man tenker på at dette er første gangen man gjennomfører en nasjonal delprøve i medisin i Norge. I internasjonal litteratur anbefaler man at man på viktige prøver som nasjonale prøver er, bør tilstrebe en reliabilitet over 0,90.

Oppgavene ble vurdert ut fra vanskelighetsgrad og diskriminerende evne. Med vanskelighetsgrad menes hvor mange kandidater som svarte rett på spørsmålet. Det vil si at en vanskelighetsgrad på 1 betyr at alle kandidatene svarte riktig på spørsmålet, mens en vanskelighetsgrad på 0 betyr at ingen besvarte spørsmålet riktig. Diskriminerende evne sier noe om hvor godt spørsmålet var til å skille mellom de 27 % av studentene med best skår på prøven fra de 27 % studentene med dårligst skår på prøven. Det vil si at jo høyere tall mellom 0 og 1 – jo bedre var spørsmålet til å diskriminere gode fra dårlig presterende studenter. En negativ diskriminerende evne forekommer når den dårligst presterende gruppen oftere svarer riktig enn den best presterende gruppen.

Det finnes ingen fasit på hva som er den ideelle vanskelighetsgrad eller diskriminerende evne for et spørsmål, men det er viktig at man har god spredning i spørsmålene, slik at man har en prøve som evner å skille studentene fra hverandre. Spørsmålene fordelte seg som følger når det gjaldt vanskelighetsgrad og diskriminerende evne på nasjonal delprøve:



Diagrammet viser god fordeling av spørsmål med ulik vanskelighetsgrad, med en overvekt av spørsmål med middels vanskelighetsgrad. Prøven kan ikke anses å ha vært lett. Hvis man sammenligner fordelingen med integrerte eksamener holdt ved NTNU er prøven vanskeligere. Ser man på historiske data fra eksamener ved NTNU vil rundt 50-60 % av spørsmålene på enhver eksamen lande i kategorien lett eller svært lett. Man skal være forsiktig med å si noe sikkert om vanskelighetsgraden av prøven sammenlignet med reelle avsluttende eksamener i medisinstudiet, da denne prøven ikke hadde noen konsekvenser for studentene og ble holdt på et tidligere tidspunkt i semesteret enn eksamener ellers. Studentene vil derfor ha lest forskjellig til nasjonal prøve og sin egen avsluttende eksamen.



Diagrammet viser en god fordeling i hvordan spørsmålene var til å diskriminere mellom studentene. Svært positivt var det at kun 2 % av spørsmålene hadde negativ diskriminerende evne, noe som ofte er et uttrykk for feil i oppgavene. Dette var også tilfellet i 2 av 3 oppgaver med negativ diskriminerende evne.

Når man vurderer oppgavekvalitet er det også viktig å vurdere hvorvidt distraktorene var gode. Med distraktorer menes de svaralternativene som ikke var riktig. På et spørsmål med høy kvalitet vil man lage gode distraktorer som er realistiske, men helt klart ikke det beste svaret. På nasjonal delprøve var det totalt 360 distraktorer og av disse var det kun 29 som aldri ble valgt av noen studenter, det utgjør kun 8 % av distraktorene. Dette er et meget lavt tall for spørsmål som ikke tidligere har vært brukt på en prøve og revidert med hensikt å gjøre distraktorene mer realistiske. Det betyr at oppgaveforfatterne har brukt tid på å utforme gode og realistiske distraktorer.

Tidsbruk

Vi har også vurdert tidsbruk på prøven. Studentene hadde god tid sammenlignet med det som er vanlig i nasjonale prøver, 120 sekunder/spørsmål mot 90 sekunder/spørsmål normalt. Ved en vurdering av tidspunkt for når studentene startet hver av de to delene og avsluttet disse, ser det ut til at få studenter brukte over 75 sekunder per spørsmål på de frittstående oppgavene og sjelden over 60 sekunder per sekvensiell oppgave. Vi mener derfor at studentene har hatt svært god tid på denne prøven.

