

EKSAMENSRAPPORT OSKE IIAB vår 2016

Innhold

1. Bakgrunn
2. Eksamensinnhold
3. Standardsetting og logging av resultater
 - 3.1 Metode for standardsetting
 - 3.2 Bruk av nettbasert programvare
4. Eksamensresultat OSKE IIAB vår 2016
 - 4.1 Resultater OSKE IIAB vår 2016
 - 4.2 Resultater stasjonsvis OSKE IIAB vår 2016

1. Bakgrunn

Det medisinske fakultet (DMF), NTNU, ved fakultetsstyret, vedtok i 2014 å innføre OSKE (objektiv strukturert klinisk eksaminasjon) som erstatning for den tidligere brukte muntlig-praktiske trestasjonseksamen på medisinstudiet. OSKE er en stasjonseksamensform for testing av klinisk kompetanse. Kommunikasjonsferdigheter, undersøkelsesteknikk og kliniske ferdigheter testes godt gjennom en OSKE. OSKE avholdes som en summativ eksamen hvor de samme stasjonsoppgavene gis til alle studentene på kullet på samme dag.

Prosjektgruppen for OSKE startet arbeidet høst/vinter 2014/15 og planla oppstart av OSKE for tredje studieår (stadium IIAB) i løpet av studieåret 2015/16.

En fullskala testeksamen (pilotOSKE) ble avholdt for 106 tredjeårsstudenter den 10.12.16. Mer informasjon om planleggingen, resultatene og anbefalingene fra prosjektgruppa finnes i dokumentet "Evaluering av pilotOSKE", som er åpent tilgjengelig på nettsiden:

<https://www.ntnu.no/dmf/eksamen-medisinstudiet>

Prosjektet ble videreført etter pilotOSKE og etter vedtak om innføring av OSKE som vurderingsmetode på førsteåret, tredjeåret og fjerdeåret.

Første summative OSKE ble avholdt fredag 3.6.16 for tredjeårsstudentene (stadium IIAB).

Et sensorteam bestående av 5 personer, med leder professor Ragnar Joakimsen, UiT; ble utnevnt av DMF i forkant av OSKE. Disse personene deltok aktivt på eksamensdagen, var med på sensurmøtet til eksamenskommisjonen etter OSKE og har levert rapport i etterkant.

2. Eksamensinnhold

For detaljer rundt lagning av eksamensoppgaver mv. vises til rapporten fra pilotOSKE, undervisningsveileder for OSKE, eksamensprosedyre og eksamensretningslinjer for OSKE.

Fagområdene i tabellen under er involvert i undervisning i stadium IIAB.

Fagoversikt:

Semester IIA (oversikt)	Semester IIB (oversikt)	Felles (oversikt)
Nevrologi	Hjerte- og karsykdommer	Patologi
Nevrokirurgi	Lungemedisin	Mikrobiologi
ØNH, sykdommer i munn, kjeve, ansikt	Sykdommer i beinmarg, blod og lymfatisk vev	Klinisk beslutningslære
Øyesykdommer	Fordøyelsesykdommer	Klinisk kjemi /venflonkurs
Hjerneslag (geriatri)	Gastrokirurgi	Farmakologi
Fysikalsk medisin/rehabilitering	Thoraxkirurgi	Radiologi
Ortopedi - rygg	Karkirurgi	A/DHLR-kurs
Svulstsykdommer (palliativ medisin)	Immunologi og transfusjonsmedisin	Suturkurs
Klinisk nevrofysiologi	Arbeids- og miljømedisin	Kommunikasjon
Rettsmedisin	Svulstsykdommer	

Kilder: Semesterhåndbøker IIA/IIB, studiehåndbok (medisinske studier) 2015/2016

(Uthevet: fagområder som hadde stasjon på pilotOSKE, i rødt fagområder som hadde stasjon på ordinær OSKE vår 2016, uthevet rødt fagområder som hadde stasjon på begge OSKE)

Stasjonskartet, eller stasjonsrammeverket, for stadium IIAB ble levert til studentene en måned før OSKE. Alle mulige OSKE-stasjoner ut i fra undervisning og læringsmål sto her. Følgelig også alle stasjonene som kom på OSKE. Utlevering av et OSKE-stasjonskart er et viktig tiltak som ikke bare veileder studentene før OSKE, men også i deres kliniske læringsarbeid videre.

