



UiO : **Det medisinske fakultet**

Digital eksamen

Organisering for kvalitet

Per Grøttum

Seksjon for medisinsk informatikk



Disposisjon

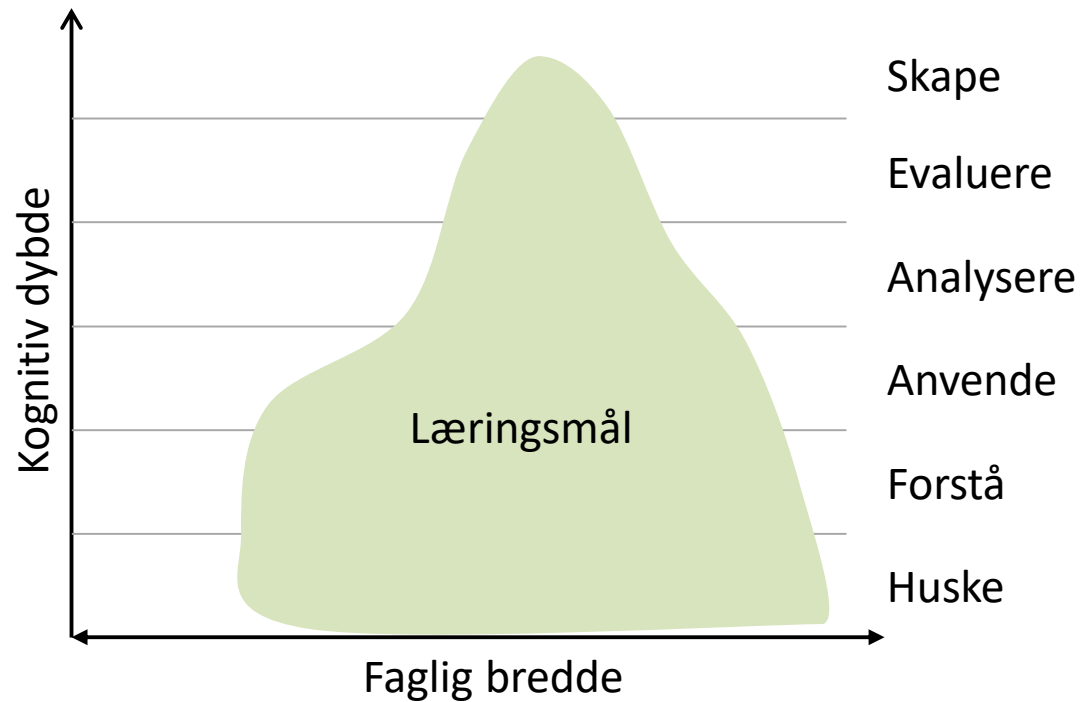
- Oppgaveproduksjon
- Oppgavetyper
- Oppgaveforfatting
- Sensur

Oppgaveproduksjon

Validitet

Måler testen det den er ment å måle?

Validitet



Eksempel: Emnet MED3300

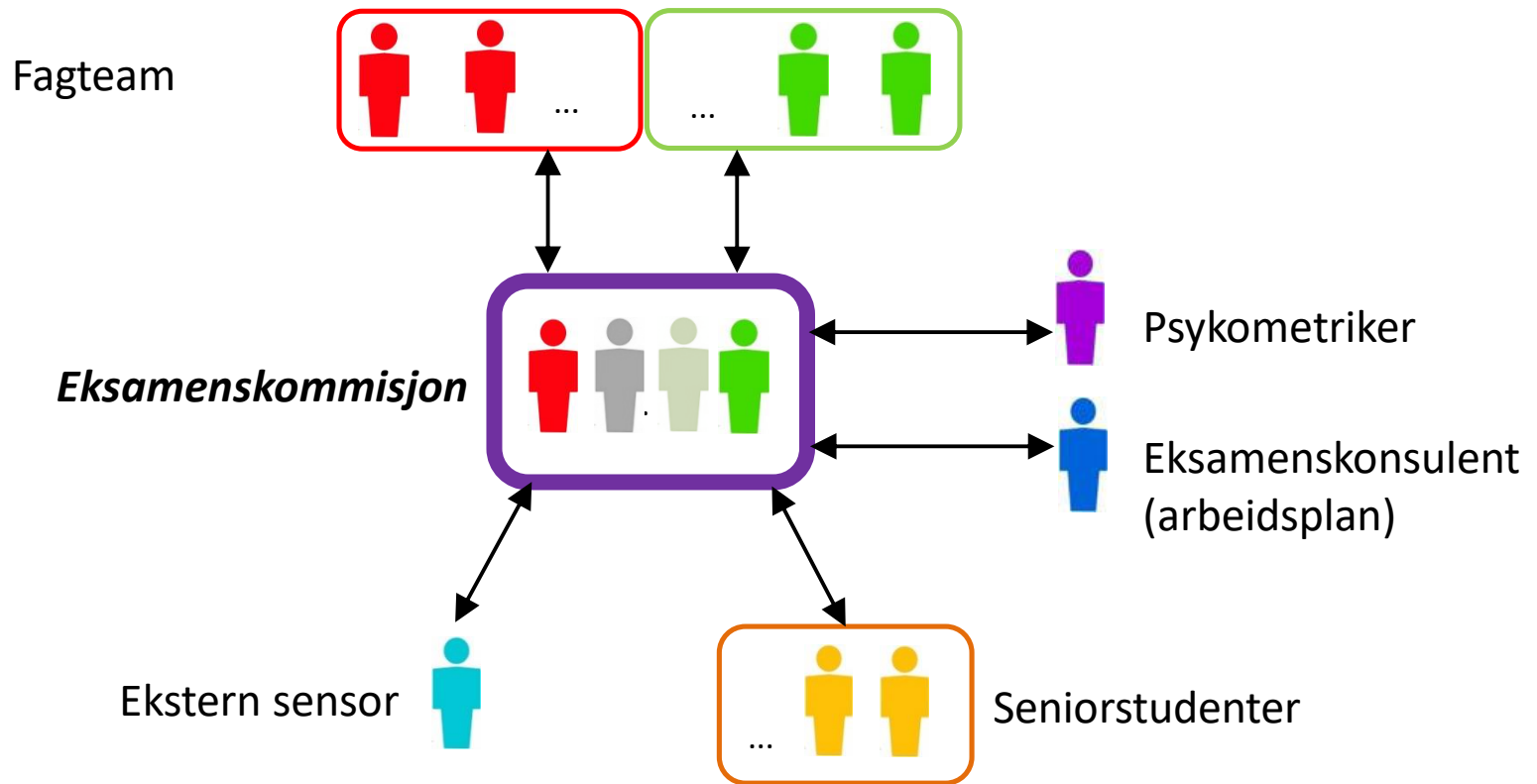
Ramme:

