



Tidsskriftet

DEN NORSKE LEGEFORENING

Eksamensformer i de norske medisinstudiene

ORIGINALARTIKKEL

OLE FARESTVEDT RAAHEIM*

ole.fr@hotmail.com

Det medisinske fakultet

Universitetet i Bergen

Forfatterbidrag: ideutvikling, datainnsamling, gjennomføring og transkribering av intervjuer, analyser, utarbeiding av manuskriptet.

Ole Farestvedt Raaheim er medisinstudent på femteåret.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

DAG FARESTVEDT RAAHEIM*

Det medisinske fakultet

Universitetet i Bergen

Forfatterbidrag: ideutvikling, datainnsamling, gjennomføring og transkribering av intervjuer, analyser, utarbeiding av manuskriptet.

Dag Farestvedt Raaheim er medisinstudent på femteåret.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

HÅKON REIKVAM

Klinisk institutt 2

Det medisinske fakultet

Universitetet i Bergen

og

Seksjon for blodsjukdommar

Medisinsk klinikk

Haukeland universitetssjukehus

Forfatterbidrag: analyse av kvantitative data, utforming av figurer, drøftinger og innspill til utformingen av manuskriptet.

Håkon Reikvam er spesialist i indremedisin og blodsykdommer og professor. Han sitter i programutvalget for medisinstudiet ved Universitetet i Bergen.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

MONIKA KVERNENES

Enhet for læring

Det medisinske fakultet

Universitetet i Bergen

Forfatterbidrag: ideutvikling, drøftinger rundt prosjektets avgrensning, innspill til intervjuguide, aktivt bidrag i prosjektets siste fase med selve manuskriptet.

Monika Kvernenes er førsteamanuensis i medisinsk pedagogikk og enhetsleder.

Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

* Ole Farestvedt Raaheim og Dag Farestvedt Raaheim har bidratt i like stor grad til denne artikkelen.

BAKGRUNN

Det finnes ingen samlet oversikt over hvilke eksamensformer som er i bruk i de norske medisinstudiene. I denne studien har vi kartlagt hvilke eksamensformer som er i bruk, og analysert studieledelsenes begrunnelser for hvorfor nettopp disse er valgt. Målet har vært å få en samlet oversikt samt å sette tenkning om vurderingspraksisen i norsk grunnutdanning i medisin inn i en internasjonal og forskningsbasert ramme.

MATERIALE OG METODE

Vi samlet informasjon om eksamensformene som brukes i de fire norske profesjonsstudiene i medisin. Videre gjennomførte vi semistrukturerte intervjuer med en representant fra studieprogramledelsen ved hvert av studiestedene. Intervjuene omhandlet styrker og svakheter ved de valgte eksamensformene.

RESULTATER

Det var til dels store forskjeller i antall eksamener, deres form og deres plassering i studieløpet i de fire studieprogrammene. Flervalgsoppgaver var den dominerende formen, supplert av ulike muntlige og praktiske eksamener. Argumentasjon for valg av de ulike eksamensformene var i stor grad knyttet til behov for standardisering, objektivitet og ressurs hensyn.

FORTOLKNING

Vurdering synes i stor grad å ha som hensikt å måle og kontrollere, mens mulighetene eksamen gir for å drive læring, i mindre grad utnyttes. Systemtenkningen rundt vurderingsformenes ulike formål, som løftes i internasjonale retningslinjer, synes i mindre grad å være en prioritet.

HOVEDFUNN

Flervalgsoppgaver dominerte som eksamensform i de norske medisinstudiene.

Valg av eksamensform syntes i stor grad å være basert på idealer om objektivitet og standardisering samt ressurs hensyn.

Eksamen syntes først og fremst å ha som hensikt å måle og kontrollere, mens muligheten for å bruke eksamen som drivkraft for læring i mindre grad ble utnyttet.

De ulike eksamensformene ble ansett å ha ulike bruksområder, men syntes i mindre grad å bli knyttet sammen for å gjøre robuste helhetsvurderinger av studentene.

Vurderingsformene i de norske medisinstudiene blir stadig debattert (1-3). Vurdering er en samlebetegnelse på all form for prøving av studenters kompetanse. Hensikten er både å sikre forsvarlig kunnskaps- og ferdighetsnivå og å påvirke studentenes læringsprosess (4-6). Ulike ekspertutvalg har utarbeidet kriterier for å bestemme hva som kan anses å være en god vurderingsform. Disse oppsummeres i stikkordene validitet, reliabilitet, ekvivalens, gjennomførbarhet, pedagogisk effekt, katalytisk effekt og troverdighet (7) (ramme 1).

Ramme 1 Kriterier for god vurderingspraksis (7).

