UiO : Universitetet i Oslo
Institutt for lærerutdanning og skoleforskning

Sensorreliabilitet på skriftlig eksamen i videregående opplæring

Estimater av sensorreliabilitet basert på analyser av karakterforslag fra to uavhengige sensorer før fellessensur - Resultater fra 40 utvalgte fag

Oppdrag som tilhører prosjektet: Ekstern kvalitetssikring av prøver og eksamen

Julius K. Björnsson, ILS/EKVA-UIO

Gustaf B. Skar, Skrivesentret-NTNU

Januar 2021

Innhold

Sammendrag 4
Rapportens formål og oppbygging 4
Bakgrunn 4
Beskrivelse av eksamenssensuren 4
Hva er sensorreliabilitet? 5
Mål på IRR (sensorsamsvar) og IRA (sensorenighet) 5
Skalaer 6
Materiale og analysemetoder 7
Materiale 7
Analysemetoder 9
Resultater fra analyse av sensorenighet og sensorsamsvar 9
Grafisk fremstilling av noen resultater 30
Hvor godt kan eksamen skille mellom elevbesvarelser av ulik kvalitet og sensorers strenghet? 33
Om «Many-facet Rasch Measurement» (MFRM) 33
Analyse av tallene 34
Sammenfatning 38
Konklusjon og veien videre 39
Referanser 40
Vedlegg 1. Antall elever i alle fag delt på år og årstid 43

Sammendrag

Denne rapporten presenterer en analyse av sensorreliabilitet ut ifra de foreløpige karakterene på skriftlig eksamen i videregående opplæring. Analysene er basert på karakterforslagene fra de to sensorene som foretok ekstern sensurering. Analysene er gjort med utgangspunkt i vurderinger av over 700 000 elevbesvarelser fra årene 2015–2019. Den endelige karakteren ble ikke inkludert i analysene. Dette er det beste estimatet vi kan få på sensorreliabilitet, siden det per i dag ikke er mulig å gjøre denne typen analyser på endelige eksamenskarakterer.

Analysen brukte klassiske metoder som beregninger av kappa, vektet kappa og intraklasse-koeffisienter for å evaluere sensorreliabiliteten. Resultatene viser at vurderingen, slik den kommer til uttrykk i karakterforslagene, i noen fag preges av svært god sensorreliabilitet. I andre fag er denne reliabiliteten så lav at vi ikke kan utelukke at eksamenskarakteren ikke bare gjenspeiler den kompetansen kandidatene har, men også vel så mye hvilke sensorer som har vurdert besvarelsen. Det er altså store forskjeller mellom fagene, men også forskjeller i hvor mye karakterene varierer innenfor hvert fag.

I tillegg ble det foretatt kasusstudier der det ble gjort MFRM-analyser av fagene, basert på utvalg der kandidater og sensorer er koplet til hverandre. Dette er en eksplorativ analyse, som må forstås som en tidlig inngang som grunnlag for videre studier. MFRM-analysen viste at eksamen generelt sett var bedre på å skille mellom sensorers strenghet enn kandidaters kompetanse[1]. Videre kunne vi i analysen av delutvalgene for MFRM-analysen notere at det ikke fantes statistisk grunnlag for å skille mellom seks nivåer av kompetanse. I gjennomsnitt klarte eksamen å utskille tre nivåer av kompetanse presist nok, ifølge MFRM analysen, men her er det store forskjeller mellom fagene.

En generell konklusjon fra disse analysene er at det er store variasjoner i sensorreliabilitet i norske eksamener, basert på analyser av foreløpige karakterer. Disse variasjonene har antakeligvis ulike årsaker i forskjellige fag. Dette må derfor utforskes nærmere for hvert fag, slik at passende tiltak kan iverksettes.

Rapportens formål og oppbygging

Bakgrunn

En gruppe nedsatt av Kunnskapsdepartementet i 2018 for å gjennomgå eksamenssystemet, konkluderer med at det er behov for mer forskningsbasert kunnskap om eksamen. De anbefalte også å utvikle et helhetlig rammeverk for kvalitetssikring av både sentralt og lokalt gitt eksamen. Et slikt rammeverk er nå under utvikling av Utdanningsdirektoratet. Direktoratet har også fått i oppdrag å utrede sluttvurderingsordningene i programfagene, og undersøkelsen av sensorreliabiliteten på noen utvalgte skriftlige eksamener i videregående opplæring er en del av dette oppdraget.

Beskrivelse av eksamenssensuren

Utdanningsdirektoratet har i samarbeid med statsforvalterne ansvaret for sensur til sentralt gitt skriftlig eksamen, og fylkeskommunen har ansvaret for sensuren til lokalt gitt eksamen. Prosessen rundt sensuren vil avhenge av hvilken eksamensform det dreier seg om, og foregår på forskjellige måter, for eksempel avhengig av om det er sentralt gitt skriftlig eksamen eller muntlig eksamen. Mens det for sentralt gitt skriftlig eksamen utvikles felles oppgaver, vurderingskriterier og gjennomføres felles sensorskolering, vil det for muntlig eksamen være ulike oppgaver, vurderingskriterier og sensorskoleringer.

Uavhengig av eksamensform sensureres eksamen av to eksterne sensorer. Ved lokalt gitt eksamen kan den ene sensoren være elevens faglærer. Det er disse karakterene, omtalt som *karakterforslag* eller *foreløpige karakterer*, som er grunnlaget for analysene i denne rapporten. Endelig karakter settes etter at sensorene har diskutert seg fram til et felles forslag i et system kalt *fellessensur.* Her er det å utvikle et sterkt tolkningsfellesskap, gjennom blant annet sensorskolering, avgjørende for kvaliteten av den endelige vurderingen. Dermed er fellessensuren en viktig del av eksamenssystemet. Fellessensur, sensorskolering, tolkningsfelleskap og endelige karakterer ved eksamen er ikke undersøkt i denne rapporten. For å kunne undersøke sensorreliabilitet på endelige karakterer, må vi organisere eksamenssensuren annerledes enn vi

[1] Den psykometriske termen er dyktighet.

gjør i dag. En analyse av de foreløpige karakterene, slik vi gjør i denne rapporten, er derfor det beste estimatet vi har på sensorreliabilitet i dagens system. Det er også av betydning hvilke karakterer sensor 1 og sensor 2 har med seg inn i diskusjonen i fellessensuren.

Denne rapporten har som formål å presentere resultater fra en undersøkelse av reliabiliteten i sensuren på eksamen i videregående opplæring i et utvalg fag (N = 40). Formålet med undersøkelsen var å dokumentere sensorreliabilitet i ulike typer av fag over tid. Fagene ble definert ut fra hvordan de organiseres når det gjelder hvem som har ansvar for å utvikle oppgaver og hvem som har ansvar for sensur (som f.eks. sentralt gitt eller lokalt gitt) og ut fra typer oppgaver som vanligvis inngår i eksamen (f.eks. langsvarsoppgaver eller oppgaver som innebærer at eleven skal skrive et kort svar).

Utvalget av fag ble gjort av Utdanningsdirektoratet og inneholder de største fagene innenfor de ulike måtene å organisere eksamen på. Analysen bygger på data fra vår- og høsteksamen fra årene 2015–2019. Utdanningsdirektoratet, ved Øyvind Lind Kvanmo, hjalp også til med beregninger i R for alle fagene.

Rapporten er bygget opp slik: Først redegjør vi for forskjellige metoder som er brukt, og så er sensorreliabilitet i alle fagene analysert med disse metodene. Til slutt trekker vi noen konklusjoner om resultatene, omtaler de som ansees å være viktigst og nevner noen mulige måter å følge opp resultatene i denne rapporten på.

Hva er sensorreliabilitet?

Det første en tenker på når det gjelder sensorreliabilitet, er om vurderinger fra to eller flere sensorer samsvarer eller ikke. Dette er oftest kalt sensorssamsvar (eng. *interrater reliability* [IRR]) eller sensorenighet (eng. *interrater agreement* [IRA]). Mens IRR, altså sensorsamsvar, kan brukes for å estimere om sensorer rangerer elevprestasjoner på en lik måte, brukes IRA, altså enighet, for å estimere hvorvidt sensorer trekker nøyaktig samme slutninger om elevers prestasjoner, dvs. om sensorer er enige om karakteren. En annen, og kanskje mer presis måte å si dette på gjenfinner vi i Tinsley & Weiss (2000):

> "The difference between reliability and agreement is: **Interrater reliability** provides an indication on the extent to which the variance in the ratings is attributable to differences among the rated subjects. […] **Interrater agreement** represents the extent to which the different judges tend to assign exactly the same rating to each subject."

En konsekvens av dette er at IRR fokuserer mest på om målingen er god nok til å avdekke forskjeller i **elevenes** ferdighet, mens IRA kan sies å fokusere mer på **sensorenes** enighet, selv om skillet mellom disse to aspektene ikke alltid er helt klart og tydelig. I denne rapporten rapporterer vi mål for både IRR og IRA.

Mål på IRR (sensorsamsvar) og IRA (sensorenighet)

Det finnes mange mål på sensorreliabilitet, alt etter formålet med undersøkelsen. Fordi denne undersøkelsen ble gjennomført for å kunne danne et godt bilde av sensorreliabiliteten, har vi brukt følgende IRR- og IRA-mål:

- **IRR:**
  - **ICC** (Intraklasse-korrelasjon): Dette målet ligner på «vanlig» (Pearson) korrelasjon, men er noe mer sofistikert og kan brukes når det er flere enn to sensorer. ICC kan også brukes for å si noe om reliabiliteten dersom karakteren baserer seg på vurderingen fra én eller to sensorer. Vanligvis ønsker en korrelasjoner som overstiger 0,70 i klasseromsvurdering og 0,90 i såkalt *high-stakes testing*-sammenheng.
- **IRA:**
  - **Prosent enighet** angir hvor stor andel av et sensorpars vurderinger som er helt like.

- **Kappa** tilsvarer prosent enighet, men kontrollerer for at noen av enighetene er et resultat av tilfeldigheter («chance-corrected»). Dette er også internasjonalt den mest brukte statistiske koeffisienten for å evaluere enighet.
- **Vektet kappa** er en variant av kappa som innebærer at en vekter sensorenighet ulikt avhengig av hvor uenige sensorene er. F.eks. vil uenigheten vektes som større om sensorparet ligger 3 karakterer fra hverandre enn om de ligger 1 karakter fra hverandre. Vanlig kappa tar ikke det hensynet og er derfor en ganske streng vurdering.

Det viktig å være klar over at det finnes forskjellige kriterier for å vurdere om koeffisientene som metodene leverer er gode nok eller ikke. For kappa-resultatene blir en såkalt Landis og Koch (1977) regel oftest brukt, og den er:

Kappa og vektet kappa:

- <0: Ingen enighet
- 0,01-0,20: Ingen til liten enighet
- 0,21-0,40: liten enighet
- 0,41-0,60: Moderat enighet
- 0,60-0,80: Substansiell enighet
- 0,80-1.00: Nesten perfekt enighet

Det finnes andre regler, f.eks. Fleiss-regelen, som sier at en kappa under 0,40 er svak, mellom 0,4 og 0,75 adekvat eller god og over 0,75 strålende. Men det må understrekes at det ikke finnes noen absolutte verdier eller metoder for å vurdere dette.

ICC-koeffisienten må også tolkes, og ofte er verdier under 0,5 ansett som svake, mellom 0,5 og 0,75 ansett som moderate, mellom 0,75 og 0,90 som meget bra og over 0,90 som strålende.

Disse koeffisientene er alle meget følsomme overfor forskjeller i distribusjon av dataene, noe som gjør sammenlikning mellom f.eks. ulike fag vanskelig, hvis ikke distribusjonene er sammenliknbare eller de samme. Dette er spesielt viktig når elevgruppene er små, fordi der kan distribusjonen av karakterene avvike fra normalitet. Hvis data er normalfordelte, så gir disse koeffisientene en nokså god estimering av reliabilitet, men hvis fordelingen er skjev, blir de ofte upresise og enten over- eller undervurderer samsvaret eller enigheten. Det er derfor viktig å evaluere distribusjonen av karakterene. Hvis den viser seg å ikke være normalfordelt, så må andre metoder tas i bruk.[2] Dette er imidlertid ikke gjort i denne første undersøkelsen, men må sees på i en senere analyse.

