

Nettbasert flervalgseksamen i sykepleierutdanning

Høgskolelektor Lene Elisabeth Blekken, Prosjektleder/høgskolelærer Trond Indergaard, Høgskolelektor Marit Olsen. Høgskolen i Sør-Trøndelag, Avdeling for Sykepleie.

Rammeplan for sykepleierutdanningen [1] beskriver innholdet i sykepleierutdanningen inndelt i fire hovedemner, der hovedemne 3 er medisinske og naturvitenskapelige emner. Kunnskap om menneskets normale utvikling og sykdomsprosesser er nødvendig for at en sykepleier skal kunne observere og i samarbeid med andre iverksette forebyggende, behandlende, lindrende og rehabiliterende tiltak. Om vurderingsordninger skriver rammeplanen at de skal sikre at studentene i løpet av studiet tilegner seg de kunnskaper og kvalifikasjoner som er skissert i målsetningene for sykepleierutdanningen [1].

Ved Hist bygger fagplanen på problembasert læring (PBL). Til 2006 har utdanningen benyttet hjemmeeksamener hvor studentene skulle trekke inn kunnskap fra alle hovedemner i en og samme eksamen. Nokut stilte i 2005 krav om flere eksamener i sykepleierutdanningen. Dette, samt retningslinjene fra rammeplanen, ble de overordnede rammene for innføring av egen eksamen i medisinske og naturvitenskapelige emner.

I Norge har flervalgsoppgaver hatt varierende utbredelse, men de siste årene har denne oppgaveformen blitt vist større interesse [2]. Det viktigste for vellykket bruk, er at vi gjennom eksamensordninger faktisk måler og sikrer det som rammeplanen ber oss om, nemlig om studentene har tilegnet seg de kunnskaper og kvalifikasjoner som er skissert i målsetningen. Medisinske og naturfaglige emner er ofte konkrete og kan være et godt utgangspunkt i arbeidet med å konstruere flervalgsoppgaver med høy reliabilitet og validitet.

Målsettingen i forhold til vårt arbeid med elektroniske flervalgstester kan deles i to:

- Å identifisere og belyse faktorer som har betydning for en vellykket bruk av denne type tester i sykepleierutdanning
- Å klargjøre om elektroniske flervalgstester er en egnet eksamensform i sykepleierutdanning

Hva kjennetegner flervalgseksamen

En flervalgseksamen består av et sett med oppgaver der hver enkelt oppgave består av ett spørsmål med tilhørende svaralternativer. Flervalgsoppgaver er lukkede oppgaver der studentene med bakgrunn i kunnskap i faget skal resonnerer seg fram til riktig svar blant flere mulige fastsatte alternativer. Dette står i motsetning til åpne oppgaver som kjennetegnes med at studenten selv skal formulere svaret på et fritt grunnlag [2]. Ifølge Sirnes [2] består en flervalgsoppgave av to hovedelementer: stammen, som inneholder selve oppgaveteksten, og svaralternativer som består av en eller flere nøkler som er det/de riktige svaret/svarene og distraktorer som da er gale svar.

Reliabilitet

Reliabilitet sier noe om påliteligheten av eksamensresultatet. Eksamen skal teste kunnskapsnivået til studenter innen et gitt fagområde. Målet er å unngå at resultatet er påvirket av faktorer utenfor studenten. Reliabiliteten vil øke jo mer lukket et spørsmål er. Når eksamensoppgaver er utformet med ett riktig svar, vil sannsynligheten for god reliabilitet være høy [3]. Ut fra dette vil vi kunne oppnå høyere reliabilitet ved å bruke flervalgsoppgaver, enn ved å bruke åpne oppgaver. Vurderingen vil være objektivt i og med at svaret er entydig

definert på forhånd. Det er likevel sider ved flervalgsoppgaver som kan redusere reliabiliteten. R.F Burton [4] peker på noen faktorer: kvalitet på spørsmålene, kvalitet på svaralternativene, antall svaralternativer og antall oppgaver.