Studiekonsulent for IIAB, Turid Beitland, og prosjektleder for OSKE har ut i fra undervisningsandeler lagd 10 forskjellige OSKE-eksamener for IIAB med 8 stasjonsoppgaver i hver. Eksamen som ble avholdt den 3.6.16 er en av disse OSKE-settene.

Oversikt/rammeverk for eksamen OSKE IIAB vår 2016:

Fagområde(r)	Stasjonstype Tema	Semester	Læringsmål
1. Hematologi Klinisk kjemi	Ferdighet Rekvirering og analyse av laboratorieprøver	IIB	2.1.7, 7.1.2
2. Radiologi	Ferdighet Radiologiske modaliteter i utredning av magesmerter	IIB	2.1.12, 6.1.7, 11.1.7
3. Kommunikasjon Almennmedisin	Anamnese/kommunikasjon Pasientsentrert metode	Felles	ID: 13.1.1, 13.1.2 IIAB: 2.1.6
4. B/DHLR	Ferdighet (Hjerte-/lungeredning)	Felles	A/DHLR-kurs
6. Hjerneslag	Klinisk undersøkelse Symptomer på hjerneslag	IIB	10.1.1, 10.2.1, 10.2.3, 10.2.3.1

7. Øye	Klinisk undersøkelse, ferdighet Skygge i synsfeltet	IIA	3.1.2
8. Kardiologi	Ferdighet Tolkning av EKG, STEMI	IIB	IB: 4.1.2 IIB: 3.1.3, 3.1.4
9. Øre-nese-hals	Klinisk undersøkelse Halssmerte/-betennelse	IIA	4.1.1, 4.1.2, 4.1.3

3. Standardsetting og logging av resultater

3.1 Metode for standardsetting

Standardsettingsmetoden "borderline regression" ble benyttet på eksamensresultatene for å definere beståttgrense på hver stasjon.

I denne metoden gjennomfører eksaminatoren skåring av studentens prestasjon i et forhåndsbestemt skåringsskjema. Hver skåring gir et visst antall poeng. Eksaminator angir også global skår på studentprestasjonen etter sitt eget skjønn. Alle sjekklisteskjårene fra alle eksaminatorene som skårer på samme stasjon blir så regressert via regresjonsanalyse mot global skår som ble gitt på den samme stasjonen. Dette gir en lineær modell hvor grensetilfellegruppen på global skår blir brukt til å predikere beståttgrensen for sjekklisteskår.

(Kilde Hejri, Sara Mortaz et al. "Assessing the Reliability of the Borderline Regression Method as a Standard Setting Procedure for Objective Structured Clinical Examination." *Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences* 18.10 (2013): 887–891.)

Prosjektgruppen satte som hovedregel for OSKE at studentene måtte bestå minst 5 av 8 stasjoner for å få beståttkarakter totalt.

3.2 Bruk av nettbasert programvare

Prosjektgruppen for OSKE utlyste en internasjonal anbudkonkurranse våren 2016 for innkjøp av nettbasert programvare for håndtering av OSKE-oppgaver, administrative funksjoner og OSKE-eksamensavholdelse.

Programvaren OMIS, fra QPERCOM, vant anbudskonkurransen og har nå inngått kontrakt med NTNU som er gyldig i tre år (til og med studieåret 2018/19).

OMIS ble brukt på eksamen med tilfredsstillende resultat. Bruk av nettbasert programvare for en klinisk eksamensform som OSKE forenkler arbeidet med resultatlogging og -analyse i stor grad. Bruken bidrar også til eksamensformens objektivitet, gjør logistikken på eksamensdagen enklere og muliggjør direkte tilbakemelding til enkeltstudenter om deres egen prestasjon og resultat.

4. Resultater

Psykometriske data slik som reliabilitet og validitet er viktige parametere for å evaluere eksamenskvaliteten. Cronbachs alfa er et mål på reliabilitetstypen intern konsistens og er ekvivalent til intraklasse korrelasjonskoeffisient. Det vil si samsvar mellom ett mål på et subjekt og ett annet mål på det samme subjektet. For eksempel: om alle skåringselementene i hver stasjonsoppgave eller om alle stasjonene i en eksamen tester det samme.