- 204 lærere
- 32 fag - én eksamen dekker 28 fag
 - Allmennedisin
 - Anestesiologi
 - Blodsykdommer
 - Endokrinologi
 - Ernæring
 - Farmakologi
 - Fordøyelsessykdommer
 - Fysikalsk medisin
 - Gastroenterologisk kirurgi
 - Geriatrici
 - Hjertesykdommer
 -

Læringsutbyttebeskrivelse (LUB):

.....gjennomføre diagnostiske vurderinger av de *mest vanlige sykdommene* i fordøyelsessystemet.

Teamarbeid



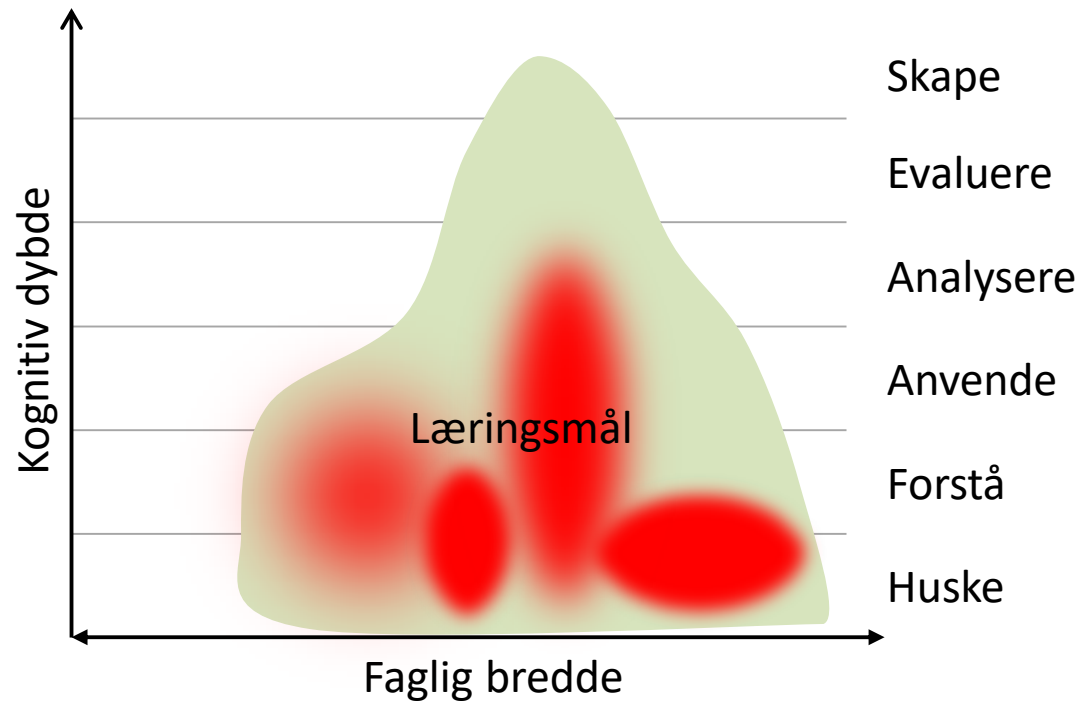
Oppgavetyper

Reliabilitet

Er resultatet stabilt og konsistent?