Validitet

Validitet handler om i hvilken grad eksamen prøver det den er tiltenkt å prøve, og om resultatet gir et sannferdig bilde av den kompetansen vi ønsker å måle [3]. Dette vil være særlig viktig i fag hvor studenten skal ta med seg den teoretiske kunnskapen over i en praktisk utøvelse av faget. Et viktig utgangspunkt for utformingen av eksamen blir da å være bevisst på hva vi ønsker å måle. Videre blir det viktig å se på ulike eksamensformer og hvilke muligheter og begrensinger som ligger i disse. I dette arbeidet kan Blooms kognitive taksonomi være en god veiviser. Målet er å teste studentens kognitive ferdigheter gjennom hele taksonomien: kunnskap, forståelse, anvendelse, analyse, syntese og vurdering. Det blir ofte hevdet av flervalgsoppgaver bare kan måle enkeltstående fakta og således kommer nederst på Bloom's taksonomi, men det er fullt mulig å lage flervalgsoppgaver som tester høyere kognitive ferdigheter [2]. Buckles og Siegfried [5] mener det mulig å måle kunnskap opp til nivå 4 i Blooms taksonomi. I arbeidet med å konstruere oppgaver til flervalgseksamen er det enklest å teste faktakunnskapene. Jo høyere du beveger deg opp i Blooms taksonomi, jo mer krevende blir det å lage oppgavene. I tillegg til dette blir det viktig å ha bevissthet om at man lager oppgaver knyttet til alle områdene som regnes som vesentlig kunnskap, og ikke bare til de områdene som lettest lar seg omforme til korte spørsmål og lett tilgjengelige svaralternativer.

Erfaringer fra gjennomføring av flervalgseksamen

Vi gjennomførte for første gang flervalgseksamen i naturvitenskapelige emner i første studieår i februar 2006 (Kull 2005). I november 2006 gjennomførte vi tilsvarende eksamen for neste kull (Kull 2006). Et av argumentene for å ta i bruk denne type tester som eksamensform er muligheten for ressursbesparelser på lang sikt. Vår erfaring så langt er at det har vært svært ressurskrevende å komme i gang med denne type eksamenstesting. Ressursbruken har vært fordelt på ulike arbeidsoppgaver i forberedende fase:

- Konstruksjon av oppgaver har først og fremst vært tillagt ulike faglærere /forelesere i naturvitenskapelige emner. Både eksterne forelesere med medisinsk bakgrunn og pedagogisk personale ved avdelingen har bidratt i dette arbeidet.
- Kvalitetssikring av oppgaver har foregått i to faser. En oppnevnt ekstern sensor har gjennomgått alle nye oppgaver og har kommentert og foreslått endringer til disse. Ekstern sensor har både sykepleiefaglig og medisinsk bakgrunn.
- Tilpassing av verktøy med tanke på rammer rundt praktisk/teknisk gjennomføring av testene. Dette er ivaretatt av personer med sykepleiefaglig og IT-faglig kompetanse (faglig IT) ved avdelingen.
- Informasjon til studentene. Dette er ivaretatt av fagansvarlig pedagogisk personell samt faglig IT.

Fra første til andre gjennomføring ble det gjort endringer i rammefaktorene for avviklingen av eksamen. Studentundersøkelser etter gjennomførte eksamener, samt våre erfaringer basert på blant annet loggføring av teknisk support under gjennomføringen, danner grunnlag for vårt videre arbeid. Spørreundersøkelsene var anonyme. De ble gjennomført i uken etter eksamen. Begge undersøkelsene består halvt om halvt av ja / nei spørsmål og med spørsmål hvor studentene skulle angi sitt syn på en 6-delt skala. Svarprosenten på spørreundersøkelsene var 74 for kull 2005 og 76 for kull 2006.

Tabell 1 Rammefaktorer for eksamen Kull2005 (feb. 2006) og Kull 2006 (nov. 2006) ¹= riktig svar. ²=feil svar

Område	Kull 2005	Kull 2006
Type spørsmål i eksamen	1 nøkkel ¹ + 2 distraktører ²	1 nøkkel + 2 distraktører
Antall spørsmål i base	196	329
Antall kategorier	6	6
Mulighet til forberedelse, student	10 spørsmål på 10 minutter	10 spørsmål på 10 minutter
Uttrekk til eksamen	Tilfeldig uttrekk av 60 spm	Tilfeldig uttrekk av 60 spm
Tid på eksamen	1 time	1 time
Sted for eksamen	Hvor som helst med internett	Datarom på skolen med vakt
Antall smtidig påloggede studenter	203	20
Karakter	Bestått / ikke bestått	Bestått / ikke bestått
Grense for bestått i %	66	60
Informasjon om resultat	Umiddelbart etter eksamen	Umiddelbart etter eksamen
Tid til å gå gjennom besvarelse	Ubegrenset	I ca. 10 minutter etter eksamen
Mulighet til å skrive ut resultat	Ubegrenset	Ingen utskriftsmulighet
Opplysning om rett svar	Om spørsmål er rett / feil	Om spørsmål er rett / feil
Tilgang Faglærer / IT støtte	På telefon	I eksamenslokalet
Hjelpemidler under eksamen	Tillatt med alle	Tillatt med bøker og web