R (lineær korrelasjonskoeffisient) er et mål på samvariasjonen mellom to variabler, her poengskår og global skår.

R² ("coefficient of determination") er denne verdien opphøyd i annen og er et mål på hvor godt den statistiske modellen kan konstruere en prediktor (her beståttgrense) ut i fra variablene.

(Kilder OSCE kurs Tromsø, "Medical statistics" Marit Veierød et al)

4.1 Resultater OSKE IIAB vår 2016

Totalt deltok 111 studenter på OSKE og 888 stasjonsresultater ble logget i OMIS.

Total Cronbachs alpha (CA) for hele OSKE på 0,54. Dette er en middels god reliabilitet for hele eksamenen. Dette korresponderer godt med lengden på eksamenen: 1 time og 20 minutter aktiv eksamenstid. Dette sammenlignet med tidligere publiserte resultater. (Cees van der Vleuten, International Advanced Assessment Course, London, 2014).

Analyse viser at alle stasjoner gir lavere total CA dersom de blir tatt ut av eksamenssettet etter tur ("Cronbach's alpha if item deleted"). Det betyr at alle stasjonene på OSKE IIAB vår 2016 bidrar positivt til totalreliabiliteten til eksamen (fig under) noe som er tilfredsstillende.

Item-Total Statistics					
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Hemtaologi	55,3241	16,414	,420	,239	,407
Radiologi	55,3331	18,630	,390	,213	,445
Kommunikasjon	56,5899	17,958	,170	,106	,511
DHLR-2	53,9201	20,520	,178	,126	,500
Hjerneslag	55,2581	19,247	,165	,052	,504
Øye	55,5223	19,534	,192	,063	,494
EKG	56,3100	16,897	,209	,102	,499
ØNH	56,2725	16,618	,272	,103	,466

Fig. Hvordan CA vil endre seg hvis hver enkelt stasjon blir tatt ut av eksamenssettet.

Det var ingen signifikante forskjeller i resultat mellom de ulike studentgruppene som hadde eksamen gjennom dagen. Det vil si, ingen fordel eller ulempe å komme først eller sist på dagen.

Noen stasjoner hadde forskjeller i gjennomsnittsskår gitt av hver enkelt eksaminator. Og på én stasjon var denne forskjellen signifikant (se resultater på stasjon 3). Men eksaminatorvariasjonen for OSKE IIAB vår 2016 var mindre enn for pilotOSKE noe som er gledelig.

Den statistiske modellen for predikering av beståttgrense per stasjon, borderline regression method, fungerte godt. R²-verdiene er gode eller tilfredsstillende for alle stasjonene, unntatt for B/DHLR-stasjonen (se resultater stasjon 4).

Elleve studenter strøk på 2 av 8 stasjoner. To studenter strøk på 3 av 8 stasjoner (disse to fikk ca 64% av poengene på hele OSKE). Disse to studentene skiller seg ut som en del svakere enn resten av kullet, men har også stasjoner i løpet av OSKE hvor de gjør det

bra. Ingen studenter strøk på 4 eller flere av de 8 stasjonene, det vil si: ingen studenter i IIAB strøk på OSKE vår 2016.

Ut i fra data fra eksamen er dette trolig ikke fordi eksamen var for lett. Dette studentkullet har vært «primet» på OSKE fra oppstart av studieåret og har fått rikelig anledning til å øve på OSKE-formatet. Prosjektgruppen for OSKE vil revidere standardsettingsnivået til neste OSKE og vurdere om 3 av 8 stasjoner som beståttgrense er for lavt.

Eksamen OSKE IIAB vår 16 fremstår som med god kvalitet totalt sett og resultatene er godkjent av sensorteam for OSKE og eksamenkommisjonen for IIAB.

4.2 Resultater stasjonsvis OSKE IIAB vår 2016

Stasjon 1: Hematologi/klinisk kjemi

Totalt deltok 5 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 77-86 %. Gjennomsnittsskår for alle eksaminatorene var 81,8%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **57,3%**.

Etter denne analysen var det 4 studenter som strøk (3,6%). Disse studentene hadde fått global skår: «klar stryk» 1 pers, «grensetilfelle» 2 pers, «bestått» 1 pers.

