

# Rapport MED 2521 vår 2017

Eksamenskommisjonen for MED 2521 har våren 2017 bestått av:

Katrine Wennevold, allmennmedisin, leder

Øystein Grimstad, hud

Jørgen Isaksen, nevrofag

Audny Anke, nevrofag

Tore Sørli frem til 1 mai/Terje Øiesvold etter 1 mai, psykiatri

Lars Marius Ytrebø, akuttmedisin

Tommy Thorsen, ekstern

Eksamenskommisjonen har hatt jevnlige møter våren 2017.

Eksamenskommisjonen leverte to eksamenssett til PEU 1 april. Ordinær eksamen bestod av 65 spørsmål, hvorav 59 MCQ av typen «single best answer» og 6 kortsvarsoppgaver, kontinuasjonseksamen bestod av 67 spørsmål, hvorav 60 MCQ av typen «single best answer» og 7 kortsvarsoppgaver. Denne rapporten omhandler ordinær eksamen. PEU bedømte ordinær eksamen til å være en middels vanskelig eksamen.

Det var på forhånd vedtatt at eksamen MED 2521 fra i år skulle være digital eksamen gitt i programmet Wiseflow.

Eksamenskommisjonen vurderte at årets eksamensoppgaver innenfor MCQ i hovedsak var bedre oppgaver enn kortsvarsoppgavene. Av den grunn ble også noen kortsvarsoppgaver skrevet om til MCQ av eksamenskommisjonen, og kvalitetssikret av faglærere etterpå. Eksamenskommisjonen ønsket en stor grad av MCQ for å unngå stor grad av skjønnsmessig sensur, noe som ofte blir tilfelle på kortsvarsoppgaver. I tillegg ønsket eksamenskommisjonen å øke antallet oppgaver fra 45 i 2016 til 2/3 av det antallet som gis på 6 times eksamen for å øke reliabilitet. I tråd med føringer fra PEU ble fordelingen av oppgaver gjort i henhold til hvor mange uker hvert fag undervises i MED 2521.

Studentene ble informert om økt andel MCQ og økt antall spørsmål 2 mai, og hadde mange spørsmål og bekymringer rundt dette. De fikk av den grunn kursprøve i delemne 4.10 allmennmedisin i Wiseflow, og med samme format på MCQ som på eksamen. I tillegg ble det på fronter lagt ut noen eksempler på gode MCQ fra nasjonal eksamen til orientering.

87 studenter møtte og fullførte årets eksamen. Den første studenten forlot eksamen etter 90 minutter, den siste brukte alle 4 timene. På selve eksamen viste det seg at det ikke var plass nok til å skrive svarene på kortsvarsoppgavene rett inn i Wiseflow. Studentene fikk da utdelt ark som ble levert inn. Disse arkene ble scannet og sendt på mail til medlemmene i eksamenskommisjonen.

Før selve eksamen fikk eksamenskommisjonen utsendt et skjema for å standardsette eksamen etter Ebels metode. Med ett unntak fylte alle i kommisjonen ut et slikt skjema (se vedlegg 1). Det viste seg å være svært spredning innad i kommisjonen på hvordan man vurderte oppgavenes viktighetsgrad og vanskelighetsgrad. På møtet hvor endelig strykgrense skulle bestemmes møtte bare ett medlem av kommisjonen opp i tillegg til leder, og det var da ikke mulig å slutføre Ebel standardsetting. Ut fra standardsettingen var spriket mellom kommisjonsmedlemmene på foreslått strykgrense mellom 31% og 62%. En sannsynlig årsak til spriket var manglende opplæring i metoden før man skulle fylle ut skjema for standardsetting, og dette bør prioriteres om metoden skal brukes senere år. Et stort sprik som vi opplevde gjør Ebel standardsetting umulig med mindre man har et møte for å enes om strykgrensen.

Det var planlagt møte for å enes om strykgrense dagen før eksamen, men møtet måtte avlyses da kun ett medlem foruten leder møtte. Resultatene fra Wiseflow for studentenes score på MCQ forelå straks etter eksamen. Studentene hadde scoret langt bedre enn forutsett av kommisjonen gjennom sin estimering av strykgrense på den uferdige Ebel standard settingen, med laveste score før tillegg av kortsvarspoeng på 37 av 59 riktige på MCQ, altså 63% rett. Det ble da åpenbart at kommisjonen ikke hadde bedømt vanskelighetsgrad av oppgavene korrekt i det spede forsøket på å bruke Ebel standardsetting, og det ble besluttet å forkaste Ebel som metode for denne kommisjonen.