Det må også poengteres at de klassiske metodene ikke vil gi svar på hva som forårsaker henholdsvis lav IRR og IRA, eller kan si noe om effektene av en gitt IRR eller IRA på reliabiliteten til oppdeling av elever i ulike kompetansesnivåer. I denne rapporten har vi derfor valgt å utvide undersøkelsene til også å inkludere en samtidig analyse av sensorer og elever (noe som kalles MFRM-analyse; se mer nedenfor). I MFRM-analysen har vi gjort studier av klynger av sensorer og elever knyttet til de 40 fagene vi nevnte innledningsvis, for å undersøke i hvilken utstrekning vi kan «separere» elevbesvarelser av ulik kvalitet uavhengig av IRA.

Skalaer

Skalaene eller vurderingskategoriene som blir brukt i forskjellige prøver og eksamener kan være av ulike typer, og kan derfor kreve forskjellige vurderingsmetoder:

[2] Det finnes også andre metoder for å evaluere dette som f.eks. bruk av «COD Coefficient of Determination», som er en kvadrert verdi av fleste av de ovennevnte korrelasjonene som tillater å konkludere om hvor mye den ene distribusjonen er forklart av den andre. Dette er i grunnen det samme som R kvadrert fra en regresjons-analyse som gir forklaringsverdien av analysen. Og i noen tilfeller bruker man standardavviket på karakterer med en standardfeil (LeBreton & Senter, 2007), men dette er ikke vanlig praksis, selv om det finnes.

1) Nominelle kategorier/skala (Nominal scale): På en skala av denne typen er vurderingskategoriene ikke nødvendigvis relatert innbyrdes, men er egentlig merkelapper for forskjellige tilstander/egenskaper/ferdigheter/kompetanser. Her vil man bruke sensorenighet (Kappa-koeffisient, % enighet).
2) Nominelle kategorier i rekkefølge (Ordered nominal scale): Her er kategoriene relatert til hverandre slik at tilstanden/egenskapen/ferdigheten/kompetansen som måles for eksempel går fra noe lite til noe stort. En karakterskala på eksamen er ofte en slik vurdering. Her brukes sensorenighet, men med metoder som tar hensyn til at en forskjell på 1 kategori er en bedre enighet enn større forskjeller (vektet kappa, ICC-Intraklasse-korrelasjon).
3) Kategorier på samme skala med lik distanse mellom alle (Interval level - same difference between categories-measure): Her dreier det seg om at alle kategorier på vurderingsskalaen tilhører samme fenomen/dimensjon/ferdighet/kompetanse og at det er like langt mellom f.eks. 1 og 2 og mellom 5 og 6 (på den vanlige karakterskalaen). Økningen i kompetanse er den samme mellom alle enheter på skalaen. Her brukes sensorsamsvar, for eksempel ICC eller en Pearson-korrelasjon.

Veldig ofte blir eksamenskarakterer sett på som det tredje alternativet her. Men det er i mange tilfeller usikkert, for eksempel når det gjelder vurderinger av frie tekster eller andre kompetanser hvor kompetanseskalaen ikke er påviselig lineær og med samme avstand mellom kategorier. Det er også mulig at sensorene bruker vurderingskategoriene på forskjellig måte ved at noen ser på dem som kategoriske mens andre anser dem for å være nominelle. Derfor er det sikrest å behandle dem som alternativ nummer to, som nominelle kategorier i rekkefølge. Dette er også forklaringen på hvorfor ICC-koeffisienten brukes både på kategori 2 og 3.

Materiale og analysemetoder

Materiale

Materialet består av karakterforslag fra sensor 1 og sensor 2 fra 726.440 elevbesvarelser vurdert i perioden 2015–2019. Endelig karakter ble ikke inkludert i disse analysene og det må en ha i tankene når man leser tallene nedenfor.

Som tidligere nevnt, kan eksamen deles i ulike grupper etter hvem som har ansvar for å lage oppgavene og ansvar for sensur. De tre gruppene er sentralt gitt eksamen med sentral sensur, sentralt gitt eksamen med lokal sensur og lokalt gitt eksamen med lokal sensur. Utvalget her består av de 20 største sentral-sentral kodene, de 10 største sentral-lokal kodene og de 10 største lokal-lokal kodene. Tabell 1 under viser en oversikt over disse fagene.

Tabell 1. Inkluderte fag og antall kandidater

Fagkode	Fagnavn	Oppgave ansvar	Sensur ansvar	Antall kandidater
AMF3102	Anleggsmaskinførerfaget	Sentral	Lokal	2.261
AUT4002	Automatiseringsfaget	Sentral	Lokal	2.307
BUA3102	Barne- og ungdomsarbeiderfaget	Sentral	Lokal	12.050
ELE3002	Elektrikerfaget,	Sentral	Lokal	8.444
ENG1002	Engelsk, Vg1	Sentral	Sentral	24.393
HEA3102	Helsearbeiderfaget	Sentral	Lokal	13.081
HSF1001	Helsefremmendearbeid	Lokal	Lokal	2.176
HSF1003	Yrkesutøvelse	Lokal	Lokal	2.055
IDR2016	Treningslære1	Lokal	Lokal	4.926
IDR2017	Treningslære2	Sentral	Sentral	7.452
LOG3102	Logistikkfaget	Sentral	Lokal	3.746
MAT1001	Matematikk1P-Y	Lokal	Lokal	7.503
MAT1001-0001	Bygg- og anleggsteknikk	Lokal	Lokal	812
MAT1001-0003	Elektrofag	Lokal	Lokal	899
MAT1001-0004	Helse -og oppvekstfag	Lokal	Lokal	1.770
MAT1001-0008	Teknikk og industriellproduksjon	Lokal	Lokal	1.200
MAT1005	Matematikk2P-Y	Sentral	Sentral	35.415
MAT1011	Matematikk1P	Sentral	Sentral	37.473
MAT1015	Matematikk2P	Sentral	Sentral	38.213
MUS2007	Musikk i perspektiv2	Lokal	Lokal	691
NOR1206	Norsk, Vg2	Lokal	Lokal	8.805
NOR1211	Norsk hovedmål, Vg3	Sentral	Sentral	154.484
NOR1212	Norsk sidemål, Vg3	Sentral	Sentral	79.527
NOR1231	Norsk hovedmål, Vg3 p.	Sentral	Sentral	51.651
NOR1232	Norsk sidemål, Vg3 p.	Sentral	Sentral	23.742
REA3002	Biologi 2	Sentral	Sentral	13.039
REA3012	Kjemi 2	Sentral	Sentral	17.828
REA3022	MatematikkR1	Sentral	Sentral	24.993
REA3024	MatematikkR2	Sentral	Sentral	23.806
REA3026	MatematikkS1	Sentral	Sentral	17.730
REA3028	MatematikkS2	Sentral	Sentral	18.024
RHO3102	Renholds-operatørfaget	Sentral	Lokal	4.097
SAM3016	Sosialkunnskap	Sentral	Sentral	14.213
SAM3020	Politikk og menneskerettigheter	Sentral	Sentral	11.351
SAM3023	Rettslære2	Sentral	Sentral	12.686
SAM3038	Psykologi2	Sentral	Sentral	17.759
SLG3102	Salgsfaget	Sentral	Lokal	3.816
SPR3008	Internasjonal engelsk	Sentral	Sentral	15.093
TMF3102	Tømrerfaget	Sentral	Lokal	3.542
YRK3102	Yrkessjåførfaget	Sentral	Lokal	3.387
Total				**726.440**

Tabellen inneholder også antallet kandidater som har tatt disse fagene fra 2015 til 2019. Som det framgår av tabellen, er dette diverse fag, store og små, og de er enten lokalt eller sentralt utviklet og lokalt eller sentralt sensurert. Det største faget var norsk hovedmål, med til sammen 154.484 kandidater og det minste var Musikk i perspektiv2 med 691.

Ansvar for oppgaveutforming og sensur framkommer i tabell 2:

Tabell 2. Ansvar for oppgaveutforming og sensur

Ansvar	Antall	Prosent
Lokalt/Lokalt	30.837	4,2
Sentralt/Lokalt	56.731	7,8
Sentralt/Sentralt	638.872	87,9
Total	**726.440**	**100,0**

Her ser vi at 88 % av eksamenene i analysen er sentralt utviklet og sentralt sensurert. Kun 4,2 % av dem er både lokalt utviklet og lokalt sensurert, mens 7,8 % er sentralt utviklet og lokalt sensurert.

Antall kandidater fordelt på høst- og våreksamen presenteres i tabell 3.

Tabell 3. Inndelingen i høst- og våreksamen

	Antall	Prosent
Høst 2015	20.236	2,8
Høst 2016	26.653	3,7
Høst 2017	27.691	3,8
Høst 2018	27.390	3,8
Høst 2019	28.074	3,9
Vår 2016	141.631	19,5
Vår 2017	146.856	20,2
Vår 2018	152.978	21,1
Vår 2019	154.931	21,3
Total	**726.440**	**100,0**

Det er altså omtrent 5 ganger flere kandidater som tar eksamen om våren enn på høsten, blant annet fordi noen fag ikke har høsteksamen. Vedlegg 1 viser fordelingen for alle fagene mellom høst og vår og for alle årene.

Analysemetoder

Vi brukte *SPSS 26* (IBM, 2019) og MS Excel for å bearbeide datafilene. Videre brukte vi R-pakken «IRR» (Gamer, Lemon, Fellows, 2019), *STATA 15* (StataCorp. 2017) og en egenprodusert Excel-fil med kappa-beregninger for å estimere IRA- og IRR-mål. MFRM-analysen ble gjennomført i programmet FACETS 3.8 (Linacre, 2020).

Resultater fra analyse av sensorenighet og sensorsamsvar

I dette avsnittet presenteres én tabell for hvert av de 40 fagene, med prosent enighet, kappa-koeffisient, vektet kappa-koeffisient og intraklasse-korrelasjon (ICC). Dette er presentert separat for vår- og høsteksamen. I tillegg vises det gjennomsnitt for vår- og høsteksamen og en total sammenfatning for faget gjennom alle årene.

Her må man huske at prosent enighet er et tall som må være ganske høyt, ettersom en helt tilfeldig distribusjon av de seks karakterene gir en sannsynlighet på omtrent 17 %, dvs. 1/6 for hver av karakterene. Derfor vil tilfeldig bruk av karakterene 1 til 6 for to sensorer være 1/6 *1/6, eller omtrent 3% sjanse for å få samme karakter, ettersom disse to karaktersettingene er uavhengige av hverandre.

En kappa-koeffisient er ganske streng og krever total enighet, og den er egentlig en binær målestokk. Som tidligere nevnt kan derfor **vektet kappa** være en mer anvendbar måling, fordi den tar hensyn til distansen mellom de gitte karakterene. Den reflekterer da at to karakterer ved siden av hverandre, f.eks. 3 og 4, reflekterer større enighet enn f.eks. 2 og 4 fra samme elevsvar. ICC er også inkludert ettersom den gir en helhetlig vurdering av samsvaret. ICC går ut ifra at karakterene er på en sammenhengende skala, men dette er kanskje ikke alltid tilfellet for eksamen. Hvis distribusjonen av karakterer er den samme eller meget lik, kan ICC i noen tilfeller også bli ganske høy, selv om ingen karakterer er de samme hos de to sensorene. Hvis den ene sensoren er konsistent én karakter under den andre, blir kappa meget lav, mens ICC kan bli høy.

Ved å bruke alle disse målestokkene på sensorreliabilitet, burde det foreligge en ganske bred evaluering av fagets sensorsamsvar og enighet. I rapporteringer av sensorreliabilitet er det vanlig praksis å bruke forskjellige koeffisienter på samme måte som vi har gjort her. Den klassiske kappa-koeffisienten er vanligst.