Hematologi/klinisk kjemi B12 Grense 5,73 (57,3%)

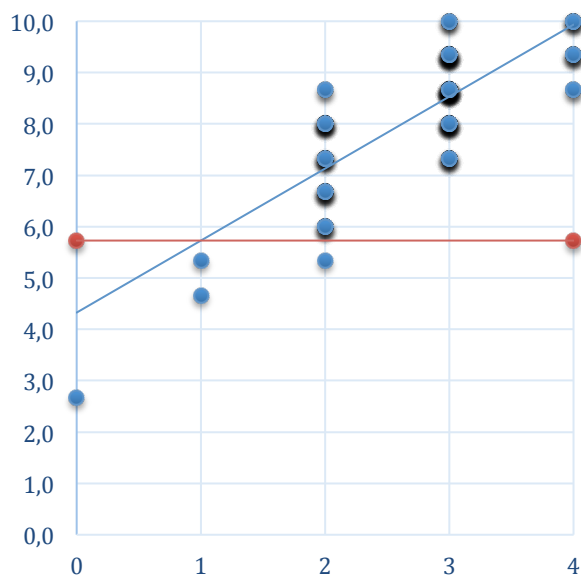


Fig: Regresjonsanalyse stasjon 1

Reliabilitet/kvalitet:

CA 0,467 (lav). Skåringselement (linje i skårings skjemaet) 2, 3, 4 og 5 bidrar negativt til den interne konsistensen i oppgaven. Få skåringselementer i oppgaven gjør CA lavere enn reelt og studentene presterer for variert innenfor samme oppgave.

R2 0,624 (god). God korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Den statistiske modellen for predikering av beståttgrense fungerer godt. Eksaminatorene enige om nivå som definerer grensetilfellestudenten, noe som gjør at modellen fungerer.

Konklusjon testkvalitet:

Lav CA trolig grunnet for stor varians i studentenes skår innenfor samme oppgave (noen skåringselementer med svært varierende svarandel). Tester disse avvikende elementene samme kunnskap som de andre elementene? Er undervisningen uklar på dette området? Modellen som predikerer beståttgrensen er god. Liten eksaminatorvariasjon i oppgaven.

Stasjonen og resultatene godkjent av eksamenskommisjonen uten endringer.

Stasjon 2 radiologi:

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 78,7-83,9 %. Gjennomsnittsskår for alle eksaminatorene var 81,7%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **64,5%**.

Etter denne analysen var det 3 studenter som strøk (2,7%). Disse studentene hadde fått global skår: «klar stryk» 1 pers, «grensetilfelle» 2 pers.

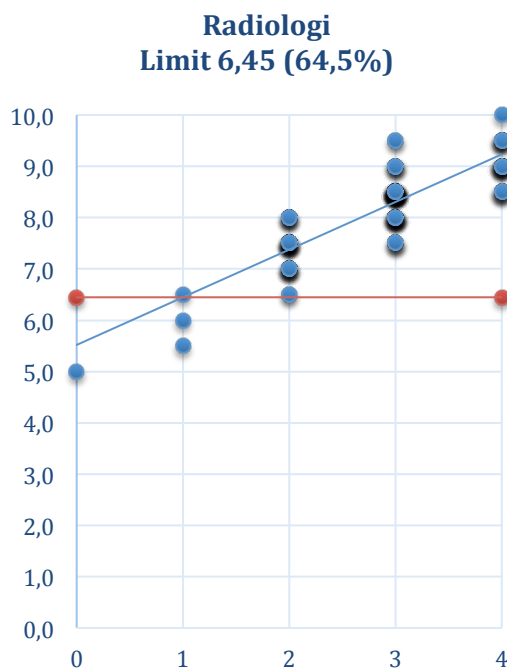


Fig. Regresjonsanalyse stasjon 2

Reliabilitet/kvalitet:

CA 0,350 (lav). For stor intern variasjon innad i oppgaven. Få skåringselementer i oppgaven gjør CA lavere enn reellt og studentene presterer for variert innenfor samme oppgave.

R2 0,740 (god). God korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Eksaminatorene er samstemte om hvilke prestasjoner som passer inn i de

forskjellige global skår-kategoriene. Den statistiske modellen for predikering av beståttgrense fungerer godt. Eksaminatorene enige om nivå som definerer grensetilfellestudenten, noe som gjør at modellen fungerer.