Kortsvarsoppgavene ble rettet av 6 av kommisjonens medlemmer. I par på to rettet de oppgavene til 1/3 av kandidatene. Leder av kommisjonen rettet noen av oppgavene hvor det var stor uenighet mellom sensorpar. Det viste seg å være stor uenighet om score på kortsvarsoppgavene, med lav interrater reliabilitet. På selve sensurmøtet ble det derfor satt av god tid til å oppnå enighet mellom sensorene på score før endelig sensur ble foretatt.

Ettersom Ebel standardsetting ikke ble gjennomført foreslo leder av eksamenskommisjonen å standardsette MCQ spørsmålene på eksamen etter Cohen's metode (se vedlegg 2), og å legge til 4 poeng for kortsvarsoppgavene (6 oppgaver, 60% er 3,6 rette, avrundet til 4). Det var ikke lagt inn krav om å faktisk klare 4 av 6 av kortsvarsoppgavene, men at det totale antall riktige svar på MCQ bedømt via Cohens metode pluss 4 ble satt som strykgrense. Ut fra Cohens metode ble det da krav om 40 riktige, og med tillegg av 4 poeng ble strykgrensen satt til totalt 44 riktige svar, eller 67,6%. Det ble vedtatt med 6 mot 1 stemme å bruke denne metoden for standardsetting. De 6 som stemte for å anvende Cohen's metode syntes det var et vesentlig poeng å øke strykgrensen når det var en hovedvekt av spørsmål hvor det var 25% sjanse til rett svar på eksamen, og dermed legge til denne sjansen i beregningene.

Formel for beregning av strykgrense var:

*For Cohen's method, the following formula was applied: Standard  $cN + 0,60(N^* - cN)$  with  $c$  being the proportion of chance performance in the test,  $N$  the total number of items in the test and  $N^*$  the score of the best performing student(s).*

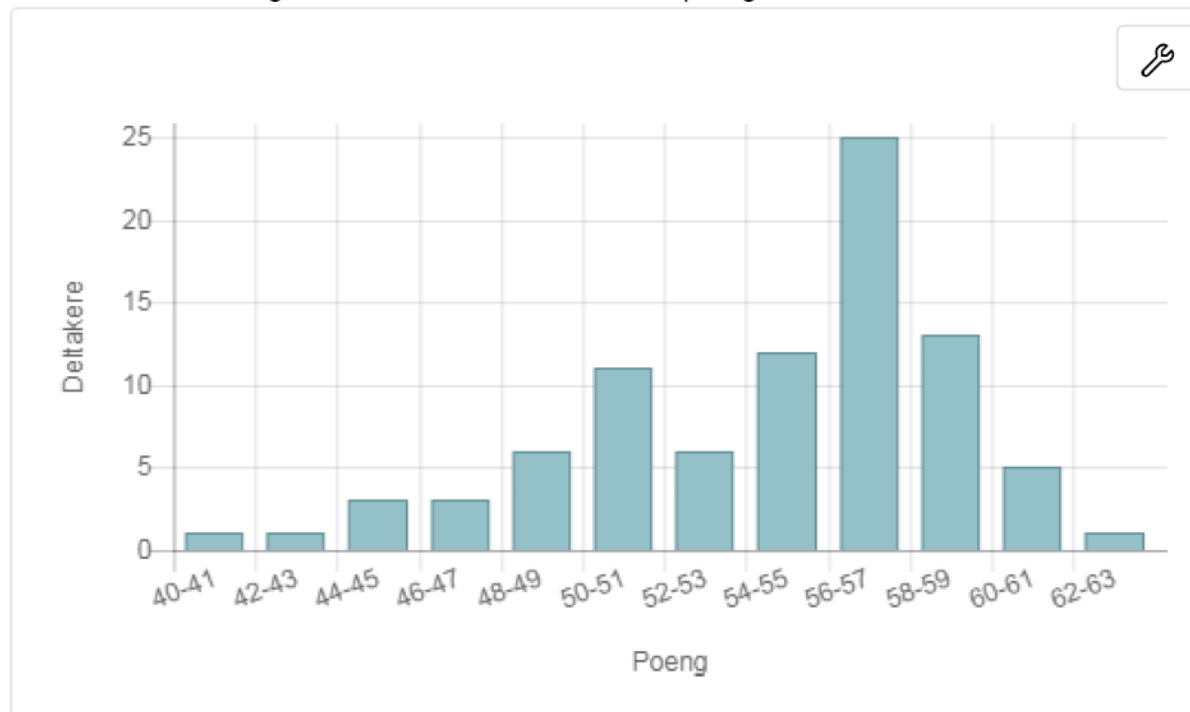
Ved denne formen for standardsetting og bestemmelse av strykgrense på 67,6 % havnet 2/87 kandidater under strykgrensen. Dette ble vedtatt som et akseptabelt resultat med 6 mot 1 stemme. Medlemmet som stemte mot ønsket at vi ikke skulle gjennomføre denne formen for standardsetting men heller skulle beholde 60% som tidligere år. Dette vil ha ført til at alle kandidatene bestod eksamen.