Tilbakemelding på oppgaver

Studentene ble gitt mulighet til å gi tilbakemelding på oppgaver både under prøven og i etterkant gjennom et nettskjema. Man ble oppfordret til å gi tilbakemelding samlet fra hvert studiested, men åpnet også for at enkeltpersoner også kunne gi tilbakemelding. Det var åpnet for tilbakemeldinger i perioden 24. mars til og med 31. mars 2017. Det ble totalt gitt tilbakemeldinger på 40 oppgaver. Alle disse tilbakemeldingene ble vurdert i sensurmøtet.

Sensurering

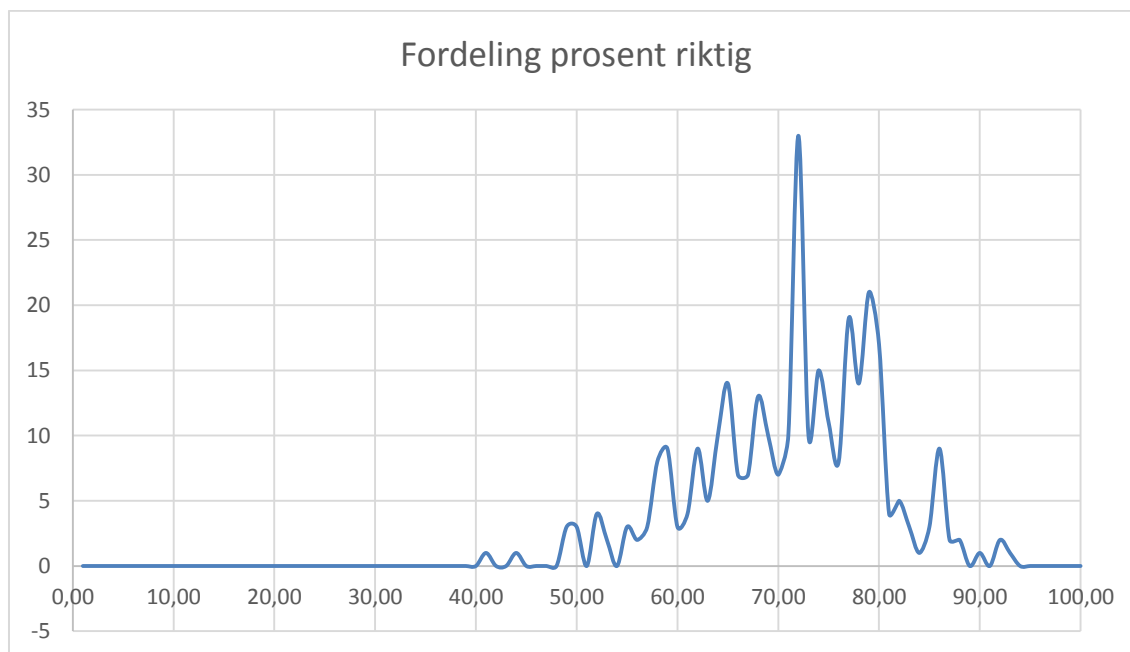
Prosjektgruppen fungerte som sensurkomité. I tillegg deltok lederne fra fagkomitéene samt studentrepresentantene⁸. Gitt at nasjonal delprøve ble gitt digitalt og med flervalgsoppgaver var det ikke behov for retting av besvarelser som sådan. Imidlertid er det fremdeles behov for en kvalitetssikring av resultater gjennom psykometri og faglige vurderinger. Psykometri og studentenes tilbakemeldinger la grunnlaget for diskusjonene i sensurmøtet. I alt sensureringsarbeid vil en sensurkomité eller kommisjon måtte vurdere vanskelige spørsmål eller ta stilling til tvilstilfeller. En ekstern sensor kan her gå inn i gi et upartisk synspunkt. Å tilknytte sensurarbeidet en ekstern sensor vil sikre legitimitet og øke kvaliteten på en nasjonal delprøve i medisin.