Konklusjon testkvalitet:

Lav CA trolig grunnet for stor varians i studentenes skår innenfor samme oppgave (noen skåringselementer med svært varierende svarandel). Tester disse avvikende elementene samme kunnskap som de andre elementene? Er undervisningen uklar på dette området?

Hvis det er elementer i oppgaven som er designet for å skille de gode studentene fra de svake studentene, vil det bidra negativt til CA – men er ikke nødvendigvis feil.

Modellen som predikerer beståttgrensen er god, beståttgrensen blir 64,5%. Liten eksaminatorvariasjon i oppgaven.

Stasjonen og resultatene godkjent av eksamenskommisjonen uten endringer.

Stasjon 3 kommunikasjon/almennmedisin

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 60,6-75,4 %. Gjennomsnittsskår for alle eksaminatorene var 69,1%. Det var én eksaminator som skåret signifikant lavere enn sine kolleger. Skåringen var imidlertid konsistent og i samsvar med global skår-kategorier.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **52,4%**.

Etter denne analysen var det 12 studenter som strøk (10,8%). Disse studentene hadde fått global skår: «klar stryk» 3 pers, «grensetilfelle» 7 pers, «bestått» 2 pers.

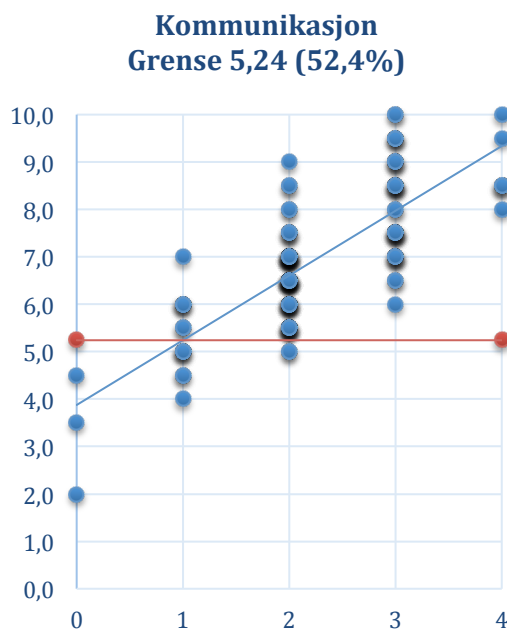


Fig. Regresjonsanalyse stasjon 3

Reliabilitet/kvalitet:

CA 0,712 (god). Skåringselement (linje i skåringsskjemaet) 4 og 5 bidrar negativt til den interne konsistensen i oppgaven. De andre elementene i skjemaet tester samme emne omtrent like godt.

R2 0,601 (ok/god). Ok korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Det er relativt stor uoverensstemmelse mellom eksaminatorene i gruppene "grensetilfelle" på 40-70% og "bestått" på 50-90%. Den statistiske modellen for predikering av beståttgrense fungerer tilfredsstillende.

Konklusjon testkvalitet:

God oppgavekvalitet med god reliabilitet og skåringen passer godt inn i statistisk modell. Signifikant eksaminatorvariasjon i oppgaven.

Stasjonen og resultatene godkjent av eksamenskommissjonen uten endringer.

Stasjon 4 B/DHLR

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 94,2-97,2%. Gjennomsnittsskår for alle eksaminatorene var 95,8%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense 85,1%.

Denne beståttgrensen ble justert ned til **84,6%** av eksamenskommissjonen etter anbefaling fra prosjektleder for OSKE.

Etter denne analysen, med justering, var det 2 studenter som strøk (1,8%). Disse studentene hadde fått global skår: «klar stryk» 1 pers (skår 46%), "bestått" 1 pers (skår 62%).