Når disse tallene vurderes, må man også huske at det kan være naturlige forskjeller mellom elevgruppene fra vår og høst, som kan ha en effekt her. Disse reliabilitetsberegningene sier følgelig ingenting om forskjeller i elevenes kompetanse eller noe om det er systematiske forskjeller mellom vår og høst.

Når disse resultatene vurderes, foreslår vi å bruke de kriteriene som er beskrevet på side 6, samt prosent enighet. I tabellene under er strålende resultater fremhevet i teksten, dvs. prosent enighet over 90 %, en kappa eller vektet kappa over 0,8 og en ICC over 0,9. I tabellene er perioden merket med en stjerne hvis noen av disse kriteriene er oppfylt.

Tabell 4.1

AMF3102	**Anleggsmaskinførerfaget, skriftlig**				**Sentral-Lokal**
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	48,14	0,32	0,48	0,64	295
H_2017	45,04	0,25	0,41	0,56	282
H_2018	37,6	0,16	0,37	0,57	242
H_2019	42,68	0,22	0,41	0,58	246
Høst:	***43,37***	***0,24***	***0,42***	***0,59***	***1065***
V_2016	61,2	0,5	0,62	0,73	299
V_2017	56,36	0,41	0,58	0,74	346
V_2018	50,17	0,31	0,48	0,66	299
V_2019	47,22	0,3	0,46	0,61	252
Vår:	***53,74***	***0,38***	***0,54***	***0,69***	***1196***
Samlet:	**48,55**	**0,31**	**0,48**	**0,64**	**2261**

Tabell 4.2

AUT4002	Tverrfaglig eksamen, automatiseringsfaget				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016*	98,61	0,98	0,99	0,99	72
H_2017*	94	0,91	0,94	0,97	50
H_2018*	94,12	0,91	0,94	0,97	51
H_2019	62,32	0,5	0,68	0,83	69
Høst:	87,26	**0,83**	**0,89**	**0,94**	**242**
V_2016	50,81	0,37	0,54	0,68	494
V_2017	64,7	0,53	0,64	0,73	541
V_2018*	80,27	0,72	0,79	0,86	527
V_2019	63,15	0,51	0,65	0,78	502
Vår:	**64,73**	**0,53**	**0,66**	**0,76**	**2064**
Samlet:	**76,00**	**0,68**	**0,77**	**0,85**	**2306**

Tabell 4.3

BUA3102	Barne- og ungdomsarbeiderfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	63,21	0,51	0,63	0,73	1079
H_2017	68,25	0,59	0,7	0,78	1244
H_2018	60,22	0,49	0,63	0,74	1111
H_2019	74,18	0,67	0,79	0,88	883
Høst:	***66,47***	***0,57***	***0,69***	***0,78***	***4317***
V_2016	63,78	0,52	0,64	0,75	2079
V_2017	54,83	0,42	0,58	0,71	1955
V_2018	60,32	0,48	0,61	0,73	2049
V_2019	63,41	0,53	0,68	0,79	1648
Vår:	***60,59***	***0,49***	***0,63***	***0,75***	***7731***
Samlet:	**63,53**	**0,53**	**0,66**	**0,76**	**12048**

Tabell 4.4

ELE3002	Elektrikerfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	56,1	0,42	0,53	0,64	615
H_2017	60,06	0,48	0,6	0,71	646
H_2018	50,19	0,35	0,5	0,63	538
H_2019	50,09	0,35	0,51	0,65	577
Høst:	***54,11***	***0,40***	***0,54***	***0,66***	***2376***
V_2016	50,21	0,35	0,48	0,61	1402
V_2017	56,13	0,43	0,57	0,69	1534
V_2018	52,86	0,39	0,54	0,67	1485
V_2019	49,09	0,33	0,49	0,64	1644
Vår:	***52,07***	***0,38***	***0,52***	***0,65***	***6065***
Samlet:	**53,09**	**0,39**	**0,53**	**0,66**	**8441**

Tabell 4.5

ENG1002	Engelsk, Vg1 studieforberedende utdanningsprogram				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	45,26	0,3	0,53	0,71	1076
H_2016	46,46	0,31	0,55	0,74	1201
H_2017	43,58	0,28	0,52	0,71	1278
H_2018	41,84	0,26	0,52	0,72	1281
H_2019	43,25	0,29	0,53	0,72	1304
Høst:	***44,08***	***0,29***	***0,53***	***0,72***	***6140***
V_2016	42,22	0,25	0,47	0,66	4180
V_2017	42,48	0,24	0,45	0,64	4772
V_2018	40,83	0,23	0,45	0,63	4746
V_2019	39,43	0,21	0,42	0,61	4547
Vår:	***41,24***	***0,23***	***0,45***	***0,64***	***18245***
Samlet:	**42,82**	**0,26**	**0,49**	**0,68**	**24385**

Tabell 4.6

HEA3102	Helsearbeiderfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	58,06	0,43	0,57	0,68	1235
H_2017	53,5	0,39	0,48	0,56	1471
H_2018	63,31	0,51	0,62	0,71	1270
H_2019	66,81	0,53	0,61	0,67	1389
Høst:	***60,42***	***0,47***	***0,57***	***0,66***	***5365***
V_2016	66,11	0,56	0,67	0,76	1782
V_2017	62,34	0,5	0,61	0,7	2005
V_2018	56,31	0,43	0,56	0,68	1877
V_2019	65,04	0,53	0,62	0,7	2048
Vår:	***62,45***	***0,51***	***0,62***	***0,71***	***7712***
Samlet:	**61,44**	**0,49**	**0,59**	**0,68**	**13077**

Tabell 4.7

HSF1001	Helsefremmende arbeid				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	65,62	0,57	0,75	0,88	317
H_2017	46,26	0,32	0,47	0,58	348
H_2018	62,5	0,52	0,67	0,79	328
H_2019	69,38	0,59	0,73	0,85	320
Høst:	***60,94***	***0,50***	***0,66***	***0,78***	***1313***
V_2016	75,51	0,69	0,81	0,9	196
V_2017	48,23	0,34	0,5	0,63	226
V_2018	65	0,56	0,73	0,84	220
V_2019	65,91	0,57	0,7	0,81	220
Vår:	***63,66***	***0,54***	***0,69***	***0,80***	***862***
Samlet:	**62,30**	**0,52**	**0,67**	**0,79**	**2175**

Tabell 4.8

HSF1003	Yrkesutøvelse				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016*	82,09	0,77	0,86	0,93	296
H_2017	67,85	0,59	0,74	0,85	339
H_2018	57,19	0,46	0,63	0,77	306
H_2019	68,24	0,6	0,74	0,85	296
Høst:	***68,84***	***0,61***	***0,74***	***0,85***	***1237***
V_2016*	86,55	0,83	0,89	0,94	171
V_2017*	89,05	0,86	0,91	0,96	210
V_2018	76,39	0,7	0,76	0,81	216
V_2019	65,91	0,56	0,69	0,8	220
Vår:	***79,48***	***0,74***	***0,81***	***0,88***	***817***
Samlet:	**74,16**	**0,67**	**0,78**	**0,86**	**2054**

Tabell 4.9

IDR2016	Treningslære 1				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	-	-	-	-	0
H_2017	53,85	0,42	0,56	0,66	78
H_2018	49,04	0,36	0,59	0,76	104
H_2019	66,67	0,57	0,75	0,88	117
Høst:	***56,52***	***0,45***	***0,63***	***0,77***	***299***
V_2016	-	-	-	-	0
V_2017	54,11	0,39	0,57	0,73	1301
V_2018	43,68	0,28	0,51	0,69	1488
V_2019	48,75	0,34	0,57	0,75	1836
Vår:	***48,85***	***0,34***	***0,55***	***0,72***	***4625***
Samlet:	**52,68**	**0,39**	**0,59**	**0,75**	**4924**

Tabell 4.10

IDR2017	Treningslære 2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	-	-	-	-	0
H_2017	-	-	-	-	0
H_2018	53,68	0,38	0,6	0,79	95
H_2019	41,59	0,21	-	0,55	113
Høst:	***47,64***	***0,30***	***0,60***	***0,67***	***208***
V_2016	-	-	-	-	0
V_2017	-	-	-	-	0
V_2018	44,01	0,26	0,46	0,65	3547
V_2019	42,9	0,25	0,45	0,63	3697
Vår:	***43,46***	***0,26***	***0,46***	***0,64***	***7244***
Samlet:	**45,55**	**0,28**	**0,50**	**0,66**	**7452**

Tabell 4.11

LOG3102	Logistikkfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	55,77	0,43	0,6	0,74	520
H_2017	44,69	0,29	0,41	0,49	414
H_2018	36,92	0,21	0,42	0,6	428
H_2019	46,39	0,31	0,51	0,67	388
Høst:	***45,94***	***0,31***	***0,49***	***0,63***	***1750***
V_2016	79,42	0,75	0,82	0,89	515
V_2017	45,7	0,31	0,5	0,66	512
V_2018	35,59	0,19	0,3	0,38	472
V_2019	49,8	0,37	0,5	0,59	496
Vår:	***52,63***	***0,41***	***0,53***	***0,63***	***1995***
Samlet:	**49,29**	**0,36**	**0,51**	**0,63**	**3745**

Tabell 4.12

MAT1001	Matematikk 1P-Y				Lokal-lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016*	84,26	0,78	0,89	0,95	591
H_2017*	77,03	0,71	0,86	0,94	653
H_2018*	79,15	0,74	0,88	0,95	753
H_2019*	73,21	0,67	0,85	0,95	836
Høst:	***78,41***	***0,73***	***0,87***	***0,95***	***2833***
V_2016*	74,65	0,68	0,82	0,91	1491
V_2017*	75,29	0,68	0,84	0,94	789
V_2018*	76,6	0,71	0,85	0,93	1346
V_2019*	72,17	0,66	0,83	0,93	1042
Vår:	***74,68***	***0,68***	***0,84***	***0,93***	***4668***
Samlet:	**76,55**	**0,70**	**0,85**	**0,94**	**7501**

Tabell 4.13

MAT1001-0001	Bygg- og anleggsteknikk				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	75	0,53	0,51	0,51	20
H_2017	60,87	0,32	0,38	0,43	23
H_2018*	78,33	0,67	0,8	0,91	60
H_2019*	97,44	0,93	0,93	0,95	39
Høst:	***77,91***	***0,61***	***0,66***	***0,70***	***142***
V_2016	-	-	-	-	0
V_2017*	63,13	0,55	0,78	0,91	160
V_2018*	71,6	0,65	0,82	0,93	243
V_2019	58,43	0,49	0,74	0,89	267
Vår:	***64,39***	***0,56***	***0,78***	***0,91***	***670***
Samlet:	**72,11**	**0,59**	**0,71**	**0,79**	**812**

Tabell 4.14

MAT1001-0003	Elektrofag				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	66,67	0,51	0,62	0,75	9
H_2017	66,67	0,5	0,62	0,77	12
H_2018*	84,62	0,73	0,8	0,88	13
H_2019*	81,25	0,73	0,82	0,91	16
Høst:	***74,80***	***0,62***	***0,72***	***0,83***	***50***
V_2016	-	-	-	-	0
V_2017	64,71	0,56	0,75	0,89	221
V_2018*	71,25	0,64	0,82	0,93	320
V_2019*	70,45	0,62	0,79	0,91	308
Vår:	***68,80***	***0,61***	***0,79***	***0,91***	***849***
Samlet:	**72,23**	**0,61**	**0,75**	**0,86**	**899**

Tabell 4.15

MAT1001-0004	Helse- og oppvekstfag				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016*	92,68	0,81	0,85	0,89	41
H_2017	83,05	0,6	0,69	0,81	118
H_2018	86,27	0,74	0,82	0,89	102
H_2019*	87,88	0,73	0,84	0,93	99
Høst:	**87,47**	**0,72**	**0,80**	**0,88**	**360**
V_2016	-	-	-	-	0
V_2017*	69,78	0,61	0,78	0,9	321
V_2018*	76,58	0,71	0,85	0,94	444
V_2019*	74,11	0,68	0,83	0,93	645
Vår:	**73,49**	**0,67**	**0,82**	**0,92**	**1410**
Samlet:	**81,48**	**0,7**	**0,81**	**0,9**	**1770**