Validitet – at vurderingsformen tester det den er ment å teste

Reliabilitet – at resultatet vil være det samme ved gjentatte forsøk under samme betingelser

Ekvivalens – at den samme vurderingen gir like resultater på tvers av lærestedene

Gjennomførbarhet – at vurderingsformen er realistisk og praktisk gjennomførbar innenfor gitte rammer og ressurser

Pedagogisk effekt – at vurderingsformen bidrar til at studentene forbereder seg på en måte som gir læringsutbytte

Katalytisk effekt – i hvilken grad vurderingen motiverer for videre læring, også etter at vurderingen er gjort

Troverdig – at ulike interessenter (studenter, læresteder, pasienter, arbeidsgivere) har tillit til vurderingssystemet

Historisk har disse kriteriene vært tillagt ulik vekt, og det har blitt foreslått at medisinsk-pedagogisk utvikling fra 1960-tallet og frem til i dag har gått gjennom tre epoker (8). Den første omtalte «vurdering som måling». Her var man opptatt av å lage så objektive og standardiserte tester som mulig, tester som skilte godt mellom gode og dårlige prestasjoner. Psykometriske metoder ble benyttet i jakten på de optimale vurderingsformene (9). Den neste perioden, som vi ser fra 1990-tallet, introduserte «vurdering som dømmekraft». Her hadde skjønnsmessige vurderinger, eksaminatorens kompetanse til å vurdere og vurderingens tilbakemeldingsfunksjon sentrale roller.

Drevet av erkjennelsen av at det ikke finnes optimale vurderingsformer, vant tenkning om «vurdering som system» i økende grad terreng rundt 2000-tallet. Argumentene omhandlet behovet for flere målepunkter og helhetstenkning rundt vurderingene, fremfor å teste isolerte kunnskaper og ferdigheter (8–12). Dette samsvarer med det som omtales som programbasert vurdering (eng. *programmatic assessment*) (13). Schuwirth og van der Vleuten legger vekt på at det ikke er en klar avgrensning mellom de ulike epokene, og at vi finner tenkning fra alle igjen i dagens vurderingspraksiser (8).

Til tross for at forskningslitteraturen om temaet vurdering er svært omfattende, finnes det ingen gode oversikter over hvilke vurderingsformer som brukes i de fire norske medisinstudiene, eller analyser som ser norsk vurderingspraksis i et forskningsbasert perspektiv. I denne studien har vi samlet informasjon om hvilke eksamensformer som brukes i profesjonsstudiene i medisin ved Universitetet i Oslo (UiO), Universitetet i Bergen (UiB), Norges teknisk-naturvitenskapelige universitet (NTNU) og UiT Norges arktiske universitet (UiT). I tillegg har vi undersøkt hvordan valg av eksamensform begrunnes. Vi har hatt hovedfokus på de summative vurderingsformene, det vil si de tellende eksamenene. Formativ vurdering, som innebærer fortløpende tilbakemeldinger og vurdering hvis hensikt er å gi retning til videre læring, er i langt mindre grad formalisert i studieplanene og har derfor kun indirekte blitt tematisert. Hensikten med studien har vært å sette tenkning om norsk vurderingspraksis inn i en større internasjonal og forskningsbasert ramme.

Materiale og metode

Denne kvalitative studien, bestående av dokumentgjennomgang og analyser av semistrukturerte intervjuer med representanter fra studieprogramledelsene, ble gjennomført høsten 2021. Dokumentgjennomgangen tok for seg de summative vurderingene. Informasjon om hvilke eksamensformer som brukes ved de ulike studiestedene, ble hentet fra universitetenes hjemmesider og supplert med opplysninger vi fikk ved å kontakte studieadministrasjonene via e-post.

Vi sendte e-post til landets fire studieprogramledere i medisin, med forespørsel om å stille til et individuelt intervju om vurderingsformer brukt i studiet. Ved UiT ble vi videresendt til assisterende studieleder. Alle ble orientert om studiens hensikt før de samtykket til å delta. Intervjuobjektene omtales i det videre som «studieprogramledere». Intervjuguiden bestod av ni spørsmål, og intervjuene hadde en varighet på ca. én time (ramme 2). Intervjuene dreide seg i stor grad om summative vurderingsformer, mens formative vurderingsformer kun ble indirekte tematisert. Resultatet fra dokumentgjennomgangen

ble lagt frem og validert i forbindelse med intervjuene. Tre av intervjuene ble gjennomført som videosamtale, ett med fysisk oppmøte. Det ble gjort lydopptak av intervjuene som dernest ble transkribert. Vi gjorde deretter en deduktiv tematisk analyse av det transkriberte materialet med den hensikt å kategorisere studieprogramledernes refleksjoner om vurdering og de ulike vurderingsformenes styrker og svakheter (14,15). Vi brukte de tre epokene i vurderingstenkning som et teoretisk rammeverk i analysen. Etter å ha drøftet hvordan kjennetegn ved de ulike epokene kunne eksemplifiseres i intervjuene, gjennomførte to av oss (OFR, DFR) analysene i gjentatte runder med sisteforfatter (MK) som diskusjonspartner og medforsker, til vi nådde konsensus. Studien ble fremlagt og godkjent av personvernombudet for forskning, Norsk senter for forskningsdata (NSD, prosjektnummer 770770).