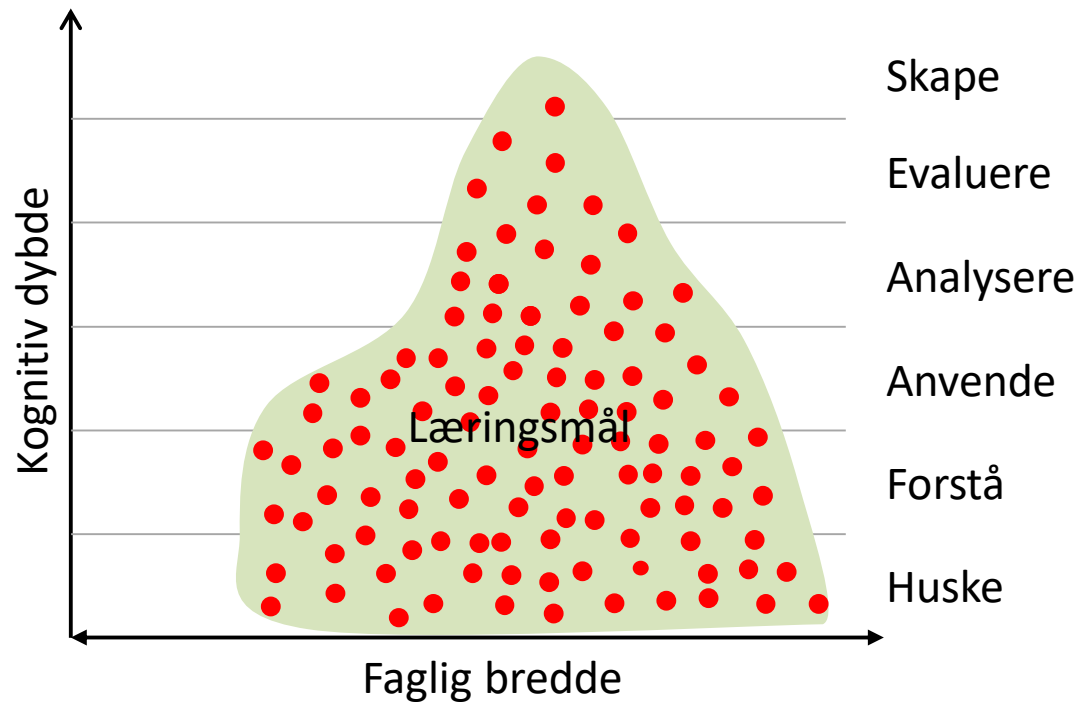
Reliabilitet

Dårlig løsning: tradisjonell essay-eksamen



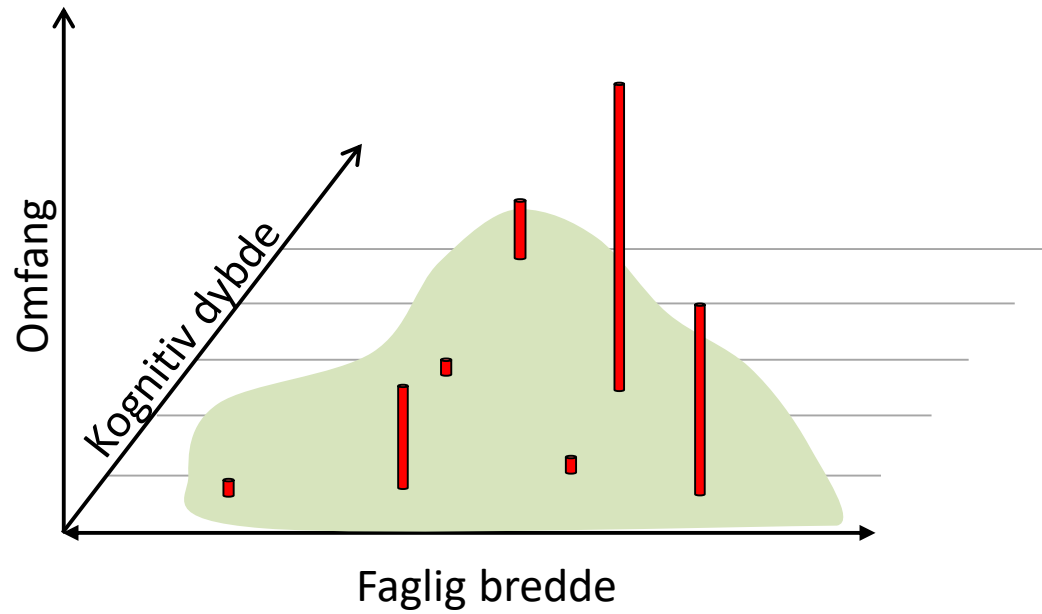
Reliabilitet

God løsning: **Mange** spørsmål (100-150) som dekker **hele** domenet



Reliabilitet

Omfang/betydning – en tredje dimensjon



Blueprint

Plan for sammensetningen av eksamensoppgavene

Tema	Faktakunnskap	Anvendelse
Affektive lidelser	3	7
Atferdsforstyrrelse		
Angst		
Depresjon		
Hyperkinetiske forstyrrelser		
Psykose		
Psykosomatikk		
...		
Sum – antall spørsmål	30	90

Spørsmålstyper

miniessay spørsmål (fritekst)

1 of 4

Hvorfor vil du gi henne diagnosen hjertesvikt? (2 linjer)

multiple response
spørsmål (flere riktige svar)

2 of 4

Hvilke medikamenter i forsiktige doser kan det være aktuelt å starte med allerede første døgnet?