Tabell 4.16

MAT1001-0008	Teknikk og industriell produksjon				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	85,37	0,57	0,67	0,81	41
H_2017	66,67	0,29	0,46	0,66	54
H_2018*	78,87	0,67	0,83	0,94	71
H_2019*	86,96	0,66	0,74	0,92	46
Høst:	**79,47**	**0,55**	**0,67**	**0,83**	**212**
V_2016	-	-	-	-	0
V_2017	70,35	0,63	0,78	0,89	199
V_2018*	66,58	0,59	0,79	0,91	404
V_2019	69,87	0,62	0,77	0,89	385
Vår:	**68,93**	**0,61**	**0,78**	**0,9**	**988**
Samlet:	**74,95**	**0,58**	**0,72**	**0,86**	**1200**

Tabell 4.17

MAT1005	Matematikk 2P-Y				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	82,86	0,7	0,79	0,88	1686
H_2016*	77,82	0,68	0,8	0,91	1623
H_2017*	78,7	0,69	0,81	0,91	1732
H_2018*	76,12	0,67	0,8	0,91	1784
H_2019*	77,63	0,69	0,81	0,91	1851
Høst:	***78,63***	***0,69***	***0,80***	***0,90***	***8676***
V_2016*	74,66	0,67	0,81	0,91	6480
V_2017*	77,35	0,71	0,84	0,93	6993
V_2018*	74,14	0,67	0,81	0,92	6803
V_2019*	74,95	0,67	0,81	0,92	6463
Vår:	***75,28***	***0,68***	***0,82***	***0,92***	***26739***
Samlet:	**77,14**	**0,68**	**0,81**	**0,91**	**35415**

Tabell 4.18

MAT1011	Matematikk 1P				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	76,92	0,7	0,84	0,93	3800
H_2016*	77,08	0,71	0,85	0,94	4054
H_2017*	77,68	0,72	0,86	0,94	4431
H_2018*	78,64	0,74	0,88	0,95	4350
H_2019*	72,08	0,66	0,83	0,93	4670
Høst:	***76,48***	***0,71***	***0,85***	***0,94***	***21305***
V_2016*	75,44	0,69	0,84	0,93	4112
V_2017*	75,3	0,68	0,82	0,92	4226
V_2018*	74,84	0,69	0,83	0,93	3705
V_2019*	75,07	0,68	0,82	0,92	4123
Vår:	***75,16***	***0,69***	***0,83***	***0,93***	***16166***
Samlet:	**75,89**	**0,70**	**0,84**	**0,93**	**37471**

Tabell 4.19

MAT1015	Matematikk 2P				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	78,85	0,68	0,79	0,89	1995
H_2016*	76,6	0,69	0,83	0,92	1816
H_2017*	76,64	0,69	0,82	0,91	1978
H_2018*	77,92	0,71	0,84	0,93	1893
H_2019*	76,84	0,7	0,83	0,92	2111
Høst:	***77,37***	***0,69***	***0,82***	***0,91***	***9793***
V_2016*	72,07	0,64	0,79	0,9	6842
V_2017*	75,12	0,69	0,83	0,92	6804
V_2018*	72,73	0,66	0,81	0,92	7396
V_2019*	72,94	0,65	0,8	0,9	7377
Vår:	***73,22***	***0,66***	***0,81***	***0,91***	***28419***
Samlet:	**75,52**	**0,68**	**0,82**	**0,91**	**38212**

Tabell 4.20

MUS2007	Musikk i perspektiv 2				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	20	0	-	0	5
H_2017	44,44	0,3	0,68	0,89	9
H_2018	66,67	0,5	0,57	0,71	6
H_2019	-	-	-	-	0
Høst:	***43,70***	***0,27***	***0,68***	***0,53***	***20***
V_2016	100	1	1	1	7
V_2017	44,19	0,29	0,47	0,62	86
V_2018	34,2	0,16	0,33	0,48	269
V_2019	44,81	0,28	0,47	0,65	308
Vår:	***55,80***	***0,43***	***0,42***	***0,69***	***670***
Samlet:	**50,62**	**0,36**	**0,59**	**0,62**	**690**

Tabell 4.21

NOR1206	Norsk, Vg2 yrkesfaglige utdanningsprogram				Lokal-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	48,64	0,34	0,56	0,73	220
H_2017	63,14	0,52	0,68	0,8	236
H_2018	47,6	0,33	0,57	0,75	229
H_2019	43,14	0,26	0,49	0,69	204
Høst:	***50,63***	***0,36***	***0,58***	***0,74***	***889***
V_2016	54,6	0,4	0,55	0,7	1456
V_2017	42,18	0,23	0,43	0,61	1944
V_2018	47,05	0,29	0,48	0,64	2170
V_2019	44,09	0,25	0,43	0,6	2345
Vår:	***46,98***	***0,29***	***0,47***	***0,64***	***7915***
Samlet:	**48,81**	**0,33**	**0,52**	**0,69**	**8804**

Tabell 4.22

NOR1211	Norsk hovedmål, Vg3 studieforb. utdanningsprogram, skriftlig				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	43,48	0,25	0,45	0,62	2261
H_2016	42,2	0,25	0,46	0,63	2199
H_2017	42,45	0,25	0,47	0,66	1981
H_2018	44,37	0,28	0,49	0,67	1918
H_2019	48,01	0,32	0,53	0,71	1985
Høst:	***44,10***	***0,27***	***0,48***	***0,66***	***10344***
V_2016	39,92	0,2	0,39	0,56	34870
V_2017	40,55	0,21	0,4	0,57	35389
V_2018	41,28	0,21	0,4	0,58	35745
V_2019	41,59	0,22	0,4	0,58	38125
Vår:	***40,84***	***0,21***	***0,40***	***0,57***	***144129***
Samlet:	**42,65**	**0,24**	**0,44**	**0,62**	**154473**

Tabell 4.23

NOR1212	Norsk sidemål, Vg3 studieforb. utdanningsprogram, skriftlig				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	45,93	0,29	0,49	0,67	1461
H_2016	43,5	0,26	0,48	0,67	1439
H_2017	43,45	0,25	0,46	0,65	1436
H_2018	43,97	0,26	0,47	0,66	1401
H_2019	45,08	0,27	0,48	0,67	1524
Høst:	***44,39***	***0,27***	***0,48***	***0,66***	***7261***
V_2016	42,03	0,23	0,42	0,6	17377
V_2017	41,72	0,22	0,41	0,59	17501
V_2018	44,12	0,25	0,44	0,62	18151
V_2019	42,2	0,23	0,43	0,61	19232
Vår:	***42,52***	***0,23***	***0,43***	***0,61***	***72261***
Samlet:	**43,56**	**0,25**	**0,45**	**0,64**	**79522**

Tabell 4.24

NOR1231	Norsk hovedmål, Vg3 påbygging til gen studiekompetanse, skriftlig				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	51,47	0,33	0,5	0,66	1053
H_2016	50,05	0,32	0,5	0,67	1017
H_2017	49,78	0,32	0,52	0,69	902
H_2018	49,13	0,32	0,51	0,68	859
H_2019	54,42	0,38	0,58	0,75	724
Høst:	***50,97***	***0,33***	***0,52***	***0,69***	***4555***
V_2016	43,47	0,23	0,39	0,56	12419
V_2017	44,1	0,24	0,41	0,58	12245
V_2018	44,47	0,24	0,43	0,61	12252
V_2019	44,81	0,24	0,42	0,59	10175
Vår:	***44,21***	***0,24***	***0,41***	***0,59***	***47091***
Samlet:	**47,97**	**0,29**	**0,47**	**0,64**	**51646**

Tabell 4.25

NOR1232	Norsk sidemål, Vg3 påbygging til gen. studiekompetanse, skriftlig				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	49,92	0,28	0,43	0,6	621
H_2016	49,29	0,29	0,48	0,65	635
H_2017	51,83	0,33	-	0,66	546
H_2018	51,53	0,33	0,49	0,66	621
H_2019	54,46	0,35	0,51	0,68	560
Høst:	***51,41***	***0,32***	***0,48***	***0,65***	***2983***
V_2016	42,97	0,22	0,39	0,56	5252
V_2017	45,71	0,24	0,41	0,57	5347
V_2018	47,85	0,27	0,44	0,61	5430
V_2019	46,58	0,26	0,44	0,61	4727
Vår:	***45,78***	***0,25***	***0,42***	***0,59***	***20756***
Samlet:	**48,90**	**0,29**	**0,45**	**0,62**	**23739**

Tabell 4.26

REA3002	Biologi 2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	53,76	0,42	0,67	0,84	279
H_2016	58,59	0,47	0,7	0,86	297
H_2017	51,31	0,38	0,64	0,82	306
H_2018*	60,55	0,49	0,75	0,9	256
H_2019*	65,52	0,57	0,78	0,91	261
Høst:	***57,95***	***0,47***	***0,71***	***0,87***	***1399***
V_2016	55,95	0,44	0,67	0,83	2917
V_2017	57,25	0,46	0,68	0,84	2980
V_2018	56,45	0,45	0,67	0,83	2900
V_2019	62,4	0,53	0,74	0,88	2843
Vår:	***58,01***	***0,47***	***0,69***	***0,85***	***11640***
Samlet:	**57,98**	**0,47**	**0,70**	**0,86**	**13039**

Tabell 4.27

REA3012	Kjemi 2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	78,64	0,73	0,87	0,95	412
H_2016*	72,84	0,67	0,84	0,94	405
H_2017*	66,87	0,58	0,79	0,91	501
H_2018*	69,44	0,62	0,81	0,93	373
Høst:	***71,95***	***0,65***	***0,83***	***0,93***	***1691***
H_2019*	77,37	0,73	0,88	0,96	380
V_2016*	70,65	0,64	0,82	0,93	3857
V_2017*	69,64	0,63	0,8	0,91	3887
V_2018*	75,84	0,7	0,85	0,94	4052
V_2019*	69,38	0,62	0,81	0,92	3961
Vår:	***71,38***	***0,65***	***0,82***	***0,93***	***15757***
Samlet:	**71,69**	**0,65**	**0,82**	**0,93**	**16137**

Tabell 4.28

REA3022	Matematikk R1				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	76,47	0,7	0,84	0,93	1823
H_2016*	76,74	0,71	0,86	0,94	1526
H_2017*	75,17	0,69	0,83	0,93	1486
H_2018*	79,55	0,74	0,85	0,93	1702
H_2019*	78,24	0,72	0,85	0,93	1636
Høst:	***77,23***	***0,71***	***0,85***	***0,93***	***8173***
V_2016*	74,71	0,68	0,83	0,93	4238
V_2017*	75,96	0,7	0,84	0,93	4259
V_2018*	75,94	0,7	0,84	0,94	4223
V_2019*	73,53	0,67	0,82	0,92	4099
Vår:	***75,04***	***0,69***	***0,83***	***0,93***	***16819***
Samlet:	**76,26**	**0,70**	**0,84**	**0,93**	**24992**

Tabell 4.29

REA3024	Matematikk R2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	69,4	0,62	0,79	0,9	634
H_2016*	74	0,68	0,83	0,93	650
H_2017*	78,78	0,73	0,87	0,95	608
H_2018*	72,11	0,65	0,82	0,93	631
H_2019*	78,13	0,72	0,85	0,94	654
Høst:	***74,48***	***0,68***	***0,83***	***0,93***	***3177***
V_2017*	76,44	0,71	0,86	0,95	5001
V_2018*	77,33	0,72	0,87	0,95	5103
V_2019*	73,08	0,67	0,82	0,93	5356
Vår:	***75,33***	***0,70***	***0,85***	***0,94***	***18637***
Samlet	**74**	**0,68**	**0,84**	**0,93**	**23805**