Kommentarer til endringer av rammefaktorer: Oppgavene ble ytterligere kvalitetssikret med tanke på at alle svar på spørsmål skulle være mulig å finne i anbefalt litteratur. Eksamen skulle avvikles under eksamensvilkår for skoleeksamen på et datarom på skolen. Studentene fikk ikke kommunisere med andre. Studentene ble i god tid før eksamen tildelt tidspunkt for eksamen. Generelt må det sies at informasjonen til studentene før eksamen var mer presis. Faglærer var tilstede i eksamenslokalet under hele eksamen. En kommentar til forberedelse til eksamen: 89 % av studentene gjennomførte testen før første eksamen. 95 % gjennomførte testen før andre eksamen. Når det gjelder stabilitet av teknologien hadde vi betydelig høyere frekvens av tekniske problemer da hele kullet (203), var pålogget samtidig enn ved eksamen i grupper på 20.

Tabell 2 Fordeling av resultater eksamen Kull 2005 og Kull 2006: Gjennomsnitt Kull 2005=79 %, Kull 2006=80%. Ikke bestått: Kull 2005: 10%, Kull 2006: 4 %

	Kull 2005		Kull 2006		Endring
	Antall	%	Antall	%	
% rette svar					
90 til 100 %	31	15	49	23	+8
80 til 89 %	70	34	71	33	-1
70 til 79 %	67	33	56	26	-7
66 til 69 %	14	7	11	5	-2
60 til 65 %	13	6	17	8	+4
Under 60%	8	4	8	4	0
Sum	203	99	212	99	

Vi antok at endringer i rammefaktorer ville føre til at eksamen for kull 2006 ble vanskeligere enn for kull 2005. På bakgrunn av dette ble grensen for bestått flyttet ned fra 66 % til 60 %. Resultatene fra eksamen tyder på at vi tok feil.

Resultater fra studentundersøkelsene

Under presenterer vi resultater fra spørreundersøkelsene som ble gjennomført i etterkant av eksamen. Materialet fra analyse av undersøkelsene er omfattende. Innen rammen av dette paperet forholder vi oss kun til noen utvalgte variabler.

Tabell 3 Hva synes du om tiden du fikk til å gjennomføre eksamen

	Kull 2005		Kull 2006	
	Antall	Prosent	Antall	Prosent
<i>God tid</i>	14	9,3	33	20,5
2	21	14,0	26	16,1
3	18	12,0	39	24,2
4	23	15,3	28	17,4
5	33	22,0	16	9,9
<i>Dårlig tid</i>	41	27,3	19	11,8
Total	150	100,0	161	100,0

Studentenes syn på tid til gjennomføring av eksamen. Studentene ble bedt om å angi hva de syntes om tiden de fikk til gjennomføring av eksamen på en skala fra 1 (God tid) til 6 (Dårlig tid). Ved første eksamen var gjennomsnittet litt over 4 og modus var 6. Ved andre eksamen var gjennomsnittet litt over 3 og modus var 3. Studentene som avla eksamen på skolen anga med andre ord at de var mer fornøyd med tiden som var avsatt til eksamen.

Tabell 4 Synes du ditt resultat på eksamen gir et godt bilde av ditt kunnskapsnivå? Prosent (antall)

	Kull 2005 (feb. 06)	Kull 2006 (nov. 06)
<i>Ja</i>	41 (61)	60 (97)
<i>Nei</i>	59 (88)	40 (64)
Total	100 (n=149)	100 (N=161)

Når det gjelder studentenes syn på om eksamen gir et godt bilde av deres kunnskapsnivå i faget mente 41 % av studentene i kull 2005 at eksamen ga et godt bilde av deres kunnskapsnivå. I kull 2006 var det 60 % som syntes eksamen ga et godt bilde.

Analysen viser relativt liten sammenheng mellom hverken alder eller kjønn og studentenes syn på hvor godt bilde eksamen ga av deres kunnskapsnivå.

Et klart flertall av studentene anga at de ønsket denne formen for eksamen senere i studiet. Henholdsvis 77% (Kull 2005) og 83% (Kull 2006). Analysen viser liten sammenheng mellom dette spørsmålet og alder. Menn er generelt mer positiv til flervalgseksamen enn kvinner.

Analysen viser liten sammenheng mellom studentenes ønske om denne formen for eksamen i andre emner senere i ditt studie og deres oppfatning om resultatet på eksamen ga et godt bilde av deres kunnskapsnivå.