- Antiarytmisk behandling med amiodaron (Cordarone)
- Beta-blokade med metoprolol (SeloZok)
- Slynge-(loop) diuretikum som furosemid (Diural, Furix) eller bumetanid (Burinex)
- Lipidsenkende behandling med statin
- Dobbel platehemmende behandling med acetylsalisylsyre (Albyl E) og clopidogrel (Plavix)
- ACE-hemmer eller angiotensin-reseptorblokker (ARB)

multiple choice
spørsmål (ett riktig svar)

3 of 4

Hvilken av følgende ekkokardiografiske beskrivelser er mest sannsynlig?

- Bicuspid aortaklaff med både stenose og insuffisiens
- Normale forhold ved ultralydundersøkelse av hjertet
- Sklerotisk mitralklaff med stor insuffisiens pga. restriksjon av bakre seil og signifikant stenose sekundært til sklerose i mitralapparatet
- Betydelig tricuspidal lekkasje pga. økt systolisk trykk i h. ventrikkel
- Gjennomgått moderat stort anteroseptalt infarkt og EF ca. 40 %

Oppgaveforfatting

en kunstart som forener
fag og spørreteknikk

Mini-essay – konstruksjon

- Utled en *høynivå kognitiv ferdighet* som essayet skal prøve. Let etter ord som *evaluere, utvikle* i læringsmålene.
- Definer, begrens, fokuser problemet (innholdet) slik at bestillingen blir presis og entydig, omfanget overkommelig og rundtom-skriving vanskelig.
- Gi informasjon om forventet lengde på svaret (hos oss sjelden over 10 linjer).
- Gi en detaljert sensurveiledning: hver karakter/score skal ha en veiledning. Det må være samsvar mellom sensurveiledningen og bestillingen i oppgaveteksten.

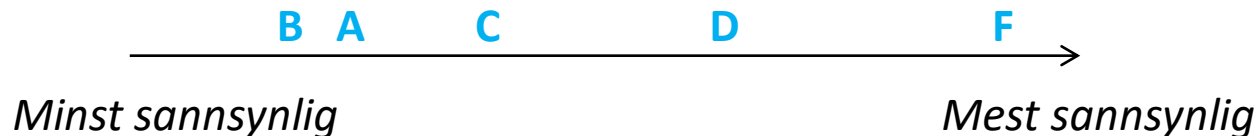
Multiple choice - single best answer

Et *rangerings*spørsmål der man skal finne det *beste* svaret blant de oppgitte - ikke nødvendigvis det helt riktige eller eneste riktige svaret.

Vanligvis 4-5 svaralternativ.

“Hvilken diagnose er **mest** sannsynlig?”

Distansen mellom det korrekte alternativet og de andre (distraktorene) er avgjørende for et godt spørsmål. Alternativene må være på samme akse (dreie seg om det samme):



Multiple choice – skrivekunst

Testklokhet – det riktige svaret kan utledes uten fagkunnskap

- *Grammatikalske* hint: distraktorene henger språklig ikke sammen med spørsmålet (f.eks. nåtid i spørsmål og fortid i ett svar)
- *Logiske* hint: spørsmål og svaralternativ henger ikke logisk sammen (spør f.eks. om medikamentell behandling og ett svar er kirurgisk)
- *Absolutte* termer: virkeligheten er sjelden “alltid” eller “aldri”
- *Gjenganger-ord*: nøkkelord i innledningen går igjen i det riktige svaret
- Et *langt/detaljert/presist* svar: det riktige svaret er mer detaljert enn de andre
- *Konvergens*: det riktige svaret/nøkkelordet gjentas

.....*Hva er riktigst å gjøre nå?*

- øke morfindosen
- redusere morfindosen
- holde morfindosen uendret
- gi midazolam