Tabell 4.30

REA3026	Matematikk S1				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	79,93	0,74	0,85	0,93	1156
H_2016*	71,82	0,65	0,83	0,93	1086
H_2017*	73,9	0,67	0,83	0,93	1088
H_2018*	77,79	0,72	0,86	0,94	1157
H_2019*	79,7	0,74	0,85	0,93	1187
Høst:	***76,63***	***0,70***	***0,84***	***0,93***	***5674***
V_2016*	73,4	0,65	0,8	0,9	2902
V_2017*	73,7	0,67	0,82	0,91	2989
V_2018*	76,13	0,7	0,84	0,93	2987
V_2019*	75,06	0,68	0,82	0,92	3176
Vår:	***74,57***	***0,68***	***0,82***	***0,92***	***12054***
Samlet:	**75,71**	**0,69**	**0,83**	**0,92**	**17728**

Tabell 4.31

REA3028	Matematikk S2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015*	76,08	0,69	0,84	0,93	556
H_2016*	76,54	0,69	0,83	0,93	439
H_2017*	80,93	0,76	0,87	0,94	388
H_2018*	80,34	0,75	0,87	0,94	468
H_2019*	83,04	0,79	0,89	0,96	395
Høst:	***79,39***	***0,74***	***0,86***	***0,94***	***2246***
V_2016*	76,48	0,7	0,83	0,92	3550
V_2017*	79,67	0,75	0,86	0,94	4102
V_2018*	76,28	0,69	0,82	0,91	3929
V_2019*	77,22	0,71	0,84	0,92	4196
Vår:	***77,41***	***0,71***	***0,84***	***0,92***	***15777***
Samlet:	**78,51**	**0,73**	**0,85**	**0,93**	**18023**

Tabell 4.32

RHO3102	Renholdsoperatørfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	62,07	0,5	0,62	0,74	435
H_2017	59,41	0,47	0,61	0,72	510
H_2018	62,87	0,51	0,63	0,73	404
H_2019	52,7	0,39	0,53	0,65	518
Høst:	***59,26***	***0,47***	***0,60***	***0,71***	***1867***
V_2016*	82,88	0,78	0,84	0,89	520
V_2017	69,69	0,6	0,71	0,8	650
V_2018	45,72	0,3	0,5	0,66	479
V_2019	47,67	0,32	0,5	0,65	579
Vår:	***61,49***	***0,50***	***0,64***	***0,75***	***2228***
Samlet:	**60,38**	**0,48**	**0,62**	**0,73**	**4095**

Tabell 4.33

SAM3016	Sosialkunnskap				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	43,28	0,27	0,48	0,66	409
H_2016	41,36	0,23	0,43	0,61	411
H_2017	34,01	0,14	0,34	0,53	344
H_2018	42,09	0,25	0,44	0,61	354
H_2019	39,76	0,21	0,46	0,67	337
Høst:	***40,10***	***0,22***	***0,43***	***0,62***	***1855***
V_2016	37,32	0,18	0,39	0,57	2996
V_2017	39,41	0,2	0,4	0,58	3002
V_2018	40,3	0,22	0,44	0,62	3035
V_2019	38,44	0,2	0,41	0,61	3322
Vår:	***38,87***	***0,20***	***0,41***	***0,60***	***12355***
Samlet:	**39,55**	**0,21**	**0,42**	**0,61**	**14210**

Tabell 4.34

SAM3020	Politikk og menneskerettigheter				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	49,15	0,35	0,55	0,72	177
H_2016	45,05	0,28	0,48	0,66	182
H_2017	34,38	0,19	0,42	0,59	128
H_2018	52,21	0,34	0,53	0,69	136
H_2019	28,85	0,09	0,32	0,55	156
Høst:	***41,93***	***0,25***	***0,46***	***0,64***	***779***
V_2016	36,7	0,19	0,41	0,59	2654
V_2017	36,56	0,19	0,39	0,57	2623
V_2018	35,76	0,17	0,38	0,58	2746
V_2019	39,36	0,22	0,43	0,62	2548
Vår:	***37,10***	***0,19***	***0,40***	***0,59***	***10571***
Samlet:	**39,78**	**0,22**	**0,43**	**0,62**	**11350**

Tabell 4.35

SAM3023	Rettslære 2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	45,45	0,31	0,55	0,75	165
H_2016	35,42	0,16	0,34	0,52	192
H_2017	45,7	0,28	0,48	0,66	151
H_2018	40,48	0,24	0,5	0,71	126
H_2019	40,52	0,26	0,52	0,72	116
Høst:	***41,51***	***0,25***	***0,48***	***0,67***	***750***
V_2016	41,99	0,23	0,42	0,6	3084
V_2017	42,37	0,24	0,43	0,61	3014
V_2018	46,27	0,29	0,49	0,67	2961
V_2019	45,6	0,27	0,46	0,63	2877
Vår:	***44,06***	***0,26***	***0,45***	***0,63***	***11936***
Samlet:	**42,64**	**0,25**	**0,47**	**0,65**	**12686**

Tabell 4.36

SAM3038	Psykologi 2				Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	41,8	0,25	0,43	0,59	256
H_2016	35,04	0,18	0,37	0,55	254
H_2017	36,98	0,19	0,39	0,55	338
H_2018	40,05	0,24	0,48	0,68	407
H_2019	36,36	0,2	0,42	0,6	363
Høst:	***38,05***	***0,21***	***0,42***	***0,59***	***1618***
V_2016	35,27	0,16	0,35	0,52	3303
V_2017	35,13	0,17	0,39	0,59	3786
V_2018	37,49	0,2	0,41	0,59	4262
V_2019	36,81	0,18	0,39	0,57	4789
Vår:	***36,18***	***0,18***	***0,39***	***0,57***	***16140***
Samlet:	**37,21**	**0,20**	**0,40**	**0,58**	**17758**

Tabell 4.37

SLG3102	Salgsfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	71,34	0,6	0,71	0,81	328
H_2017*	80,2	0,74	0,83	0,9	500
H_2018	73,63	0,65	0,76	0,85	364
H_2019*	77,48	0,7	0,8	0,88	515
Høst:	***75,66***	***0,67***	***0,77***	***0,86***	***1707***
V_2016	77,4	0,69	0,77	0,85	646
V_2017	70,46	0,61	0,74	0,85	474
V_2018	76,09	0,68	0,78	0,87	435
V_2019*	81,95	0,76	0,84	0,91	554
Vår:	***76,48***	***0,69***	***0,78***	***0,87***	***2109***
Samlet:	**76,07**	**0,68**	**0,78**	**0,87**	**3816**

Tabell 4.38

SPR3008	Internasjonal	engelsk,	skriftlig		Sentral-Sentral
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	37,62	0,19	0,41	0,61	412
H_2016	37,7	0,17	0,42	0,64	382
H_2017	43,21	0,27	0,49	0,67	324
H_2018	37,82	0,2	0,45	0,66	312
H_2019	38,98	0,21	0,41	0,61	313
Høst:	***39,07***	***0,21***	***0,44***	***0,64***	***1743***
V_2016	41,81	0,22	0,42	0,6	3542
V_2017	41,04	0,22	0,41	0,59	3475
V_2018	39,73	0,21	0,4	0,57	3295
V_2019	40,99	0,23	0,44	0,63	3035
Vår:	***40,89***	***0,22***	***0,42***	***0,60***	***13347***
Samlet:	**39,88**	**0,21**	**0,43**	**0,62**	**15090**

Tabell 4.39

TMF3102	Tømrerfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016	60,47	0,49	0,69	0,83	387
H_2017	63,47	0,51	0,65	0,77	334
H_2018	71,85	0,61	0,71	0,79	437
H_2019	63,26	0,51	0,7	0,84	430
Høst:	***64,76***	***0,53***	***0,69***	***0,81***	***1588***
V_2016	55,06	0,39	0,57	0,71	425
V_2017	55,12	0,42	0,6	0,74	488
V_2018	66,94	0,53	0,66	0,75	493
V_2019	74,04	0,66	0,79	0,89	547
Vår:	***62,79***	***0,50***	***0,66***	***0,77***	***1953***
Samlet:	**63,78**	**0,52**	**0,67**	**0,79**	**3541**

Tabell 4.40

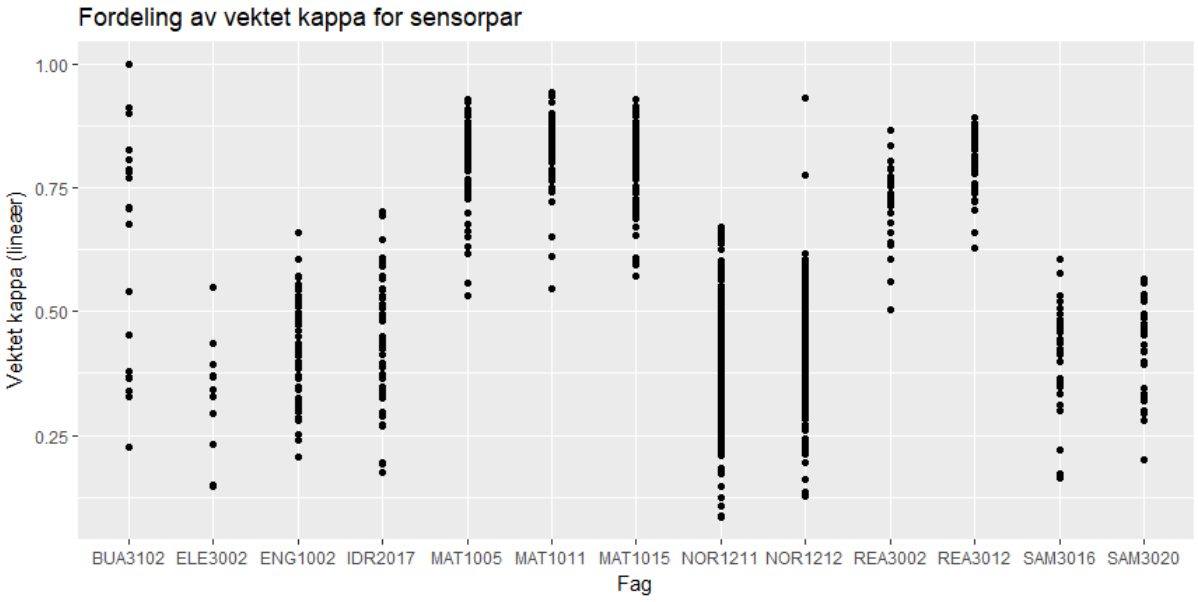
YRK3102	Yrkessjåførfaget, skriftlig				Sentral-Lokal
Periode	Enighet %	Kappa	Vektet kappa	ICC	Ant. kandidater
H_2015	-	-	-	-	0
H_2016*	99,11	0,99	0,99	1	338
H_2017	72,08	0,65	0,76	0,84	419
H_2018*	79,15	0,74	0,83	0,9	446
H_2019	50,57	0,36	0,45	0,54	439
Høst:	***75,23***	***0,69***	***0,76***	***0,82***	***1642***
V_2016	76,32	0,69	0,77	0,82	380
V_2017*	82,96	0,78	0,88	0,95	487
V_2018	72,38	0,65	0,77	0,87	467
V_2019	69,34	0,6	0,71	0,8	411
Vår:	***75,25***	***0,68***	***0,78***	***0,86***	***1745***
Samlet:	**75,24**	**0,68**	**0,77**	**0,84**	**3387**

Det er ganske klart når en ser på disse tabellene, at noen fag skiller seg ut. Norsk, språkfag og samfunnsfagene har generelt lav enighet mellom sensorer, og matematikk og andre realfag har generelt høyt sensorsamsvar og enighet. Det er også klart at noen av de mindre fagene har ganske store variasjoner fra år til år, og at dette i noen tilfeller er knyttet til at de har få kandidater, eller at de er både lokalt utviklet og lokalt sensurert. Dette kan muligens også knyttes til at disse gruppene er små og følsomme for endringer i gruppen, noe som kan føre til skjeve fordelinger av karakterene som igjen kan gjøre evalueringen av sensorreliabiliteten usikker. Det er derfor alltid viktig å ha i tankene hvor store grupper det er som ligger bak reliabilitetsberegningene og prøve å evaluere hvorvidt dette kan ha en effekt på resultatene.