Faktorer som har betydning for en vellykket bruk av denne type tester i sykepleierutdanning

Vellykket bruk av elektroniske flervalgstester forutsetter at vi ivaretar interessene til både pedagogisk personale, studentene og de teknisk ansvarlige. Vi har identifisert mange faktorer som kan ha betydning for dette arbeidet, men vil her nøye oss med å diskutere noen av disse som vi anser som spesielt viktige i denne sammenhengen.

I forhold til reliabilitet vil oppgaver som består av lukkede spørsmål i seg selv bidra til reliabilitet. Dette krever imidlertid at spørsmålet er tydelig. I tillegg vil svaralternativene og løsningsrommet vi gir studentene bli svært viktig. Vi må sørge for at vi hverken snevrer inn for mye eller trekker inn alternativer som ligger for langt til siden for temaet i spørsmålet. Svaralternativene må også være mange nok til at vi reduserer sannsynligheten for gjetting.

Når det gjelder validitet er en av fordelene at vi får mulighet til å teste store deler av pensum, men dette krever igjen at de som har ansvar for å lage oppgaver greier å konstruere gode spørsmål innen alle temaer anser som viktig. Oppgavene må også utfordre studentene på ulike nivå i Blooms taksonomi. Dette betyr at vellykket bruk krever en del ressurser og da

spesielt når det gjelder utforming av oppgaver. I tillegg må alle oppgaver gjennom en kvalitetssikring fra ekstern sensor. Faglærere må også ha et nært samarbeid med de ansvarlige for teknisk gjennomføring. To av forfatterne har både IT-faglig og sykepleiefaglig kompetanse, disse var teknisk ansvarlige ved gjennomføringen av eksamen. Vi anbefaler at tilsvarende kombinert kompetanse er representert i arbeidet rundt en tilsvarende eksamen.

Faktorer som er nevnt i forhold til lærere vil også ha betydning for studenter. For våre studenter er dette en ny eksamensform og for mange er det første gang de skal benytte elektronisk verktøy i en eksamenssituasjon. God og entydig informasjon på forhånd blir dermed ekstra viktig. I tillegg vil mulighet til å prøve ut eksamensformen gjennom realistiske tester bidra til trygghet både i forhold til verktøy og eksamensform. Det er nødvendig at den tekniske rammen rundt eksamen fungerer problemfritt og at testverktøyet ikke oppleves forstyrrende i eksamenssituasjonen. Det er viktig at studentene opplever at resultatet gir et riktig bilde av opplevd kunnskapsnivå. Sett i forhold til de svært gode resultatene er det betenkelig at bare 40 % og 60 % av kullene angir at eksamen gir et godt bilde av deres kunnskapsnivå. En del studenter kan oppfatte det positivt at vi med denne testformen får testet studentens kunnskap innen et emne og ikke deres evne til å uttrykke seg skriftlig. Sensuren er objektiv og innebygget i systemet. Det gjøres ingen skjønsmessige vurderinger, noe som eliminerer muligheten for forskjellsbehandling ved sensurering. Når det gjelder tid til gjennomføring må denne tilpasses i forhold til antall spørsmål, vanskelighetsgrad på spørsmål og i forhold til i hvor stor grad vi lærere ønsker at de skal kunne slå opp i litteratur. Det er viktig at studentene er "passe" presset i forhold til tid. Sett opp mot resultatene som ble oppnådd på eksamen er det forfatterens oppfatning at tiden studentene hadde til disposisjon ved eksamen var for god i forhold til antall og type oppgaver de skulle besvare.

For teknisk/praktisk administrasjon og gjennomføring er det viktig å ha en databaseløsning som gjør det enkelt å vedlikeholde oppgavebasen. En database for alle oppgavene og en standardisert mal for oppretting av nye oppgaver er nødvendig. Databaseløsningen må være uavhengig av system slik at oppgavene kan brukes i ulike typer testverktøy. Metadata knyttet til oppgavene er nødvendig for organisering og gjenfinning av oppgaver samt deling av oppgaver med andre avdelinger/institusjoner.

Flexibilitet i testverktøyet er en annen viktig faktor. Det må være enkelt å sette sammen et oppgavesett der en for eksempel kan kombinere randomisering med faste oppgaver. Det må være enkelt å håndtere brukertilganger knyttet til testverktøy. Også for teknisk ansvarlige er det viktig å utelukke muligheten for at studenter blir kastet ut av systemet. I tilfeller der dette allikevel skjer, er det viktig at teknisk ansvarlige har mulighet til å styre brukertilganger også under gjennomføring av tester. Vi har erfart at studentene kan bli kastet ut fra systemet under gjennomføring. I slike tilfeller er det vesentlig at allerede besvarte oppgaver er lagret i systemet. Mulighet for enkel navigering mellom oppgaver i eksamenssituasjonen er etterspurt blant studentene og må tas hensyn til ved valg av testverktøy.