Studentenes totalscore på eksamen er gitt ved følgende figur:

## Oversikt og statistikk

### Poengddistribusjon

Her kan du se fordelingen av deltakere i forhold til samlet poengsum



Deltakere

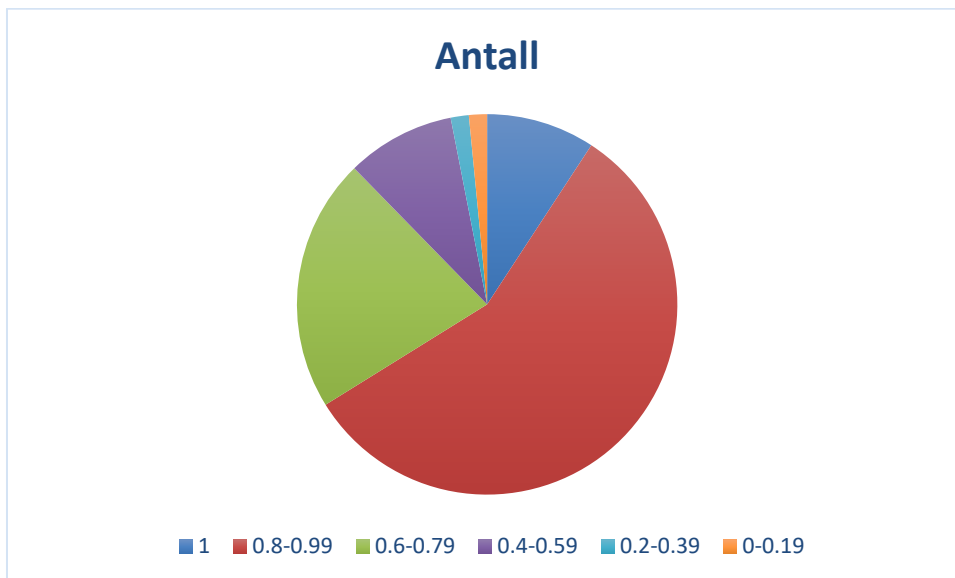
Resultat

Oppgavesett

Valgte svar

Psykometrien for årets eksamen viser at MCQ oppgavene har en reliabilitet gitt ved Cronbach's alfa på 0,65, noe som er lavt. Dette skyldes at det var flere oppgaver hvor alle hadde rett, slik at det ble få oppgaver igjen til å kalkulere reliabilitet. I tillegg er 59 oppgaver totalt så lavt at det vanskelig å oppnå høy reliabilitet.

Item difficulty (hvor mange som har rett svar per oppgave) viser at det er mange spørsmål hvor majoriteten av studentene har rett svar. Anbefalt verdi for item difficulty er 0,6-0,8 for MCQ. For vårens eksamen er fordelingen slik for alle oppgavene, inkl kortsvarsoppgavene:



Dette tyder på at oppgavene i hovedsak var for enkle for studentene. Oppgavene skulle vært plassert i grønn sone, mens majoriteten er i rød sone, og en nokså stor andel i blå (alle har rett).

Eksamenskommisjonen konkluderte med en lærerik periode. Med 6 mot 1 stemme ønsker vi 100% MCQ for senere år, da kortsvarsoppgavene er vanskelige å rette på grunn av upresise spørsmål og upresis fasit. I tillegg er det atskillig enklere å ha kun MCQ på en digital eksamen. Kommisjonen vil satse på høyere vanskelighetsgrad på MCQ, og ønsker skolering i hvordan skrive gode oppgaver. For å øke reliabiliteten ønsker vi å øke antall oppgaver til 100. Det vil mest sannsynlig være tid nok til å løse 100 oppgaver på 4 timer, da studentene beskriver at de på utveksling måtte løse 100 oppgaver på 90 minutter. Et av medlemmene i kommisjonen påpeker at det likevel er noen som brukte tida helt ut, og at man bør øke eksamen til 6 timer om man øker til 100 spørsmål. For å sikre reliabiliteten er det viktigste antallet spørsmål, og så kan man gjerne bruke 6 timer på å løse dem om det er nødvendig.

*26.06.17 Katrine Wennevold, leder av eksamenskommisjonen*

En av medlemmene har ønsket tillagt sin dissens:

Merknad for sensur eksamen Med 2521, vår 4. året: Metode for utregning av strykgrense ble bestemt og gjennomført først etter vurdering av resultatene. Undertegnede mener en i framtiden bør ha en standardsetting av sensur. Siden en i år ikke fikk gjort dette før sensurmøtet mener jeg strykgrensen burde vært satt til 60%, som har vært standard de siste årene, og ikke høyere slik det ble gjort.