Lokal sensurering ser ikke ut til å være svakere enn den sentrale. Dette er ikke helt lett å forklare. For å gjøre det på en god måte, vil det være nødvendig å gå inn i enkelte fag og se nærmere på hvordan sensorene organiserer sitt arbeid. Det vil også bli viktig i fortsettelsen å se på forskjellige eksamensformer og forskjellige oppgaveformater, ettersom disse aspektene muligens kan forklare noe av denne ganske store variasjonen. Flervalgsoppgaver med entydige svar er lettere å vurdere enn lange skrevne tekster. Realfagsoppgaver med tydelige rette og feil svar er også lettere å vurdere og fører sannsynligvis til høyere sensorreliabilitet. For å kunne forklare disse forskjellene i reliabilitet, er det derfor som allerede understreket, nødvendig å gå inn i hvert fag og se på hvordan kompetansene i fagene er testet, hva slags svar sensorene får å vurdere og hvordan oppgaveformatene varierer, i tillegg til å se nøye på hvilke veiledninger og instrukser sensorene får å arbeide ut ifra. Her er det også viktig å peke på at mange av fagene ligger på grensen til å være gode nok ut ifra kriteriene som er skissert ovenfor.

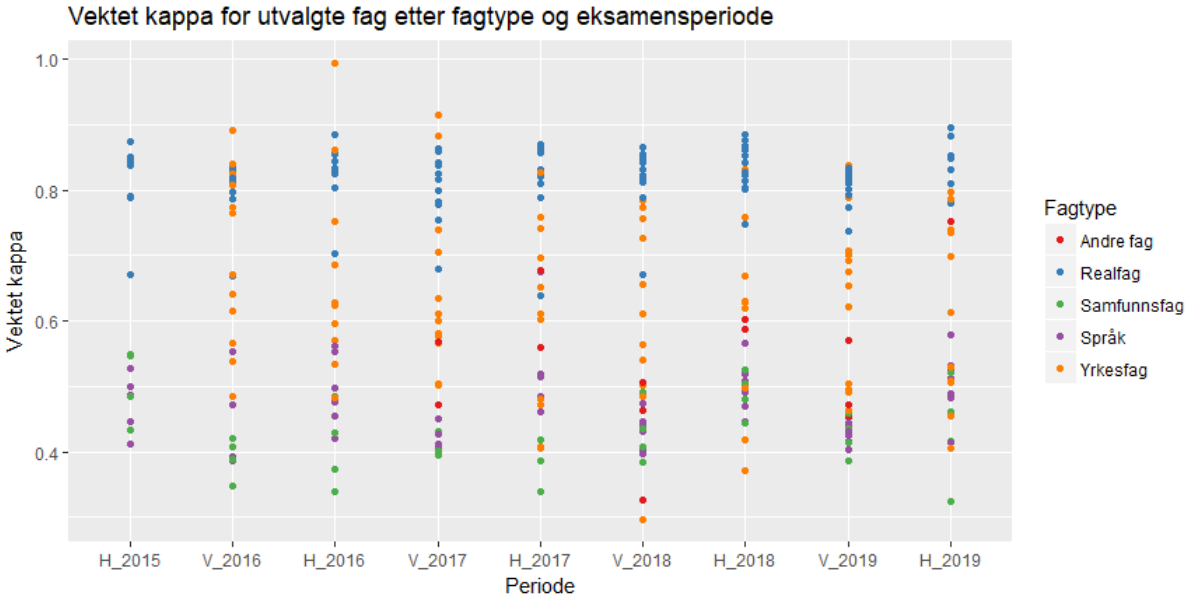
Grafisk fremstilling av noen resultater

Alle disse resultatene og tallene kan være litt uoversiktlige, og derfor har vi også inkludert en grafisk fremstilling som viser dem klassifisert og oppsummert for analyser gjort av sensorpar, overordnet type fag og høst- våreksamen.



Figur 1. Sensorpar

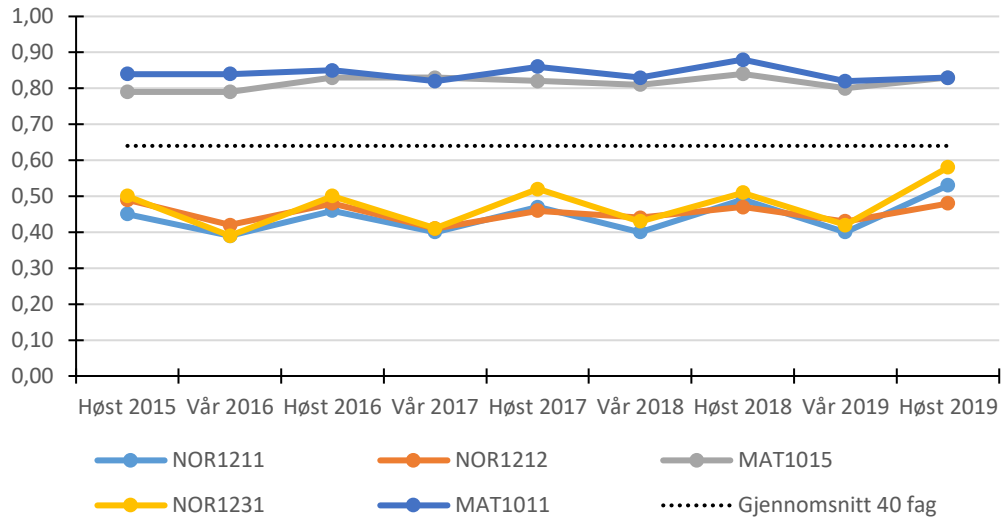
I figur 1 har vi valgt å vise bare noen fag for å illustrere variasjonen mellom dem. Her ser vi norsk og matematikk, men også noen mindre fag som har både stor og mindre variasjon. Vektet kappa er sannsynligvis den koeffisienten som gir mest mening. Den tar hensyn til at karakterforslag som er nærme hverandre, selv om sensorene ikke er helt enige, er bedre enn de som er lengre fra hverandre. For eksempel betyr en forskjell mellom 3 og 4 større enighet enn en forskjell mellom 2 og 4. Figur 1 viser at det er vesentlig større enighet i matematikk og realfag enn i språkfag og norsk, noe som kanskje ikke overrasker. Ikke bare er koeffisientene høyere, dvs. større enighet, men det er også mindre variasjon i karaktergivingen hos de ulike sensorparene. Figuren viser helt tydelig at norsk, språk og samfunnsfag har størst variasjon, der varierer koeffisienten vesentlig mer enn i de andre fagene. I disse fagene går kappa-koeffisienten nesten aldri over 0,75, og det er også mange tilfeller hvor den ligger meget lavt. Også noen av de små fagene har en stor variasjon, som for eksempel BUA3102-barne og ungdomsarbeiderfaget. Andre små fag har mindre variasjon, som for eksempel ELE3002-elektrikerfaget.



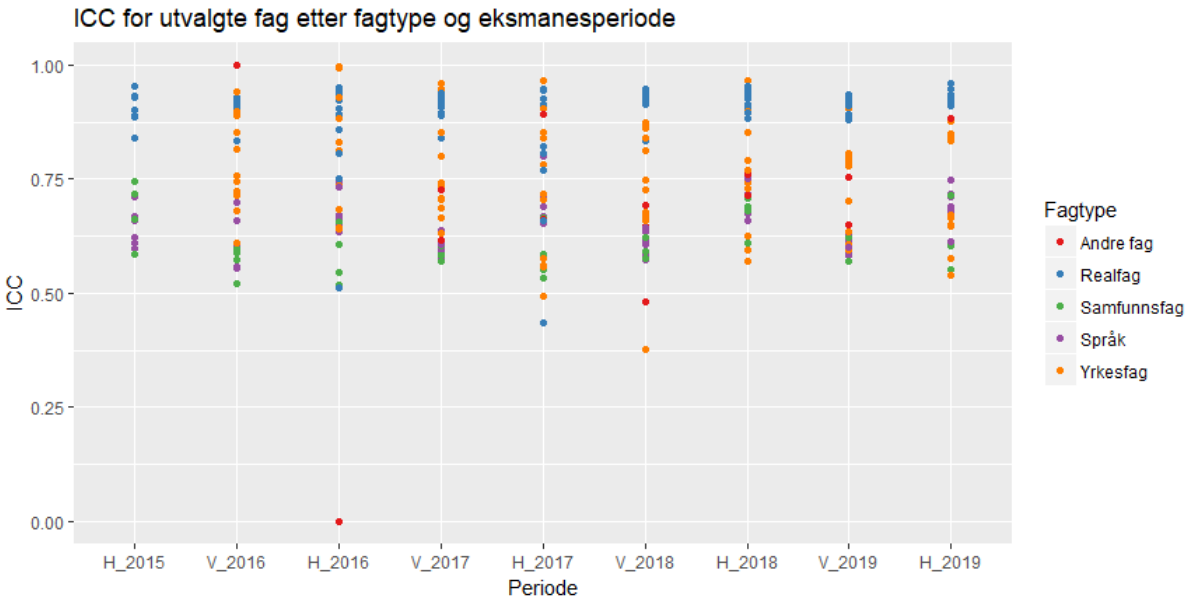
Figur 2. Vektet kappa

Figur 2 viser fagtypene høst og vår. Forskjellene mellom høst og vår ser ikke ut til å være påfallende. I tabellene ovenfor kan man likevel se at enigheten i noen fag endrer seg fra høst til vår, dvs. det er forskjeller, men det ser ikke ut til å være en generelt systematisk endring mellom høst og vår på tvers av fag. I noen fag ser enigheten ut til å være større om høsten enn om våren, og i andre fag er dette omvendt.

I figur 3 presenteres vektet kappa for de fem største fagene og et gjennomsnitt for alle fagene i perioden høsten 2015-våren 2019. Selv om vi ikke ser noen store systematiske forskjeller på tvers av alle fag, viser figur 3 en «bølgete» kurve for NOR1211 og NOR1231, og til en viss grad for NOR1212, mens MAT1011 og MAT1015 ser forholdsvis stabile ut. Bølgene for norsk-kodene indikerer lavere enighet på våren, noe vi også kan se i tabellene ovenfor. En mulig forklaring på dette er at antallet kandidater, og dermed også arbeidsbyrden for sensorene, er vesentlig større på våren enn på høsten.



Figur 3. Variasjoner i tid for de fem største fagene



Figur 4. ICC fra utvalgte fag

Når en evaluerer figur 4 og sammenlikner den med figur 2, så kan det se ut som samsvaret (ICC) er større enn enigheten (vektet kappa). Men her må vi huske at disse koeffisientene ikke måler det samme. En ICC-

koeffisient viser en korrelasjon som kan bli høy selv om absolutt enighet kan være lav, mens vektet kappa måler hvorvidt sensorene bruker de samme eller like karakterer for samme elever. Det er også viktig å huske på at karakterskalaen ikke betyr nøyaktig det samme i alle fag eller blir brukt eller kan brukes på samme måte på tvers av fagene.

Disse sammenfatningene viser uansett tydelige forskjeller mellom fag ut ifra innhold. Når det gjelder hvordan oppgavene er utviklet og sensurert, er ikke bildet like klart. Der finner vi variasjoner som ikke kan forklares med disse tallene. Det er derfor nødvendig å gå inn i hvilke metoder som er brukt og hvordan sensureringen er iverksatt, i tillegg til å se på hvilke oppgaveformater og vurderingskriterier som er brukt i hvert enkelt fag. Kun ved å se på alle disse aspektene samlet, kan vi få et mer fullstendig bilde av kvaliteten på ulike eksamener.

Hvor godt kan eksamen skille mellom elevbesvarelser av ulik kvalitet og sensorers strenghet?

I dette avsnittet presenteres kasusstudier fra de 40 fagkodene våren 2019. Disse studiene bygger på analyse innenfor det statistiske rammeverket «*Many-facet Rasch Measurement*» (MFRM).

Om «Many-facet Rasch Measurement» (MFRM)

Forenklet kan MFRM sies å bygge på følgende logikk: resultatet til en elev vil være avhengig av elevens kompetanse, oppgavens vanskegrad og sensors strenghet. Et konkret eksempel kan være at sensor X og sensor Y og sensor Z har fått i oppgave å vurdere flere elevbesvarelser, som er gitt som svar på de to oppgavene A og B. Noen få elever har bare svart på den ene oppgaven. De tre sensorene har fordelt arbeidet slik at de har et lite antall fellestekster som alle tre vurderer. Resten av tekstene vurderes sammen med én makker. I dette eksemplet skiller elevbesvarelsene seg i kvalitet, og sensorene skiller seg i strenghet. I tillegg er begge oppgavene av ulik vanskegrad.