Endringer i fasit basert på oppgaveanalyse og studenttilbakemeldinger

Basert på lav eller negativ diskriminerende evne, diskuterte sensurkomitéen 7 spørsmål, de fleste av disse var også kommentert av studentene. Basert på vurdering i sensurkomitéen ble 6 spørsmål trukket fra prøven grunnet uklarhet, feil eller at spørsmålet forekom to ganger i oppgavesettet. 5 spørsmål fikk endret sin fasit. Av disse fikk ett spørsmål nytt riktig svar og på fire spørsmål godkjente man to svaralternativer som riktig. At 9 % av spørsmålene ble endret eller trukket fra prøven er å forvente på en helt ny prøve med utelukkende nye spørsmål som aldri har vært brukt før. På tross av grundig kvalitetskontroll innad i fagkomitéene og ved fagfellevurdering av enkelte spørsmål, vil feil forekomme. En del kunne nok vært unngått hvis fagfeller jobbet direkte i databasen slik at de også hadde oppdaget feil ved innlegging av spørsmål i databasen. Systemet er derfor laget slik at dette vil være mulig ved neste gjennomføring. Ved en senere prøve burde man hatt noen til å lese gjennom den endelige versjonen av prøven slik den forelå i eksamensverktøyet studentene skulle bruke. Dette kunne ha luket ut enkelte feil.

Resultater

319 kandidater avla prøven og deres skår fordelte seg fint mellom 41 og 93 prosent riktig. Basert på sensurmøtets vurdering av at 6 spørsmål skulle trekkes fra prøven, har sensur derfor vært basert på de 114 gjenværende spørsmålene på prøven. Snittskår var 70,7 %. Fordelingen i prosentandel var som vist under, og viser en god spredning:



Gjennomsnittsskår ved de ulike studiestedene varierte mellom 68,8 % og 73,2 %, med noe ulik skår innen de seks fagområdene på tvers av studiestedene. Ettersom oppmøteprosentene var forskjellige er det vanskelig å tolke om prøven viser reelle forskjeller mellom studiestedene. Fakultetene har fått detaljerte resultater som danner basis for benchmarking, videre analyse og oppfølging internt.

En prøve som er obligatorisk for alle studenter og evt. konsistente forskjeller over tid vil være nødvendig for å si noe om det er akademiske forskjeller i studentenes prestasjoner mellom de fire studiestedene.

Standardsetting

Årets nasjonale delprøvepilot var frivillig og det var ingen form for standardsetting involvert. Studentene fikk rapport om hvordan de gjorde det sammenlignet med kohort på eget universitet og landsgjennomsnittet. Vi har likevel gjort noen vurderinger på hva utfallet hadde blitt hvis man skulle sette en grense mellom bestått og ikke-bestått.

Per i dag bruker de fleste norske medisinske fakulteter forutbestemte absoluttgrenser som setter grensen mellom bestått/ikke-bestått eller ulike karakterer. Med utgangspunkt i det har vi beregnet antall stryk og strykrate for absolutt beståttgrenser mellom 50 og 70 %. Vi har i tillegg vurdert hva beståttgrensen ville ha blitt hvis man benytter seg av modifisert Cohens metode, en metode som tar høyde for prøvens vanskelighetsgrad og benytter studentene på 90-persentilen (på denne prøven hadde de 80 % riktig) som en referansegrense. K-verdi angir hvor mange prosent av poengsummen til 90-persentilen som blir beståttgrensen. Resultatene er presentert i tabellen under.