Audny Anke

# Vedlegg 1 Ebel standardsetting

Spørsmål I	viktighet					vanskelighetsgrad					Item difficulty
	ØG	AA	KW	JI	TØ	ØG	AA	KW	JI	TØ	
1	e	v	v	e	e	l	v	l	l	l	1
2	g	g	v	g	v	m	v	m	m	m	0,99
3	e	v	v	e	v	m	m	m	l	m	
4	e	g	e	e	e	m	v	m	l	v	0,77
5	v	g	v	e	v	m	v	v	m	v	0,75
6	e	v	v	e	v	l	v	l	l	l	0,7
7	e	v	v	v	v	l	v	l	l	l	0,98
8	v	v	e	e	e	l	m	l	l	l	1
9	v	g	e	e	e	v	v	m	m	m	0,4
10	v	g	v	v	e	m	v	m	v	m	0,91
11	v	v	g	v	g	m	v	v	l	v	0,92
12	g	g	g	g	g	m	v	m	v	v	0,94
13	v	g	v	v	g	m	v	m	l	v	0,64
14	v	v	v	v	v	v	v	v	m	m	0,98
15	v	v	e	e	e	m	m	m	l	l	0,98
16	v	g	v	v	v	v	v	m	m	m	0,98
17	v	v	v	v	v	m	v	m	m	m	0,97
18	v	v	e	v	v	m	v	l	m	v	0,92
19	v	g	v	e	v	m	v	l	m	v	0,95
20	v	g	v	v	e	m	m	l	l	l	0,98
21	e	v	v	e	e	m	v	m	m	m	0,82
22	e	e	v	e	e	l	v	m	m	m	0,84
23	e	v	v	v	v	m	v	m	v	m	0,95

24	e	v	v	e	v	l	v	l	l	l	0,67
25	e	v	v	e	e	m	v	m	m	l	0,94
26	g	g	v	v	g	m	v	l	m	m	0,98
27	v	v	v	v	e	l	v	l	l	l	0,97
28	v	v	v	v	v	l	m	l	l	l	0,64
29	v	v	g	v	e	m	m	m	l	m	1
30	v	g	v	v	e	l	v	m	m	m	0,98
31	g	g	g	g	g	v	v	v	m	v	0,57
32	g	g	v	v	v	v	v	v	m	m	0,74
33	v	v	v	v	e	v	v	m	v	v	0,78
34	e	v	v	e	e	l	m	l	l	l	0,89
35	g	g	v	g	v	v	v	m	m	m	0,98
36	v	v	v	v	v	m	v	m	l	l	0,99
37	v	v	v	v	g	m	v	m	l	m	
38	v	v	v	e	v	m	m	m	l	m	
39	v	g	v	e	e	m	v	l	l	m	
40	e	v	e	e	e	m	v	l	l	m	
41	v	v	v	v	v	m	v	l	v	m	0,16
42	e	v	v	e	v	m	v	m	m	m	0,85
43	e	v	v	e	v	l	m	l	l	l	1
44	e	v	e	e	v	m	m	l	l	l	0,89
45	v	v	v	e	v	m	v	m	m	m	0,36
46	e	v	v	e	v	m	v	m	m	m	0,91
47	v	g	v	e	e	v	v	m	v	v	0,76
48	g	g	v	v	e	v	v	m	v	v	0,85
49	g	g	v	v	e	v	v	v	v	m	0,92
50	g	g	v	g	g	v	v	l	v	v	0,48

51	v	v	v	v	v	m	v	l	m	m	0,52
52	v	v	v	e	e	m	m	l	l	m	0,94
53	e	v	v	e	v	l	v	l	l	m	0,92
54	v	v	v	v	v	m	v	l	l	l	0,95
55	g	g	v	v	v	v	m	m	m	m	0,98
56	e	v	v	e	v	m	v	m	m	m	
57	v	v	e	e	v	l	v	l	l	v	1
58	e	g	v	e	e	m	v	m	l	l	0,93
59	e	v	e	e	v	m	v	l	l	l	1
60	v	g	v	v	v	m	v	m	m	m	0,89
61	g	v	v	v	g	v	v	v	m	v	0,78
62	g	g	v	v	g	m	v	m	m	m	0,97
63	g	v	v	v	v	v	v	m	l	l	0,85
64	v	v	v	v	v	v	v	l	m	l	0,68
65	v	v	v	e	e	v	v	v	v	m	0,84

Tentativ strykgrense gitt HS sitt oppsett

Øg	AA	KW	JI	TØ
53	31	55	64	57
%	%	%	%	%

## Vedlegg 2

Link til artikkel om Cohen standardsetting

<http://www.tandfonline.com/doi/pdf/10.3109/01421590903196979?needAccess=true>