Ramme 2 Intervjuguiden som ble brukt i studien.

Hvilke vurderingsformer brukes ved studieprogrammet i medisin?

- Formative?
- Summative?
- Hva innebærer en skriftlig eksamen?
- Hva innebærer en muntlig/praktisk eksamen?

Kan du si litt om hvorfor dere har akkurat disse vurderingsformene?

- Pålitelighet
- Gyldighet
- Innvirkning på fremtidig læring og praksis
- Aksept fra studenter og fakultet
- Kostnader

Hvis du ser på de ulike vurderingsformene dere bruker – hva er styrkene og svakhetene ved de enkelte metodene?

Hvordan opplever studentene disse vurderingsformene?

- Hva sier de om dem? Klager? Synspunkter som formidles dere?

Hvis dere ikke trengte å ta hensyn til ressursbruk – hva tenker du ville vært den ideelle måten å drive vurdering på?

Har det vært diskutert eller prøvd ut andre vurderingsformer enn de som er i bruk i dag? I så fall, hva er grunnen til at det ikke har blitt noe av?

Bruker dere graderte karakterer? Hvorfor / hvorfor ikke?

Det er mange som har meninger om vurdering og eksamen – hvilke diskusjoner har dere hatt i programstyret eller andre fora om dette? Er det noen tilbakevendende diskusjoner, uenigheter eller sterke krefter som vil det ene eller det andre?

Er det noe mer du vil si om temaet vurdering før vi avslutter? Noe du vil legge til?

Resultater

EKSAMENSFORMER VED DE FIRE LÆRESTEDENE

På bakgrunn av det som kom frem i intervjuene og dokumentanalysen, lagde vi følgende seks eksamens kategorier: flervalgsoppgaver (+ kortsvar), mappevurdering, essayeksamen, hovedoppgave, muntlig/klinisk og objektiv strukturert klinisk eksamen (OSKE).

Disse eksamensformene kan også grovt inndeles i skriftlige og muntlige eksamener (ramme 3). Det totale antallet eksamener gjennom studiet varierte fra 11 (UiT) til 22 (UiO). Antall skriftlige eksamener varierte fra 7 (UiT) til 15 (UiB) (figur 1).

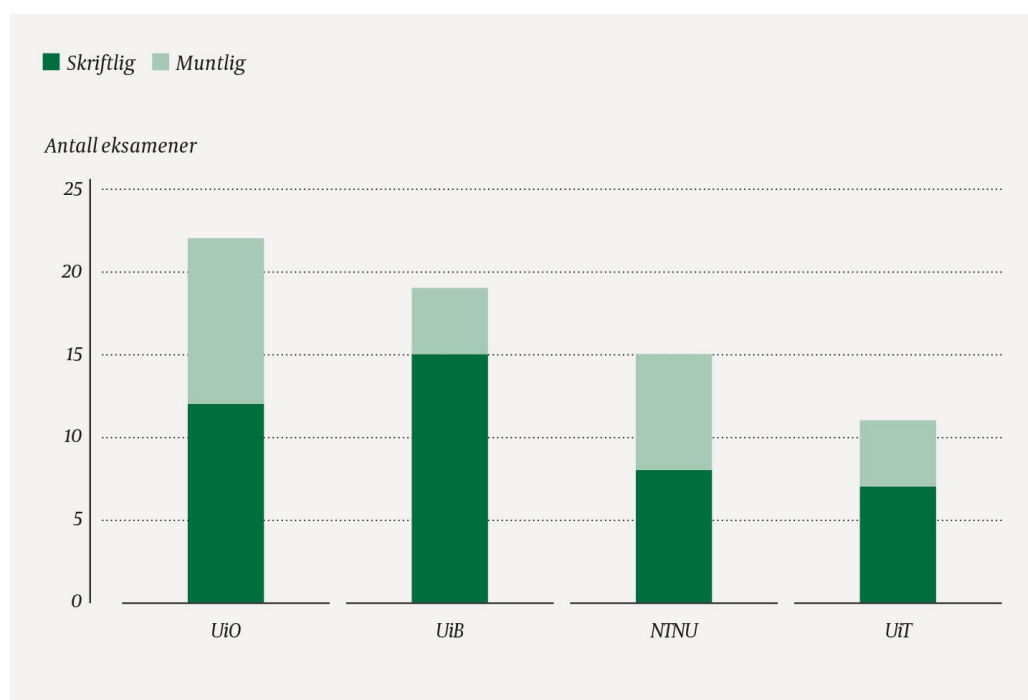
Ramme 3 De ulike eksamensformene som ble identifisert ved gjennomgang av studieplanene ved de ulike lærestedene høsten 2021.