Multiple response - pick N options item

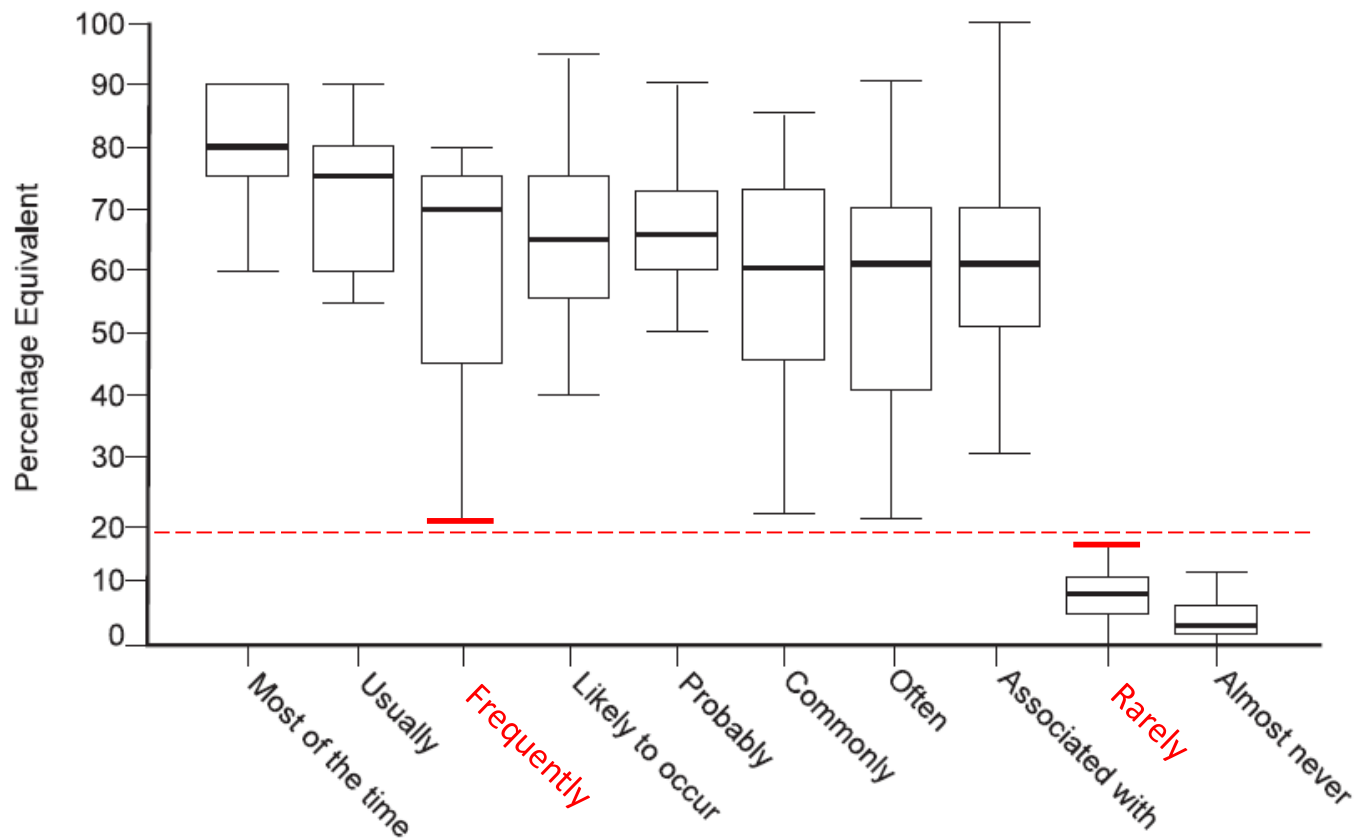
- Kan ha flere korrekte svar.
- Gir flere nyanser/større presisjon i svaret = økt validitet

Hva er **de mest sannsynlige** diagnoser hos denne pasienten?

- A
- B
- C
- D
- E



Multiple response - forståelse av relative begrep



Multiple response

Rangeringsbegrep må kvantifiseres

Hva er de **2** mest sannsynlige diagnoser hos denne pasienten?

- A
- B
- C
- D
- E



Multiple response – antall svaralternativ

For å unngå for høy poengsum ved tilfeldig gjetting:

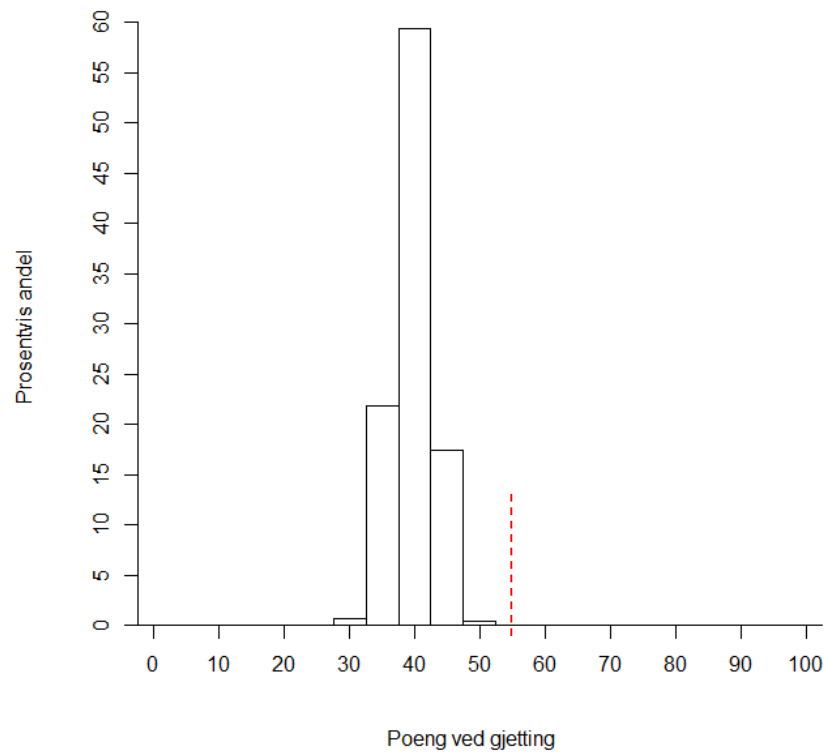
Det må **minst** være 2,5 ganger så mange svaralternativ som antall riktige svar