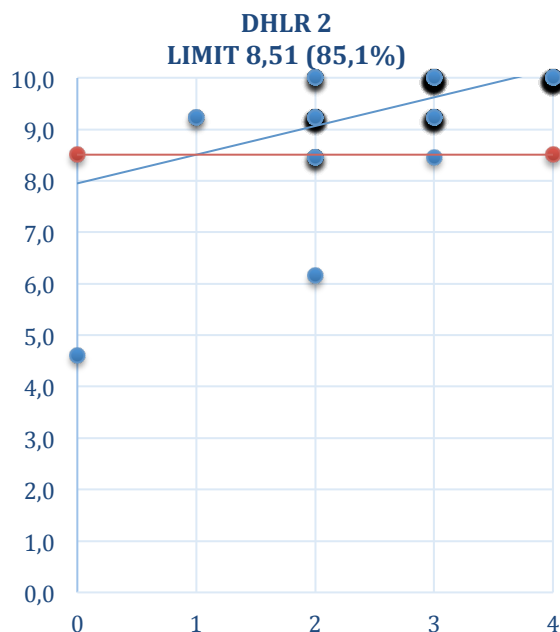


Fig. Regresjonsanalyse stasjon 4

Reliabilitet/kvalitet:

CA 0,499 (ok). Skåringselement (linje i skåringsskjemaet) 3 og 10 bidrar negativt til den interne konsistensen i oppgaven – for stor variasjon i om studentene klarer dette.

R2 0,390 (dårlig). Den statistiske modellen fungerer dårlig fordi studentene vet at denne oppgaven kommer på eksamen og det har vært vanskeligere for eksaminatorene å bruke global skår-skalaen. En student ble plassert i "grensetilfelle"-gruppen, men denne

studenten hadde hele 92% i poengskår. Regresjon med predikering via grensetilfellegruppen fungerer derfor ikke godt.

Konklusjon testkvalitet:

Akseptabel reliabilitet i denne oppgaven, og ingen eksaminatorvariasjon.

Beståttgrensen justert til 84,6% slik at 6 studenter ikke fikk stryk hvis grensen var 85,1% (urimelig strengt når resultatet så nært til beståttgrense). Justeringen godkjent av eksamenskommisjonen og resultatendringen hadde ingenting å si for totalresultatet til kullet.

HLR på fremtidige eksamener

Prosjektgruppa vil utrede HLR sin posisjon i fremtidens OSKE. Det er foreløpig ikke planlagt som obligatorisk stasjon ved OSKE. Kanskje det var mulig å gjøre en AHLR-utsjekk som obligatorisk arbeidskrav *før* man fikk ta OSKE? Vi kan snakke mer om dette til høsten!

Stasjon 6 hjerneslag

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 79,2-84,9%. Gjennomsnittsskår for alle eksaminatorene var 82,5%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **59,7%**.

Etter denne analysen var det 5 studenter som strøk (4,5%). Disse studentene hadde fått global skår: «grensetilfelle» 1 pers, "bestått" 4 pers.

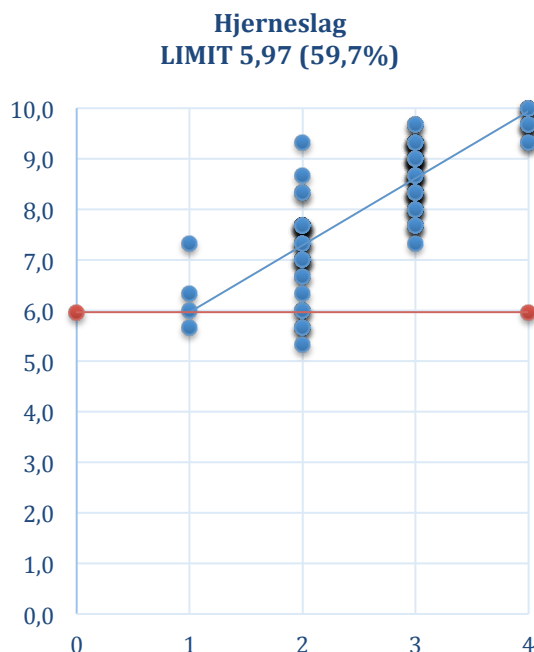


Fig. Regresjonsanalyse stasjon 6

Reliabilitet/kvalitet:

CA 0.637 (ok/god). Tilfredsstillende til god reliabilitet og god intern konsistens.

R2 0,648 (ok/god). God korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Gruppen "bestått" viser litt for stor variasjon i poengskår, 53,3-93,3%. Den statistiske modellen for predikering av beståttgrense fungerer godt.