SKRIFTLIGE

- Flervalgsoppgaver (+ kortsvar) Består enten utelukkende av flervalgsoppgaver eller en kombinasjon av flervalgsoppgaver og kortsvarsoppgaver. Den mest brukte varianten av flervalgsoppgaver er *single best answer*, der kun ett av totalt fire alternativer er riktig. Kortsvarsoppgaver er fritekstopp-gaver der man svarer med en kort tekst.
- Mappevurdering Studenten leverer inn arbeid underveis i semesteret i en «mappe» og får tilbakemeldinger underveis (formativ vurdering), før en eventuell endelig innlevering med bestått / ikke bestått (summativ vurdering).
- Essayeksamen Består utelukkende av fritekstopp-gaver, ligner kortsvarsoppgaver. En misvisende term, da essay i denne sammenhengen henviser til en kort, presis og konsis tekst, og ikke til det en kjenner som essay som sakprosa-sjanger (3).
- Hovedoppgave En større skriftlig oppgave som strekker seg over flere semestre, der hensikten er at studentene skal anvende en vitenskapelig arbeidsmetode. Studenter og veiledere foreslår selv tema til oppgave.

MUNTLIGE

- Muntlig/klinisk Med eller uten pasient til stede og med eller uten klinisk tilnærming. Eksempler kan være anatomi-eksamen eller en mer klinisk rettet case-eksamen i øyefaget.
- Objektiv strukturert klinisk eksamen (OSKE) En standardisert muntlig eller praktisk eksamen med et bestemt antall stasjoner som studenten skal gjennom.



Figur 1 Antall skriftlige og muntlige eksamener ved de ulike lærestedene, basert på gjennomgang av studieplanene for høsten 2021. UiO = Universitetet i Oslo, UiB = Universitetet i Bergen, NTNU = Norges teknisk-naturvitenskapelige universitet, UiT = UiT Norges arktiske universitet.

Tabell 1 viser *når* i studieløpet de skriftlige og muntlige eksamenene ble avholdt. I motsetning til de skriftlige eksamenene, som ble gjennomført regelmessig gjennom hele studiet, var det flest muntlige eksamener i siste halvdel av studiet. Ved UiT ble det ikke

gjennomført noen muntlige eksamener før på sjetteåret. UiB hadde muntlig eksamen på tredje-, fjerde- og sjetteåret. UiO og NTNU gjennomførte muntlig eksamen også tidligere i studieforløpet.

Tabell 1

Tidspunkt i studieløpet antall skriftlige og muntlige eksamener avholdes, basert på gjennomgang av studieplanene høsten 2021. S = skriftlig eksamen, M = muntlig eksamen, UiO = Universitetet i Oslo, UiB = Universitetet i Bergen, NTNU = Norges teknisk-naturvitenskapelige universitet, UiT = UiT Norges arktiske universitet.

	Førsteåret		Andreåret		Tredjeåret		Fjerdeåret		Femteåret		Sjetteåret	
	S	M	S	M	S	M	S	M	S	M	S	M
UiO	3	-	1	2	2	2	3	2	2	2	1	2
UiB	3	-	2	-	2	1	2	2	3	-	3	1
NTNU	1	1	1	1	1	1	2	2	1	-	2	2
UiT	1	-	1	-	1	-	2	-	1	-	1	4

En mer detaljert oversikt over eksamensformer ved de ulike lærestedene er samlet i tabell 2. Flervalgsoppgaver var dominerende ved alle studiestedene. UiB hadde som eneste studiested en egen essayeksamen. I tillegg skilte UiB seg ut ved å ha mappevurdering på femte- og sjetteåret. Som avsluttende eksamen ble det avholdt en nasjonal deleksamen felles for alle lærestedene, som besto utelukkende av flervalgsoppgaver. Av muntlige vurderingsformer var både muntlige/kliniske eksamener og OSKE-eksamen i bruk. UiT var det eneste lærestedet som ikke hadde OSKE-eksamen som summativ vurdering, men kun som obligatorisk arbeidskrav på tredjeåret.

Tabell 2

Antall eksamener med ulike eksamensformer ved de fire lærestedene, basert på gjennomgang av studieplanene høsten 2021. OSKE = objektiv strukturert klinisk eksamen, UiO = Universitetet i Oslo, UiB = Universitetet i Bergen, NTNU = Norges teknisk-naturvitenskapelige universitet, UiT = UiT Norges arktiske universitet.

	Flervalgsoppgaver (+ kortsvar)	Muntlig/klinisk	OSKE	Mappevurdering	Essay	Hovedoppgave ¹
UiO	11	8	2	-	-	1
UiB	11	2	2	2	1	1
NTNU	7	3	4	-	-	1
UiT	6	4	-	-	-	1

¹Kalles masteroppgave ved UiT.