Metode	Beståttgrense	Antall stryk	Prosent stryk
Absolutt grense	50 %	6	1,9
Absolutt grense	55 %	17	5,3
Absolutt grense	60 %	42	13,2

Absolutt grense	65 %	84	26,3
Absolutt grense	70 %	128	40,1
Relativ mCohen (k=0,60)	48,4 %	5	1,6
Relativ mCohen (k=0,65)	52,5 %	12	3,8
Relativ mCohen (k=0,70)	56,5 %	22	6,9

Vurderinger:

- Resultatene var tilnærmet like mellom studiestedene og de små forskjellene som er vanskelig å tolke i lys av ulik oppmøteprosent ved de fire studiestedene.
- Man bør tilstrebe størst mulig reliabilitet og validitet på nasjonale prøver. Dette kan oppnås ved å øke testtiden og antall spørsmål på prøven. Vi anbefaler at man går ned på 90 sekunder per spørsmål og øker testtiden til 6 timer per prøve. Da vil man kunne enten dekke flere fag og/eller dekke de fagene man tester bredere, noe som sikrer høyere validitet. Testen vil da bestå av 240 spørsmål.
- Det er nødvendig å vinne mer erfaring med nasjonal delprøve, som et obligatorisk arbeidskrav, før det kan være aktuelt å innføre et krav til bestått.
- Prosjektgruppen fungerte som sensurkomité. Det vil være ønskelig å introdusere en ekstern sensor som kan involveres i studentenes tilbakemeldinger og som kan konsulteres på sensurmøtet.

8. Studentenes erfaringer

Prøven og dens formål

Alle studentene fikk tilsendt et evalueringsskjema som tok for seg prøvens innhold, gjennomføring og formål. Av de 319 studentene som gjennomførte prøven, svarte 152 (48 %) på skjemaet de fikk tilsendt. I tillegg til å gi tilbakemeldinger på prøven, har mange studenter gjort seg noen tanker om hva prøven bør være i fremtiden. Det synes å være en viss enighet om at en nasjonal delprøve i seg selv er et positivt tiltak (84 %), mens 68 % svarer at de ønsker at det skal jobbes mot en felles skriftlig avsluttende eksamen. Utfordringer som ulikheter mellom lærestedene, prøvens form og vektlegging, samt muligheten for at dette vil vise seg gjeldende i jobbsammenheng, synes å være fellesnevnerne i de kvalitative tilbakemeldingene. Enkelte tar også til orde for at en bør legge mer vekt på å enes om hva som faktisk forventes av medisinstudenter måneder før de skal ut i arbeidslivet. Skal de kunne vurdere differensialdiagnoser? Vurdere alvorlighetsgrad og hastegrad? Skal de legge en videre plan utover det som forventes av en allmennlege? Og hvor dyptgående kunnskap skal de ha om diagnoser, behandling og utredning som er forbeholdt spesialisthelsetjenesten?

Prøvens form

Studentene synes tidvis å være svært uenige i hvilken form de ønsker at prøven skal ha videre. Flere av tilbakemeldingene går på at de enten ønsker flere eller færre sekvensielle oppgaver, ett tillegg av kortsvarsoppgaver eller en endring i antallet flervalgsoppgaver. Her ser vi klare forskjeller på tilbakemeldingene fra de ulike lærestedene, og det kan virke som at studentene som er vant med de ulike formene også foretrekker disse. Den kvantitative tilbakemeldingen viser at nærmere halvparten (45 %) var svært fornøyde med de sekvensielle oppgavene, mens ca. en tredjedel (30 %) var like fornøyde med flervalgsoppgavene.

Det er også verdt å nevne at studentene har kritisert korrekturfeil, og flere har gjort et poeng av at antall pasientkasus er for høyt i prøvens flervalgsdel. Oppgavene oppfattes av flere som en ”wall of text/tekstvegg” som resulterer i utfordringer med konsentrasjonen. I tillegg var det én av oppgavene som kom med to ganger i prøvesettet, og én oppgave som hørte hjemme i sekvens havnet i flervalgsdelen.