Gjennomsnittsscore ved å gjette
2 riktige svar av 5 svaralternativ:

40 poeng

Stågrense:

55 poeng



Skriveseminar

- For lærerne i hvert emne
- Ved starten av hvert semester
- Drøfter spørsmål som falt godt og dårlig ut i forrige eksamen
- Går igjennom spørsmålsteknikk (aldri for mye repetisjon...)
- Går igjennom spørsmålsutkast lærerne har laget

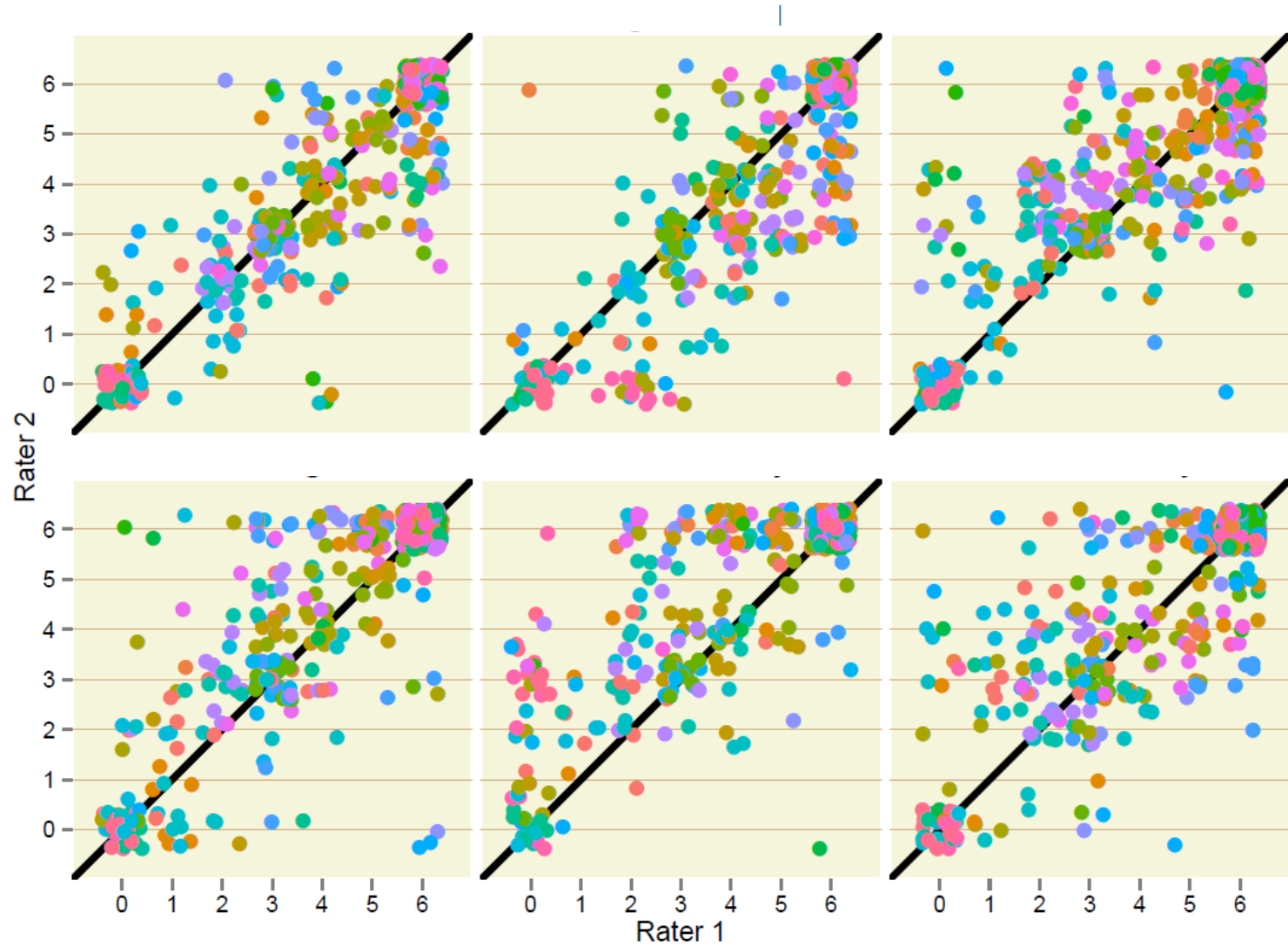
Sensur

Psykometri:

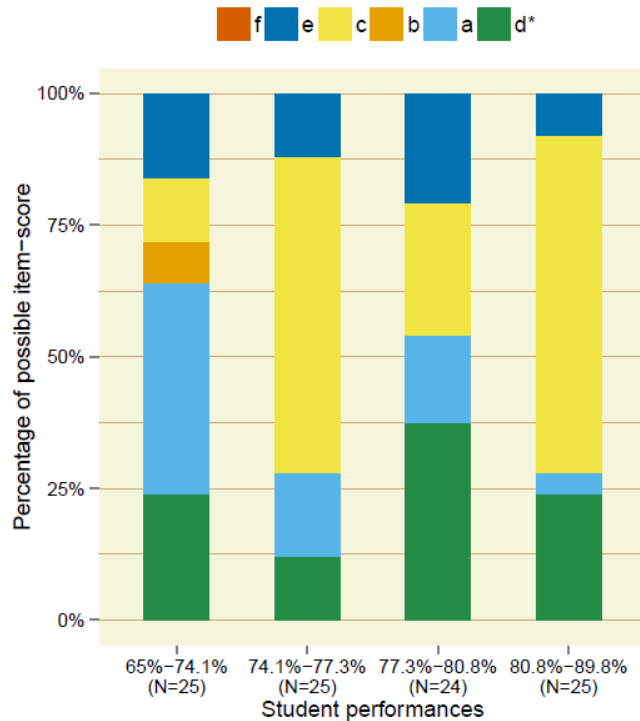
Statistisk analyse av testresultater

- Inter-rater analyse av essay-spørsmål
- Item analyse av alle spørsmål

Interrater reliabilitet - essays



Item analyse

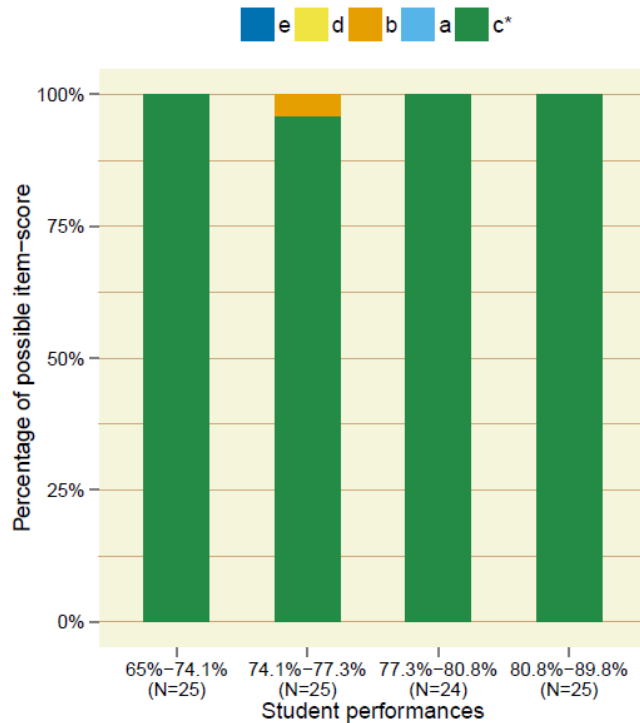


En urimelig lav gjennomsnittscore (item facility) er fakultetets problem, ikke studentenes: temaet kan ligge utenfor læringsmålene, eller være mangelfullt undervist, eller det foreligger «curriculum overload».

Item statistics

- Correlation with totalscore: 0.05
- Item facility: 24.24%

Item analyse – “portalspørsmål”