BEGRUNNELSER FOR VALG AV EKSAMENSFORM

Flervalgsoppgaver

Bruken av flervalgsoppgaver ble begrunnet blant annet med at vurderingsformen er utbredt internasjonalt, mens noen pekte på at «det bare har blitt sånn». Styrker som ble trukket frem, var at flervalgsoppgaver gir mulighet for å teste bredt, har høy reliabilitet og er objektive og rettferdige. Dessuten ble slike oppgaver ansett som «en enkel eksamen å bedømme»; oppgavene rettes automatisk, og det kreves lite ressurser. Samtidig hadde alle tanker rundt formens svakheter, som at det kan være utfordrende å lage adekvate oppgaver, og at oppgaveformen i liten grad tester studentenes resonnering, holdninger, ferdigheter og generelle kompetanse.

Noen uttrykte bekymring for om flervalgsoppgaver har god nok diskriminerende evne, det vil si om grunnlaget for å vurdere om studenter skal bestå eller stryke på eksamen, er godt nok. I tillegg ble gjenbruk av oppgaver og det at studentene kan gjette seg til riktig svar, trukket frem som mulige problemer. En av studieprogramlederne mente at flervalgsoppgaver alene som vurderingsform «ikke vil være en fullgod testing uten at du da i tillegg har en klinisk eksamen».

Kortsvarsoppgaver

Blant fordelene ved kortsvarsoppgaver sa en av informantene at «de tvinger studenten til å vise at han eller hun kan analysere en situasjon», og at «oppgavene krever resonnering, og studentene må presentere svarene aktivt fra egen hukommelse». En av studieprogramlederne pekte på at kortsvarsoppgaver på denne måten gir en tilleggsverdi til flervalgsoppgaver, og at de til en viss grad kan kompensere for begrensningene ved flervalgsoppgaver. En av de andre studieprogramlederne var imidlertid usikker på om oppgavene har noen betydning i bedømmelsen av hvorvidt noen skal stryke eller stå på eksamen: «Vi bruker en god del ressurser på å sensurere dette. Og den innsatsen tror ikke jeg er balansert mot betydningen av kortsvar.»

Å differensiere studentene fra hverandre ble av flere trukket frem som et av målene med eksamen, og en av studieprogramlederne beskrev det som «et nederlag» dersom ingen stryker. Ulemper som ble trukket frem, var at vurderingsformen er ressurskrevende, at den er krevende å rette, og at sensuren vil påvirkes av subjektive vurderinger. Noen trakk frem at det var uheldig at andelen kortsvarsoppgaver var blitt redusert de siste årene, samtidig som flervalgsoppgaver hadde fått en stadig større plass.

Objektiv strukturert klinisk eksamen (OSKE)

Programlederne anså OSKE-eksamen som en allsidig vurderingsform som tester en variert bredde av basalkunnskap, praktiske ferdigheter og kommunikasjonsferdigheter på en standardisert og objektiv måte. Flere pekte på at OSKE-eksamen er anerkjent som en god vurderingsform med høy reliabilitet – «OSKE er rettferdig og reproduserbar, og det opplever studentene også».

Noen la vekt på at OSKE-eksamen er et godt verktøy for å stryke studenter som bør strykes. Studieprogramlederne var samstemte om at vurderingsformen er ressurskrevende, noe som ble trukket frem som forklaring på hvorfor UiT foreløpig har OSKE-eksamen kun som et obligatorisk arbeidskrav, og ikke som tellende eksamen. Noen pekte på at bruken av skuespillere, som ofte benyttes som pasientkasus (markører) ved ulike stasjoner, kan gi kunstige kliniske problemstillinger.

Muntlige/kliniske eksamensformer

Begrunnelsen for bruk av muntlige/kliniske vurderingsformer var at det er viktig at studenten evner å analysere og håndtere en større mengde informasjon og sette det i sammenheng: «Den kliniske eksamen skal jo teste nettopp måten å fungere på som lege, ved å se deg i relasjon til pasient og også den praktiske utførelsen av det.» Det ble trukket frem at vurderingsformen gir mulighet for å teste studentene bredt med en helhetlig vurdering av deres ferdigheter, kunnskaper og profesjonelle kompetanse.

Det ble imidlertid også pekt på at studentene kan ha flaks eller uflaks med hvilke oppgaver som blir gitt: «Du har litt flaks/uflaks om du får akkurat den sykdommen i leveren som du kan.» Av andre svakheter ble det vektlagt at denne eksamensformen ikke er standardisert på samme måte som skriftlig eksamen og OSKE-eksamen, samt at den er ressurskrevende å gjennomføre.