Ved å bruke MFRM-analyser vil det være mulig å utnytte det faktum at elever, sensorer og oppgaver bindes sammen gjennom utvalgte fellestekster. Gjennom hundretalls, iblant tusentalls iterasjoner, vil MFRM-analysen estimere elevbesvarelsers kvalitet, sensorers strenghet og oppgavers vanskegrad uavhengig av hverandre og på samme skala (som i tillegg er en intervallskala). Dessuten er det mulig å bruke informasjonen for å justere for «urettferdigheter». I dette eksemplet var sensor X strengest, Y moderat og Z minst streng. Oppgave A var vanskeligere enn oppgave B. Elever som kun gjennomførte oppgave A og som ble vurdert av sensor X og Y vil gjennom MFRM-analysen kunne få en noe høyere skår, som tar høyde for at de sannsynligvis ville klart seg bedre om de også hadde gjennomført oppgave B og blitt vurdert av sensor Z.

En viktig funksjon ved MFRM-analyser er å estimere hvor presist en vurdering skiller mellom kvaliteten til elevbesvarelser, strengheten til sensorer og vanskegraden til oppgaver. Dette måles med en separasjonsindeks (strata). Hvis presisjonen for elevbesvarelser er lav, vil det ikke være mulig å si at to elevbesvarelser faktisk er av ulik kvalitet – forskjellen vil ikke være statistisk signifikant (for utdypende informasjon om MFRM, se Linacre, 2018, Rasch, 1980).

En målsetning med en slik MFRM-analyse som er anvendt her, er derfor å separere elevenes kompetanse, oppgavenes vanskegrad og sensorenes strenghet til tre størrelser, slik at vi får kunnskap om f.eks. elevers kompetanse uavhengig av nøyaktig hvilken sensor som har vurdert elevsvaret.

En forutsetning for å skape meningsfulle resultater i MFRM-analyser er at alle «fasetter», dvs. kandidater, sensorer og fagkoder, er lenket til hverandre. Lenkingen kan være relativt svak, men det må finnes en empirisk kobling mellom alle fasetter. Denne koblingen kan for eksempel oppstå hvis besvarelsene til kandidat A er blitt vurdert av Sensor 1 og Sensor 2, som i sin tur også har vurdert kandidat B. Kandidatene A og B og Sensor 1 og 2 er dermed koblet til hverandre. Kandidat C, som er blitt vurdert av sensor 3 og 4, er

derimot ikke koblet til kandidatene A og B, liksom sensorene 3 og 4 ikke er koblet til sensorer 1 og 2. Ett av trinnene i analysene har derfor vært å lage utvalg hvor denne koblingen er til stede. Dette har vi gjort ved å analysere data i FACETS for å så undersøke størrelsen på delutvalgene (såkalte subsets). Vi har så eksportert filer fra FACETS til Excel for å manuelt identifisere størrelsen på delutvalgene. Vi vil ikke presentere tekniske detaljer om disse tidskrevende pre-analysene, men nøyer oss med å konstatere at vi i hvert tilfelle har valgt å gå videre med det utvalg som har bestått av størst antall sensorer og så brukt disse i en ny analyse.

Det er flere resultater som er av interesse fra en MFRM-analyse. Vi vil her sette søkelyset på følgende:

- Den empiriske lenkingen mellom sensorer og elever som et bilde på det empiriske tolkningsfelleskapet våren 2019.
- Reliabiliteten i diskriminering av besvarelser av ulik kvalitet og hvorvidt sensorer er statistisk signifikant ulikt strenge i sin vurdering. Diskriminering av elevbesvarelser og sensorstrenghet kalles i MFRM-terminologi for *separasjon* og dreier seg enkelt sagt om vurderingen ser ut til å kunne separere besvarelser/strenghet i distinkte kategorier. Dette vil være avhengig av presisjon i vurderingen og antall vurderinger (se også Huebner & Skar, 2021). Hvis vurderingene er konsistente, vil data passe MFRM-modellen godt og separasjonsindeksen vil øke. Få observasjoner av en enkeltelev i tillegg til et lite utvalg vil gjøre estimatene fra MFRM-analysen usikre og separasjonsindeksen vil bli lav.

Analyse av tallene

Vi vil presentere følgende mål:

- deskriptiv statistikk, inklusive antall i hvert utvalg,
- *strata*, som uttrykker antall statistisk distinkte klasser av prestasjon som er mulig å utskille, og
- reliabiliteten *R*, som er en analogi til *Cronbach's alpha,* og som går fra 0–1.

Det er viktig å huske at R-verdien gjenspeiler et tenkt scenario hvor elevers resultat er summen av to karakterforslag og ikke karakteren som blir satt etter fellessensur. For sensorer vil vi også presentere prosent samsvar. *Det er viktig å understreke at vi ønsker høye strata- og R-verdier for elevbesvarelser og lave for sensorer.* Det er også viktig å presisere at strata for elever kan være lavt, selv om samstemmigheten mellom sensorer er høy. Dette kan eksempelvis skyldes små forskjeller i elevprestasjoner og at antall observasjoner av elevprestasjoner er få. Høye strata og R-verdier for sensorer innebærer at vi med høy sikkerhet kan si at de er ulikt strenge. Resultatene presenteres i tabellene 5 og 6 og kommenteres nedenfor.

Tabell 5. Deskriptiv statistikk

	Subsets	Kand.			Sensorer		
	N	*N*	*M*	*SD*	*N*	*M*	*SD*
Alle fag	860	92 095	3.19	1.10	2 389	3.28	0.51
AMF3102	6	252	3.03	1.02	12	3.02	0.30
AUT4002	4	502	3.05	1.06	9	3.07	0.31
BUA3102	26	1 648	2.92	1.15	53	2.91	0.38
ELE3002	19	1 644	3.01	1.04	39	3.00	0.46
ENG1002	19	4 547	3.88	1.07	80	3.88	0.32
HEA3102	33	2 048	2.24	1.11	67	2.25	0.45
HSF1001	10	220	3.04	1.23	20	3.07	0.52
HSF1003	10	220	3.06	1.19	20	3.05	0.55
IDR2016	4	1 836	3.59	1.14	108	3.68	0.55
IDR2017	19	3 697	3.32	1.05	76	3.32	0.34
LOG3102	6	496	3.08	1.15	12	3.06	0.62
MAT1001	10	1 042	3.17	1.47	60	3.29	0.70

MAT1001-0001	4	267	3.35	1.38	29	3.29	0.84
MAT1001-0003	5	308	3.99	1.26	31	4.16	0.68
MAT1001-0004	7	645	3.28	1.35	56	3.30	0.77
MAT1001-0008	5	385	3.18	1.23	38	3.24	0.38
MAT1005	38	6 463	2.50	1.23	163	2.47	0.29
MAT1011	15	4 123	2.89	1.27	55	2.88	0.18
MAT1015	38	7 377	2.79	1.24	163	2.80	0.28
MUS2007	9	309	4.09	1.07	32	4.15	0.44
NOR1206	33	2345	3.43	0.96	236	3.48	0.50
NOR1211	131	38 125	3.51	0.97	520	3.51	0.27
NOR1212	65	19 323	3.25	0.99	261	3.25	0.29
NOR1231	135	10 175	2.90	0.91	479	2.87	0.41
NOR1232	70	4 728	2.68	0.92	258	2.65	0.38
REA3002	9	2 843	3.58	1.29	35	3.58	0.21
REA3012	11	3 961	3.46	1.42	43	3.46	0.24
REA3022	14	4 099	3.29	1.30	52	3.29	0.21
REA3024	17	5 356	3.32	1.38	64	3.31	0.23
REA3026	10	3 176	3.20	1.22	38	3.19	0.16
REA3028	13	4 196	3.18	1.23	48	3.18	0.17
RHO3102	6	579	2.49	1.19	12	2.50	0.46
SAM3016	10	3 322	3.41	1.07	37	3.41	0.29
SAM3020	7	2 548	3.50	1.11	29	3.49	0.30
SAM3023	8	2 877	3.55	0.99	31	3.55	0.32
SAM3038	13	4 789	3.53	1.08	51	3.52	0.31
SLG3102	5	554	3.09	1.07	13	3.09	0.24
SPR3008	10	3 036	3.37	1.06	40	3.37	0.19
TMF3102	6	548	2.40	1.17	12	2.39	0.33
YRK3102	4	411	3.01	1.10	8	3.01	0.49

NB. Kand. = kandidater. M = gjennomsnitt, SD = standardavvik.

Samlet var de 92 095 elevene og 2 389 sensorene oppdelt i 890 delutvalg. I enkelte fag var elevene samlet i få, større utvalg (som IDR2016), mens i andre fag var utvalgene relativt sett flere og mindre (som HSF1001). Konsekvensen av at kandidater og sensorer inngikk i distinkte delutvalg, er at resultatene mellom elever og sensorer ikke per se er sammenlignbare. Manglene på empirisk lenking innebærer at vi ikke kan vite noe om en enkeltsensor i relasjon til hele korpset, kun hans eller hennes kollegaer i det aktuelle delutvalget. Dermed kan vi ikke heller, strengt tatt, sammenligne elever mellom utvalg; vi vet ikke om elevene i et utvalg har fått en vurdering som er på linje med andre elever, eller om de havnet i et «snilt» eller «strengt» korps.

Ser vi til R og strata for elevene (tabell 6), kan vi notere at fem delutvalg har en R-verdi på 0,90 eller høyere. Det betyr at i disse fagene ville reliabiliteten våren 2019 vært på et tilstrekkelig høyt nivå hvis en tok i bruk karakterforslagene (se **fet stil** i tabell 6). Samtidig ser vi at ikke noen fag oppviste en strata-verdi på mer enn 5,0. Enkelt uttrykt betyr det en mangel på underlag for å separere elever i seks kompetansenivåer. *Dette skyldes med høy sannsynlighet at elevene er «observert», som dette gjerne kalles, svært få ganger – vanligvis kun to ganger per besvarelse.*

Hvis vi studerer R og strata for sensorene, hvor det er ønskelig med lav separasjon (jfr. beskrivelse ovenfor), blir bildet delvis det motsatte. For 15 fagkoder er reliabiliteten mindre enn 0,90, men for resterende 25 fagkoder er det mulig å med høy presisjon separere sensorene i distinkte klasser av strenghet. For eksempel plasserer sensorene (*n* = 15) i NOR1212-utvalget seg i 12 distinkte strata, mens forskjellen mellom sensorene i MAT1001-utvlaget (*n* = 21) er ubetydelig. I figur 5 visualiseres strata for kandidater og sensorer, og grafen gjengir det som tallene i tabell 6 viser, nemlig at eksamen i disse 40 fagene, samlet sett og

spissformulert, er bedre på å skille mellom strengheten til sensorer enn kompetansen til elever. I figur 6 visualiseres reliabiliteten på tvers av fag.