Prøvens brukersnitt og praktiske gjennomføring

Tilbakemeldingene på prøvens brukersnitt viser stor enighet blant studentene, også på tvers av lærested. Programmet som ble brukt (Questionmark) ble oppfattet som komplisert og rotete. Utfordringene med programmet og en lite fordelaktig layout ser ut til å være en gjenganger i tilbakemeldingene fra studentene. Enkelte av bildene og EKG-ene virker å ha vært i for lav oppløsning, og man er ikke konsekvente i bruken av referanseområder og normalverdier. Det foreslås også at flervalgsspørsmålene deles opp over flere sider, slik at den visuelle opplevelsen blir bedre, samt at det blir enklere å beregne tid.

Når det kommer til den praktiske gjennomføringen av prøven påpeker flere at det ved enkelte studiesteder var utfordringer i eksamenslokalet knyttet til både bordplassering og tilgang på IKT-personell ved oppstarten av prøven. Flere mener at informasjonen i forkant av prøven både var mangelfull og at den kom noe sent. I tillegg forteller enkelte studenter at de opplevde byttet mellom flervalgsdelen og de sekvensielle oppgavene som en usikkerhetsfaktor.

Prøvens relevans og vanskelighetsgrad

Et viktig moment i tilbakemeldingene har vært at enkelte av spørsmålene er for ”spesialiserte”, altså at de retter seg mer mot hva en spesialist skal kunne, enn det en skal kunne forvente av en nyutdannet lege. Denne kritikken knyttes til dels opp mot at enkelte spørsmål har en form som bærer preg av å være ”hva tenker jeg på nå”-spørsmål, og dels ulikheter i hva som blir undervist på de forskjellige lærestedene. Det blir påpekt at prøven med hell kunne være bedre forankret i kliniske veiledere, og at dette burde bli opplyst om på forhånd. Vanskelighetsgraden oppfattes, med unntak av enkelte konkrete spørsmål, å være lagt på et godt nivå. Over halvparten av studentene (51 %) synes de hadde god tid, og flere takket også for en god prøve som var relevant og ga nyttig eksamenstrening.

Helhetsinntrykk og vurderinger:

Etter en gjennomgang av studentenes tilbakemeldinger fant vi flere helt konkrete tiltak som det bør jobbes med før en eventuell ny prøve gjennomføres, dette inkluderer:

- Bedre korrekturlesing
- Mer vekt på at spørsmålene tester hva en nyutdannet lege skal kunne, altså mindre grad av spesialistmateriale
- Velge en annen digital løsning, inkludert forbedre layouten, samt øke oppløsningen på bilder/EKG
- Sikre at eksamenslokalene og eksamensvaktene er satt i stand til å gjennomføre en slik prøve
- Hvis prøven skal belyse reelle forskjeller mellom de ulike lærestedene, så bør den også være obligatorisk.

Videre ser vi at det også er punkter der det er stor uenighet blant studentene i hva som er den beste løsningen, blant annet er det svært varierende tilbakemeldinger på hvorvidt prøven bør inneholde flere eller færre flervalgsoppgaver, sekvensielle oppgaver og/eller kortsvarsoppgaver. Dette er noe arbeidsgruppa må ta med seg i sin videre vurdering av nasjonal delprøve. Studentene har ellers kommet med mange positive tilbakemeldinger, og det synes rimelig å tolke tilbakemeldingene dithen at de synes dette har vært nyttig tidsbruk, selv i siste semester.

9. Videre arbeid

Vi foreslår gjennomføring av ny nasjonal delprøve i mars 2018 og at denne etableres som et obligatorisk arbeidskrav. Prosjektgruppen opprettholdes.

Det foreslås utvidelse med ytterligere to fagkomiteer, forslagsvis i to medisinske fagområder:

- lunge/endokrinologi/infeksjonsmedisin
- muskel-skjelett/ortopedi

Det vil gi en økning i antall spørsmål. Testtiden foreslås derfor utvidet til 4-5 timer.

Datoer for nasjonal delprøve 2018 vil være 13. april, noe forskjøvet pga. påsken som faller 29. mars til 2. april. Forutsatt at prosjektet videreføres til høst 2018 vil dato for delprøve ved UiB og UiO være 11. oktober.