Mappevurdering

UiB var eneste lærested som brukte mappevurdering som vurderingsform. Begrunnelsen for bruken av denne vurderingsformen var at «det ligger litt i tiden, men den viktigste grunnen er at studentene er aktive gjennom hele semesteret». En styrke som ble trukket frem, var at mappevurdering gir rom for varierte typer oppgaver, mens det kan anses som svakhet at den kan være ressurskrevende å følge opp.

Hovedoppgave

Refleksjoner rundt styrker og svakheter ved denne vurderingsformen ble ikke diskutert i intervjuene, og omtales derfor ikke nærmere.

Vurderingspraksis sett under ett

Helhetlig sett ga informantene uttrykk for at de var tilfredse med vurderingsformene som brukes i eget studium. Ingen ga uttrykk for at de ønsket å endre på de summative vurderingene slik de er i dag. Imidlertid var flere tydelige på at innholdet og kvaliteten, spesielt på flervalgsoppgavene, har forbedringspotensial. Studieprogramlederne argumenterte for at fremfor å sette i gang et større reformarbeid av eksamensregimet, er ressursene bedre anvendt i forbedring av undervisningen og innarbeiding av mer formativ vurdering. Flere pekte på at de har en vei å gå når det gjelder å gi tilbakemeldinger til studentene gjennom studiet.

Diskusjon

Vår studie viser at flervalgsoppgaver er den dominerende eksamensformen i medisinstudiet ved de fire lærestedene, supplert av ulike muntlige og praktiske vurderingsformer. Studieprogramlederne syntes å anerkjenne at ulike vurderingsformer har ulike bruksområder, mens den overordnede argumentasjonen for valg av eksamensform domineres av standardisering, rettferdig eller objektiv vurdering og ressurs hensyn. Reliabilitet og ekvivalens synes å bli brukt som de overordnede kriteriene. I tillegg demonstreres kriteriet om gjennomførbarhet ved at studieprogramlederne viste til flervalgsoppgavers egenskaper som selvrettende og gjenvinnbare. Dette er tankegodt som passer med epoken «vurdering som måling» skissert av Schuwirth og van der Vleuten (8).

Vi fant et syn på vurdering hvor subjektivitet omtales som noe en ønsker å minimere. Imidlertid har det psykometriske fokuset i arbeidet med vurdering fått kritikk fra ulike hold. Særlig fremheves faren for forsterkning av et reduksjonistisk syn på kommunikative og sosiale ferdigheter, som det er vanskeligere å lage objektive målbare kriterier for. Brian Hodges er blant dem som har tatt til orde for at vi derfor bør tillate subjektivitet, i betydningen kvalifiserte skjønnsmessige vurderinger, å spille en sentral rolle i vurderingen av studenter (9). Han viser til van der Vleuten og Schuwirths funn (12) om at det ikke er standardisering som er avgjørende for en tests reliabilitet, men andre faktorer som total testtid og antall vurderinger. Ifølge Hodges kan flervalgsoppgaver, essay, tradisjonelle muntlige eksamener, OSKE-eksamen og andre metoder alle oppnå høy reliabilitet så lenge eksamenstiden er lang nok (9).

Bruken av muntlige/kliniske eksamener kan plasseres i epoken «vurdering som dømmekraft». At andelen kortsvarsoppgaver reduseres, ble av noen trukket frem som uheldig. Dette er et syn som også kan plasseres i samme epoke. Flere studier peker på utfordringene med å la flervalgsoppgaver dominere eksamenene (16). Sammenlignet med

fritekstoppgaver stimulerer flervalgsoppgaver til overflatelæring med fokus på memorering, mens avanserte ferdigheter som refleksjon, skriftlig formuleringsevne, resonnering og kritisk dømmekraft nedprioriteres (10, 16, 17).

Moderne tilnærminger til vurdering er opptatt av å vurdere studentenes kompetanse helhetlig og ved hjelp av varierte kilder og måleinstrumenter (8). En av våre informanter var opptatt av å bruke flere metoder for å veie opp for enkeltmetoders begrensninger. Ellers syntes det helhetlig sett å være mindre fokus på å se de ulike vurderingsformenes styrker og svakheter i kombinasjon. Nettopp trianguleringen av metoder, som vil si å trekke en beslutning basert på flere vurderingsformer, er et sentralt kjennetegn i systemtenkningen. Ingen av informantene i vårt materiale nevnte begrepet programbasert vurdering (13), ei heller ble det ytret ønske om å innføre for eksempel vurderingspanel som kan ta helhetlige beslutninger om bestått, stryk eller karakter fremfor å la hver enkelt deleksamen telle for seg.