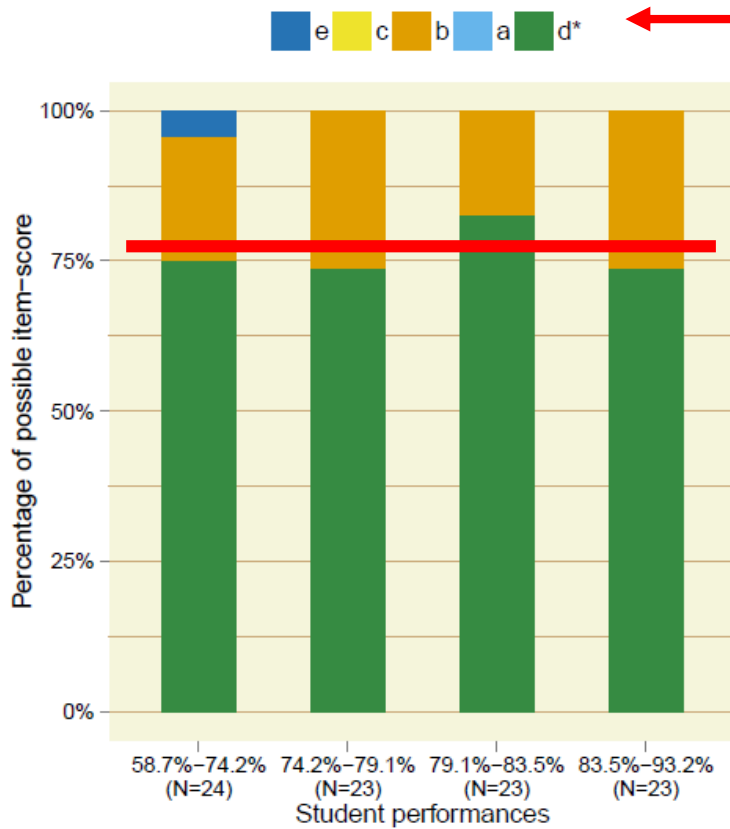


En svært høy gjennomsnitsscore er OK hvis det er et “portalspørsmål” = sentral og avgjørende kunnskap - som alle studentene viser at de kan. Hvis ikke, har spørsmålet ingen verdi for eksamen.

Item statistics

- Correlation with totalscore: 0.08
- Item facility: 98.99%

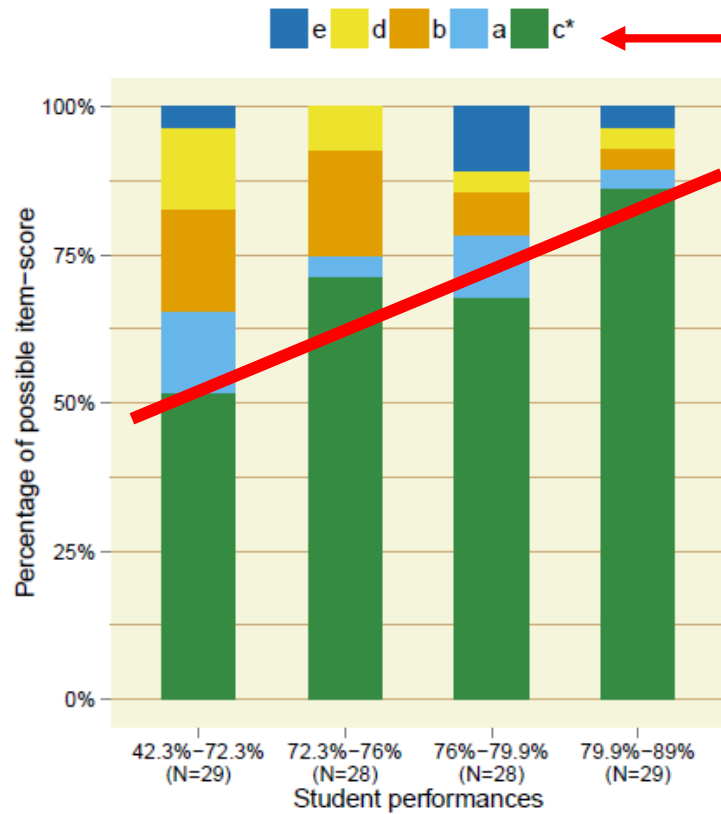
Item analyse – et “dårlig” konstruert spørsmål



Bare 2 av 5 opsjoner ble brukt

Gjennomsnittsscore er normal (76%), men spørsmålet skiller ikke mellom «gode» og «dårlige» studenter.

Item analyse – et “godt” konstruert spørsmål



Alle 5 opsjoner er brukt

Gjennomsnittsscore 67%
Meget godt skille mellom «gode»
og «dårlige» studenter betyr at
karaktersettingen blir reliabel.

Prosess:

- Primærsensur gjøres ferdig 2-3 dager før sensurmøte
- Psykometrisk analyse lages og sendes sensorene
- Sensorene konfererer parvis og retter diskrepanser fra inter-rater analysen
- Sensurmøte:
 - Item-analysen gjennomgås og eventuelle spørsmål ekskluderes fra sensuren
 - Endelige poengsummer reberregnes
 - Karakter settes

Høy kvalitet → gjenbruk av spørsmål → innsynsbegrensning



- Oppgaver som skal gjenbrukes, kan unntas offentliggjøring
- Retten til innsyn i besvarelser må opprettholdes, men innsyn kan skje i kontrollerte former

- Ca 20% av eksamensoppgavene offentliggjøres = treningsoppgaver
- Ca 80% gjenbrukes (bank) **etter revisjon** basert på psykometrien

Organisering for kvalitet

- Validitet gjennom bredt teamarbeid
- Reliabilitet gjennom mange og presise spørsmål
- Spørsmålskvalitet gjennom kontinuerlig opplæring i spørsmålsskriving
- Sensurkvalitet gjennom psykometrisk analyse og avviksbehandling
- Eksamenskvalitet gjennom gjenbruk og forbedring av oppgaver