Konklusjon testkvalitet:

God oppgavekvalitet med god reliabilitet og skåringen passer godt inn i statistisk modell. Liten eksaminatorvariasjon i oppgaven.

Stasjonen og resultatene godkjent av eksamenskommissjonen uten endringer.

Stasjon 7 øye

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 74-84.8 %. Gjennomsnittsskår for alle eksaminatorene var 79,8%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **64,7%**.

Etter denne analysen var det 5 studenter som strøk (4,5%). Disse studentene hadde fått global skår: «klar stryk» 2 pers, «grensetilfelle» 3 pers.

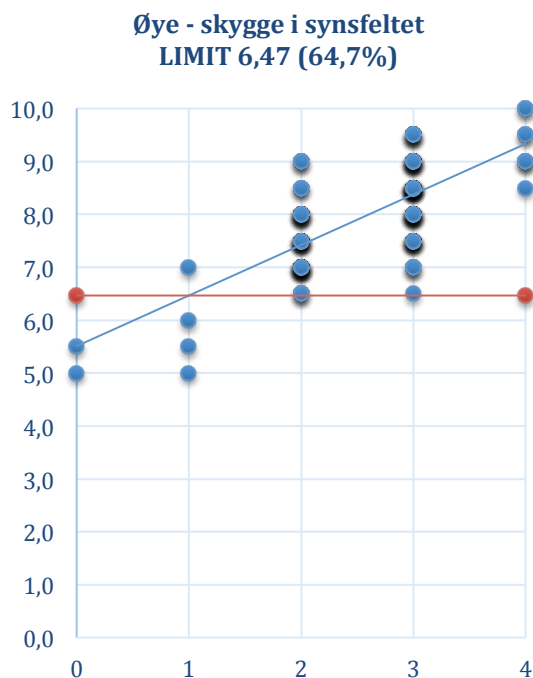


Fig. Regresjonsanalyse stasjon 7

Reliabilitet/kvalitet:

CA 0, 235 (lav). Det er for stor variasjon i skåring mellom de ulike skåringselementene/linjene i skåringsskjemaet. Linje 6 (RAPD/swinging flash light) bidrar negativt til reliabiliteten – svært varierende om studenten gjorde dette. De flinkeste studentene gjorde ikke alltid denne undersøkelsen. Få skåringselementer i oppgaven gjør CA lavere enn reelt og studentene presterer for variert innenfor samme oppgave.

R2 0,516 (ok). Tilfredsstillende korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Gruppene "bestått" og "god bestått" er skåret veldig likt,

henholdsvis 65-90% og 65-95%. Dette betyr at R2 ikke blir høy. Den statistiske modellen for predikering av beståttgrense fungerer tilfredsstillende/godt.

Konklusjon testkvalitet:

Noe dårlig reliabilitet/intern konsistens, men eksaminatorene er godt enige og det er sammenheng mellom poengskår og global skår.

Stasjonen og resultatene godkjent av eksamenskommissjonen uten endringer.

Stasjon 8 kardiologi

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 64,8-76,2 %. Gjennomsnittsskår for alle eksaminatorene var 71,9%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **50%**.

Etter denne analysen var det 6 studenter som strøk (5,4%). Disse studentene hadde fått global skår: «klar stryk» 2 pers, «grensetilfelle» 2 pers, "bestått" 2 pers.

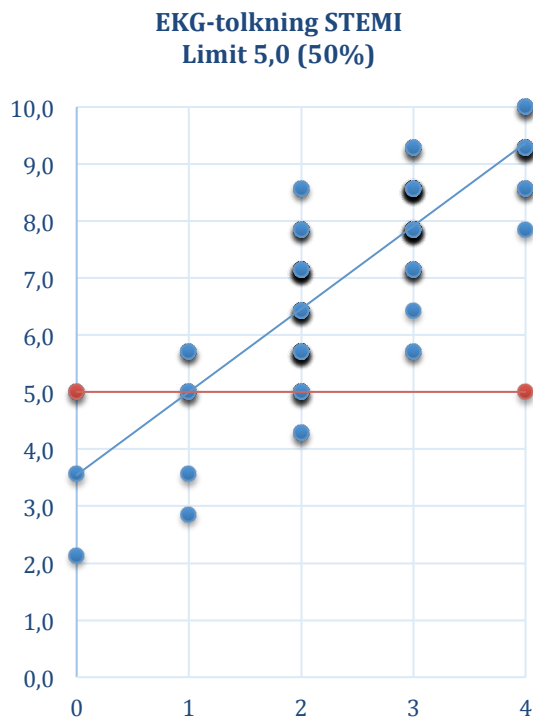


Fig. Regresjonsanalyse stasjon 8

Reliabilitet/kvalitet:

CA 0,610 (ok/god). Skåringselement (linje i skåringsskjemaet) 3, 6, 9 og 10 bidrar negativt til den interne konsistensen i oppgaven. Studentene presterer for variert innenfor samme oppgave slik at reliabiliteten går noe ned.

R2 0,667 (god). God korrelasjon mellom poengskåring og global skår (skjønnsmessig vurdering). Den statistiske modellen for predikering av beståttgrense fungerer godt.

Konklusjon testkvalitet:

Akseptabel oppgavekvalitet med ok/god reliabilitet og skåringen passer godt inn i

statistisk modell. Liten eksaminatorvariasjon i oppgaven.
Stasjonen og resultatene godkjent av eksamenskommissjonen uten endringer.

Stasjon 9 ØNH

Totalt deltok 4 eksaminatorer på denne stasjonen. Gjennomsnittsskår fra eksaminatorene var 65,6-78,4 %. Gjennomsnittsskår for alle eksaminatorene var 72,3%.

Etter borderline/grensetilfelle-regresjonsanalyse ble beståttgrense **51,6%**.

Etter denne analysen var det 14 studenter som strøk (12,6%). Disse studentene hadde fått global skår: «klar stryk» 1 pers, «grensetilfelle» 4 pers, "bestått" 9 pers.

Variasjonsbredden i poengsum i "bestått"-gruppen er hele 36-86%, som kan bety at eksaminatorene har hatt varierende forståelse av hva som går inn i den skjønsmessige skårkategorien "bestått".

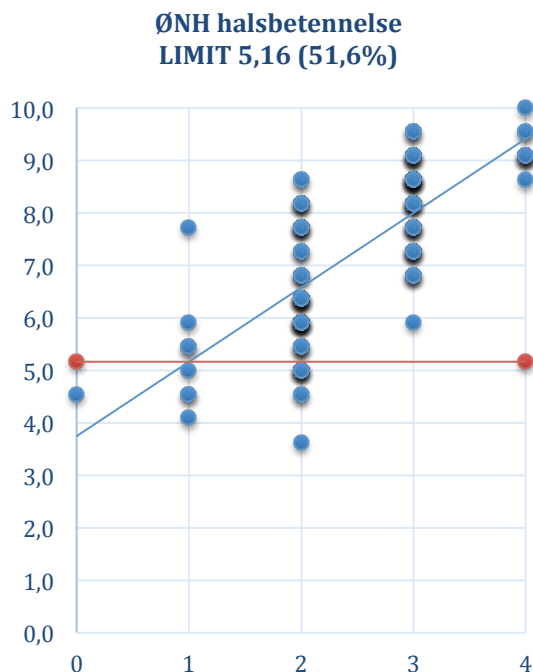


Fig. Regresjonsanalyse stasjon 9

Reliabilitet/kvalitet:

CA 0,772 (god). Skåringselement (linje i skåringsskjemaet) 5 og 8 bidrar negativt til den interne konsistensen i oppgaven. God reliabilitet i oppgaven, alle elementene i oppgaven tester samme kunnskap.

R2 0,539 (ok). Ok korrelasjon mellom poengskåring og global skår (skjønsmessig vurdering). Den statistiske modellen for predikering av beståttgrense fungerer ok, men det er for stor uoverensstemmelse mellom eksaminatorene for hva slags prestasjon som skal gi "grensetilfelle" eller "bestått" i global skår.

Konklusjon testkvalitet:

Akseptabel oppgavekvalitet med god reliabilitet og skåringen passer greit inn i statistisk

modell. Noe eksaminatorvariasjon i oppgaven, men ikke signifikant.
Stasjonen og resultatene godkjent av eksamenskommissjonen uten endringer.

Marie Thoresen
Prosjektleder OSKE

DMF, NTNU
Juni 2016