Ønsket om mer formativ vurdering kan imidlertid tyde på at en ser viktigheten av å styrke vurderingspraksisene i retninger som bedre understøtter studentenes læring. Imidlertid kan ønsket om å beholde de summative vurderingene og samtidig videreutvikle de formative, indikere et manglende blikk for hvordan formativ og summativ vurdering henger sammen og påvirker hverandre. For eksempel vil økt bruk av formativ vurdering redusere behovet for summativ vurdering, slik nederlandske forskere har vist (18, 19).

Det er også verdt å ta med at studentene i mindre grad er like tilfredse med dagens vurderingspraksis. I studiebarometeret til Nasjonalt organ for kvalitet i utdanningen (NOKUT) viser ferske tall at alle studieprogrammene i medisin skårer lavt når det gjelder temaene tilbakemeldinger og vurderingsformer, med indeksverdier på hhv. 2,7 og 3,7 (på en skala fra 1 til 5, der 5 er best) (20). Til sammenligning er snittskåren for alle utdanninger i landet hhv. 3,3 og 4,0 (21).

At studentene synes å være mindre tilfredse med vurderingsordningene enn studieprogramlederne, tyder på at studenter og studieledelse ikke tenker likt om hva som kjennetegner gode vurderingsformer. Både studenter og undervisere har uttrykt seg kritisk til dagens vurderingspraksis der flervalgsoppgaver er så sterkt dominerende (13). Bekymringen handler blant annet om at flervalgsoppgaver inviterer til pugg, og at studenter kan gjøre det bra på eksamen, men samtidig ha relativt store kunnskapshull (1, 2). Studentene selv forteller hvordan de kan gjette seg til riktige svar, kjenne igjen oppgaver fra tidligere gitte eksamener og strategisk velge bort å lese på enkelte mindre fag, men samtidig få en god nok skår til å bestå eksamen (1, 2). Det er også uttrykt bekymring for at flervalgsoppgaver foretrekker eksamensformer som gir studentene trening i å skrive og uttrykke seg, da presis og riktig bruk av medisinske faguttrykk og terminologi er essensielt videre i legeyrket (22). Dette utgjør sterke argumenter for å øke bruk av fritekstoppgaver.

En rekke ytre drivkrefter kan tenkes å bidra til å forklare disse utviklingstrekkene. Mangel på ressurser, lite kunnskap om alternative vurderingsformer og strenge regelverk for eksamensavvikling kan være mulige forklaringer på hvorfor oppfatningene om eksamen ikke synes å være oppdaterte i henhold til internasjonal forskning. I tillegg kan flyttingen av autorisasjonspunkt for legene fra etter turnus (nå LIS1-stilling) til etter endt grunnutdanning ha bidratt til behovet for økt kontroll. En kunne tenke seg at innføring av nasjonal delprøve (23, 24), kunne bli brukt som anledning både til å ivareta kontrollfunksjonen og samtidig åpne for økt bruk av pedagogisk stimulerende og triangulerende vurderingsformer underveis i studiet.

STUDIENS BEGRENSNINGER OG STYRKER

Funnene i denne studien er basert på dokumentgjennomgang samt intervjuer med studieprogramlederne. Det er gjort et strategisk utvalg, og vi vet ikke om andre nøkkelpersoner eller undervisere ved studiestedene har andre tanker om vurdering og eksamensformer enn de som har fremkommet her. Vi har ikke samlet kunnskap om

formative vurderingsformer, noe som ville krevd andre informanter med tettere kontakt med studentene gjennom undervisning og veiledning sammenlignet med studieprogramlederne.

Studiens styrke ligger først og fremst i at den bidrar med en systematisk oversikt over eksamensformer som er i bruk i de norske medisinstudiene, og at begrunnelsene for valg av eksamensform er satt inn i en forskningsbasert ramme.

KONKLUSJON

Det var relativt store forskjeller mellom lærestedene når det gjelder antall eksamener og fordelingen mellom muntlige og skriftlige eksamensformer. Flervalgsoppgaver er i utstrakt bruk ved alle studieprogrammene, mens kortsvarsoppgaver og andre former for fritekstoppgaver er lite utbredt. Tankesettet rundt temaet vurdering synes likevel å være nokså likt ved de fire studieprogrammene, sett fra studieprogramledernes perspektiv. Alle har refleksjoner rundt styrker og svakheter ved de ulike eksamensformene og anerkjenner at ulike eksamensformer har ulike bruksområder. Likevel: Når vi ser på hvilke eksamensformer som er mest utbredt, og hvilke kriterier som i størst grad vektlegges, synes eksamenspraksisen i stor grad å være drevet av behovet for måling og kontroll, mens mulighetene eksamen gir for å drive læring, i liten grad utnyttes. Det er behov for mer forskning på temaet vurdering i norsk medisnutdanning generelt og på bruken og virkningene av formative vurderingsformer spesielt.