Er elektroniske flervalgstester en egnet eksamensform i sykepleierutdanning?

Et klart svar på spørsmålet er det ikke mulig å gi. I den grad faktorene vi har påpekt over blir ivarettatt, mener forfatterne at det er aktuelt å gå videre med utvikling og utprøving av flervalgseksamen som en av flere eksamensformer ved vår avdeling. I tillegg til faktorer knyttet til testverktøy, reliabilitet, validitet og ressursbruk er forholdet til pedagogisk konsept av stor betydning. Som nevnt innledningsvis bygger fagplanen ved Hist på PBL (problembasert læring). Vi må vi ha en grunnleggende drøfting i forhold til om, og eventuelt hvordan, innføring av denne type eksamen står i forhold til PBL som grunnleggende konsept.

Planer og mål for videre arbeid med nettbaserte flervalgseksamener i sykepleierutdanningen

Formatert: Tal
11,75 cm, Left

For at vi skal kunne høste gevinst i form av ressursbesparelser over tid er gjenbruk et nøkkelord. Ved første gjennomføring hadde vi 196 antall oppgaver i vår database. Ved andre gjennomføring var antallet oppgaver økt til 329. Vi planlegger å fortsette med å utvide databasen. Jo flere oppgaver i databasen jo bedre muligheter er det for å variere oppgavesettene. Ønsket om en stor oppgavebase må ikke gå på bekostning av kvaliteten på den enkelte oppgave. Det må sikres at oppgavene i databasen er representative for kunnskapen det forventes studentene skal inneha. Kontinuerlig evaluering av oppgavene i databasen må ivaretas. Vi vil jobbe systematisk med å konstruere oppgaver som tester ulike nivå i Blooms taksonomi. Vi planlegger også å konstruere oppgaver med flere enn tre svaralternativer og andre typer oppgaver som for eksempel paringsoppgaver og dra og slipp oppgaver.

Et mål er å se på mulighetene for et nasjonalt samarbeid om en felles oppgavebase for flervalgsoppgaver. Dersom en felles oppgavebase skal realiseres vil det være nødvendig å utarbeide et minste felles metadatasett for oppgavene. Metadata (her: beskrivende data om den enkelte oppgave) er nødvendig for å muliggjøre deling av oppgaver mellom avdelinger og institusjoner. Eksempel på aktuelle metadata i denne sammenheng kan være: Forfatter, institusjon, kategori, vanskelighetsgrad, hvor ofte brukt.

Til nå har vi benyttet testverktøyet i it's learning ved gjennomføring av eksamen. Våre erfaringer så langt tilsier at dette verktøyet ikke dekker våre behov godt nok. Vi ønsker et system som er mer fleksibelt på flere områder. Her nevnes kort:

- Muligheten til å kombinere faste oppgaver med tilfeldig utvalgte oppgaver.
- Mulighet for å navigere mellom oppgaver i tillegg til å kunne bla fram og tilbake.
- Vising av alle oppgaver på en side samtidig som det settes tidsbegrensning

Det er et mål for oss å finne og å ta i bruk et testverktøy som dekker våre behov. Et nytt testverktøy er under utvikling ved Avdeling for Informatikk og e-læring ved HiST og vår avdeling vil være med på å teste ut dette verktøyet. Vi har også fått muligheten til å stille krav til verktøyet som gir oss fleksibilitet i forhold til våre behov. Samarbeidet så langt har gitt positive inntrykk med tanke på både fleksibilitet og stabilitet og vi planlegger å fortsette dette samarbeidet.

Kilder:

1. Utdannings- og forskningsdepartementet, *Rammeplan for sykepleierutdanning*. 2005, Oslo. 14.
2. Sirnes, S.M., *Flervalgsoppgaver - konstruksjon og analyse*. 2005, Bergen: Fagbokforl. 137 s.
3. Lauvås, P. and A. Jakobsen, *Exit eksamen - eller? former for summativ evaluering i høgre utdanning*. 2002, Oslo: Cappelen akademisk forl. 243 s.
4. Burton, R.F., *Multiple choice and true/false tests: reliability measures and some implications of neaktiv marking*. *Assessment & Evaluation in Higher Education*, 2004. **29**(5): p. 585 - 595.
5. Buckles, S.J.J.S., *Using Multiple-Choice Questions to Evaluate In-depth Learning of Economics*. *Journal of Economic Education*, 2006. **37**(1): p. 48-57.

Kontaktadresser: lene.blekken@hist.no , trond.indergaard@hist.no , marit.olsen@hist.no