Forfatterne takker Harald Wiker (Universitetet i Bergen), Maja-Lisa Løchen (UiT Norges arktiske universitet), Ivar Skjåk Nordrum (NTNU) og Elin Olaug Rosvold (Universitetet i Oslo) for deres velvilje til å bidra med opplysninger som danner grunnlaget for denne artikkelen.

Intervjuobjektene har gitt samtykke til å bli identifisert i artikkelen.

Artikkelen bygger på en studentoppgave av Ole Farestvedt Raaheim og Dag Farestvedt Raaheim ved Det medisinske fakultet ved Universitetet i Bergen.

Artikkelen er fagfelleverdert.

REFERENCES

1. Wessel J. Eksamenseksperter med svake kunnskaper. Tidsskr Nor Legeforen 2021; 141. doi: 10.4045/tidsskr.20.0946. [PubMed][CrossRef]
2. Steen VM. Fritt frem for kunnskapssvake medisinstudenter. Tidsskr Nor Legeforen 2020; 140. doi: 10.4045/tidsskr.20.0733. [PubMed][CrossRef]
3. Gjersvik P. Eksamenslære for dummies. Tidsskr Nor Legeforen 2020; 140. doi: 10.4045/tidsskr.20.0142.
4. Raaheim A. Eksamensrevolusjonen: Råd og tips om eksamen og alternative vurderingsformer. 2. utg. Oslo: Gyldendal, 2019.
5. Epstein RM. Assessment in medical education. N Engl J Med 2007; 356: 387–96. [PubMed][CrossRef]
6. Assessment methods in medical education. Int J Health Sci (Qassim) 2008; 2: 3–7. [PubMed]
7. Norcini J, Anderson B, Bollela V et al. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Med Teach 2011; 33: 206–14. [PubMed][CrossRef]
8. Schuwirth LWT, van der Vleuten CPM. A history of assessment in medical education. Adv Health Sci Educ Theory Pract 2020; 25: 1045–56. [PubMed][CrossRef]
9. Hodges B. Assessment in the post-psychometric era: Learning to love the subjective and collective. Med Teach 2013; 35: 564–8. [PubMed][CrossRef]
10. Biggs J, Tang C. Teaching for Quality Learning at University. 4. utg. New York, NY: Open University Press, 2011.

11. Boud D. Sustainable assessment: rethinking assessment for the learning society. *Stud Contin Educ* 2000; 22: 151–67. [CrossRef]
12. van der Vleuten CPM, Schuwirth LWT. Assessment of professional competence: From methods to programs. *Med Educ* 2005; 39: 309–17. [PubMed][CrossRef]
13. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011; 33: 478–85. [PubMed][CrossRef]
14. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. *Med Teach* 2020; 42: 846–54. [PubMed][CrossRef]
15. Clarke V, Braun V. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *Psychologist* 2013; 26: 120–3.
16. Vasan CAM, DeFouw DO, Holland BK et al. Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anat Sci Educ* 2018; 11: 254–61. [PubMed][CrossRef]
17. Gjersvik P. Å skrive må læres. *Tidsskr Nor Legeforen* 2017; 137. doi: 10.4045/tidsskr.17.0537. [PubMed][CrossRef]
18. Watling CJ, Ginsburg S. Assessment, feedback and the alchemy of learning. *Med Educ* 2019; 53: 76–85. [PubMed][CrossRef]
19. van der Vleuten C, Heeneman S, Schuwirth LWT. Programmatic assessment. I: Dent J, Harden R, Hunt D, red. *A Practical Guide for Medical Teachers*. 6. utg. Elsevier, 2021.
20. NOKUTs studiebarometer. Medisinstudiet. <https://studiebarometeret.no/no/student/finn?query=medisin&p=1&i=0&s=0&f=0&g=0> Lest 22.3.2022.
21. NOKUTs studiebarometer. Rapport 2021. https://www.nokut.no/globalassets/studiebarometeret/2021/hoyere-utdanning/studiebarometeret-2020_hovedtendenser_1-2021.pdf Lest 22.3.2022.
22. Gjersvik P. Evnen til å uttrykke seg. *Tidsskr Nor Legeforen* 2021; 141. doi: 10.4045/tidsskr.21.0747. [PubMed][CrossRef]
23. Valestrand EA, Slørdahl TS, Bærheim A et al. Mot en nasjonal delprøve i medisinstudiet. *Tidsskr Nor Legeforen* 2016; 136: 390–1. [PubMed][CrossRef]
24. Frich JC, Andersen S, Bærheim A et al. En felles prøve for alle legestudenter i Norge. *Tidsskr Nor Legeforen* 2017; 137. doi: 10.4045/tidsskr.17.0536. [PubMed][CrossRef]

Publisert: 12. august 2022. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.22.0337

Mottatt 1.5.2022, første revisjon innsendt 28.5.2022, godkjent 13.6.2022.

Publisert under åpen tilgang CC BY-ND. Lastet ned fra tidsskriftet.no 27. november 2022.