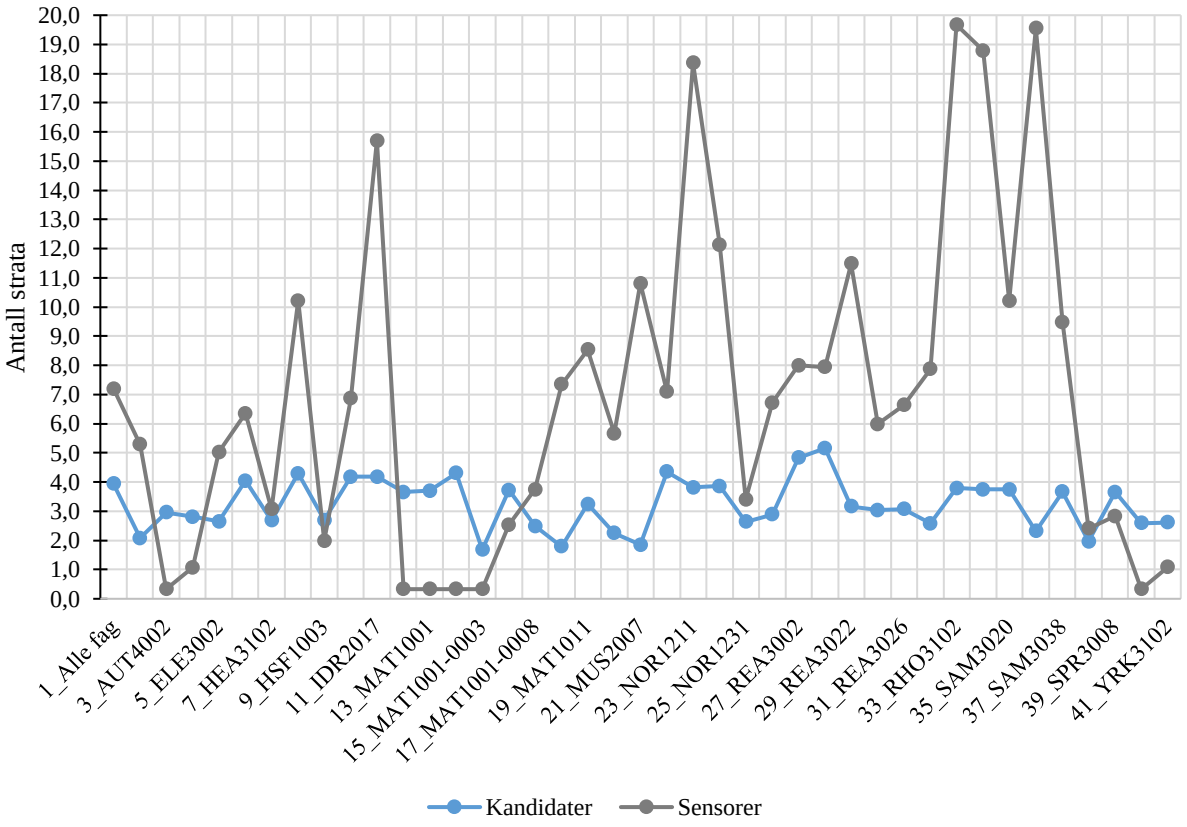
Det er viktig å poengtere at store sprik mellom sensorer i seg selv ikke er negativt, så lenge en kan kontrollere for dem og så lenge det er mulig å separere elever med høy presisjon. Gitt hvordan eksamen vurderes per nå, så finnes det imidlertid få omstendigheter som tilsier at det er tilfelle. MFRM-analysen viser at det er høy sannsynlighet for at enkeltelever er prisgitt sensors vurderingsatferd. Videre kan resultatene fra analysen tyde på at informasjonen fra sensuren ikke er presis nok til å bruke en skala med seks trinn for å separere ulike nivåer av den viste fagkompetansen.

Tabell 6. Reliabilitet, MFRM-analyse

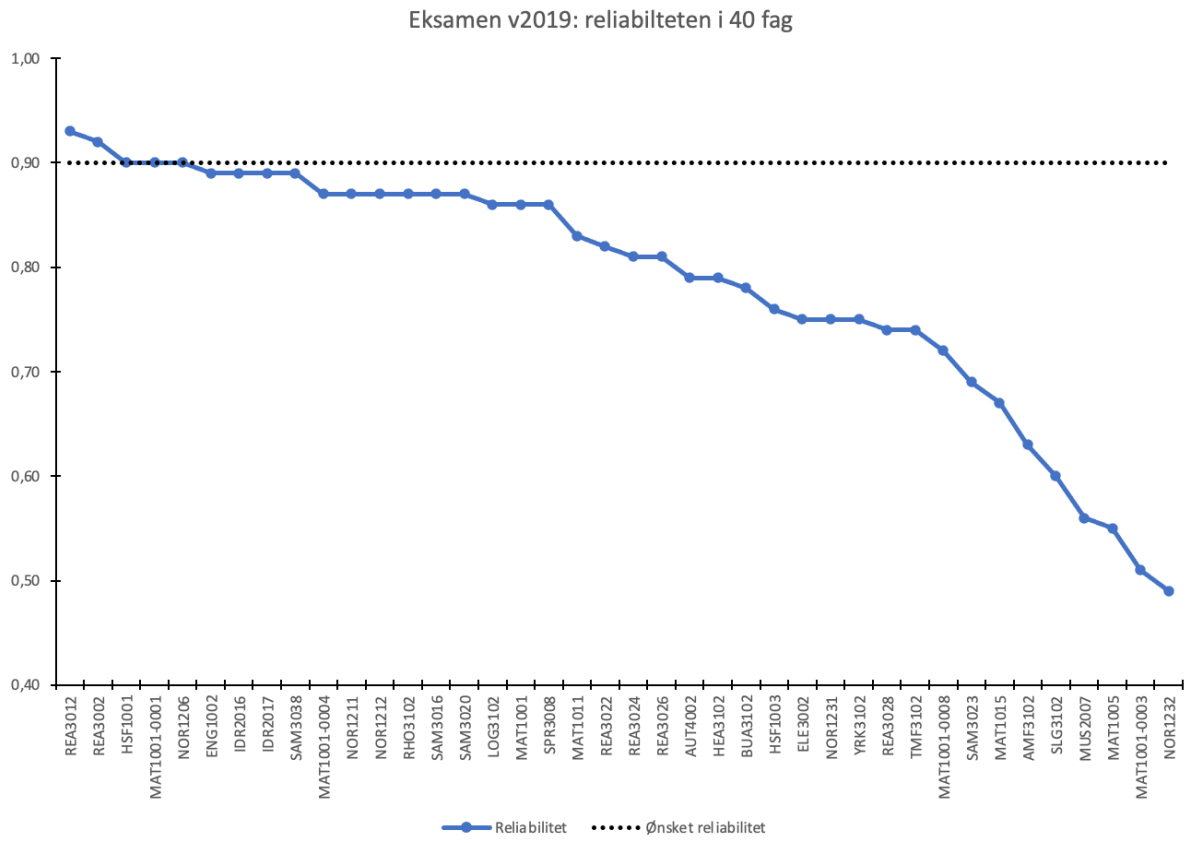
	Kand.					Sensor					
	N	*M*	*SD*	*R*	Strata	*N*	*M*	*SD*	*R*	Strata	*%*
Alle fag[a]	87 401	3.21	1.09	0.88	3.94	2 196	3.30	0.48	0.96	7.20	51.9
AMF3102	34	3.10	0.88	0.63	2.08	2	3.04	0.22	0.93	5.29	61.8
AUT4002	157	3.39	1.10	0.79	2.96	2	3.44	0.08	0.00	0.33	38.6
BUA3102	99	2.72	0.94	0.78	2.81	3	2.73	0.09	0.24	1.07	67.7
ELE3002	162	3.18	1.00	0.75	2.65	3	3.21	0.28	0.93	5.03	75.3
ENG1002	689	4.04	1.00	0.89	4.04	12	4.08	0.33	0.95	6.35	38.2
HEA3102	125	2.05	0.94	0.79	2.69	3	2.06	0.18	0.81	3.07	48.0
HSF1001	23	3.04	1.32	**0.90**	4.29	2	3.04	0.80	0.98	10.22	17.4
HSF1003	17	2.24	1.02	0.76	2.70	2	2.81	0.07	0.61	1.99	88.2
IDR2016	1 121	3.64	1.12	0.89	4.18	63	3.71	0.56	0.96	6.87	49.6
IDR2017	206	3.50	1.01	0.89	4.18	4	3.50	0.51	0.99	15.70	35.0
LOG3102	79	2.82	1.18	0.86	3.66	2	2.82	0.03	0.00	0.33	63.3
MAT1001	234	3.03	1.39	0.86	3.70	21	3.24	0.94	0.00	0.33	75.5
MAT1001-0001	120	3.55	1.33	**0.90**	4.31	12	3.32	1.01	0.00	0.33	54.2
MAT1001-0003	54	4.36	1.09	0.51	1.68	7	4.24	0.64	0.00	0.33	66.7
MAT1001-0004	134	3.40	1.45	0.87	3.72	13	3.38	0.94	0.73	2.54	71.6
MAT1001-0008	96	3.09	1.19	0.72	2.49	9	3.10	0.29	0.87	3.75	64.6
MAT1005	652	2.27	1.16	0.55	1.79	15	2.27	0.13	0.97	7.35	76.7
MAT1011	321	3.02	1.26	0.83	3.24	4	3.02	0.14	0.97	8.53	68.5
MAT1015	678	2.81	1.29	0.67	2.25	15	2.76	0.20	0.94	5.67	78.9
MUS2007	54	3.64	1.11	0.56	1.85	6	3.69	0.17	0.98	10.81	28.3
NOR1206	120	3.53	0.95	**0.90**	4.35	11	3.68	0.31	0.95	7.09	45.0
NOR1211	352	3.48	0.92	0.87	3.81	4	3.47	0.36	0.99	18.36	35.5
NOR1212	1 295	3.18	0.98	0.87	3.86	15	3.18	0.30	0.99	12.14	39.5
NOR1231	47	3.03	0.73	0.75	2.64	4	2.98	0.21	0.84	3.40	38.3
NOR1232	59	2.65	0.85	0.49	2.89	4	2.77	0.52	0.96	6.72	33.9
REA3002	322	3.48	1.33	**0.92**	4.83	4	3.47	0.22	0.97	8.00	63.4
REA3012	378	3.42	1.34	**0.93**	5.16	4	3.43	0.24	0.97	7.94	70.6
REA3022	303	3.49	1.24	0.82	3.16	4	3.49	0.20	0.99	11.48	67.3
REA3024	320	3.14	1.38	0.81	3.04	4	3.14	0.22	0.95	5.99	74.7
REA3026	333	3.24	1.19	0.81	3.07	4	3.23	0.14	0.96	6.64	74.5
REA3028	353	3.33	1.20	0.74	2.58	4	3.33	0.10	0.97	7.87	79.6
RHO3102	97	3.08	1.16	0.87	3.79	2	3.08	0.82	0.99	19.67	22.7
SAM3016	369	3.23	1.05	0.87	3.75	4	3.24	0.54	0.99	18.79	31.4
SAM3020	367	3.59	1.15	0.87	3.75	5	3.55	0.45	0.98	10.21	32.4
SAM3023	366	3.81	0.96	0.69	2.32	4	3.81	0.42	0.99	19.55	40.7
SAM3038	358	3.20	1.06	0.89	3.67	4	3.20	0.19	0.98	9.48	36.9

SLG3102	225	3.05	1.14	0.60	1.96	5	3.10	0.34	0.71	2.41	85.3
SPR3008	307	3.39	0.95	0.86	3.65	4	3.39	0.05	0.78	2.83	45.3
TMF3102	86	2.56	1.01	0.74	2.59	2	2.56	0.01	0.00	0.33	47.7
YRK3102	101	2.97	1.17	0.75	2.62	2	2.97	0.03	0.24	1.08	80.2

NB. Kand. = kandidater. M = gjennomsnitt, SD = standardavvik, R = Rasch-reliabilitet, % = prosentuell fullstendig samstemmighet. [a]For «Alle fag» er reliabiliteten estimert uten at det er tatt hensyn til forskjeller i vanskegrad mellom fagene.



Figur 5. Strata for hhv. kandidater og sensorer



Figur 6. Reliabilitet fra MFRM analysen for alle 40 fagene

Sammenfatning

Formålet med denne rapporten var å presentere en undersøkelse av sensorreliabilitet ut ifra de foreløpige karakterene for et utvalg på 40 fag årene 2015–2019. I analysen bruktes flere mål på sensorreliabilitet for å sikre en best mulig forståelse av reliabiliteten på eksamen. Analysen baserer seg på karakterforslagene fra to uavhengige sensorer. Dette er det beste estimatet vi kan få på sensorreliabilitet siden det per i dag ikke er mulig å gjøre denne typen analyser på endelige eksamenskarakterer.

Sammenfattet kan vi si at påliteligheten i den eksterne sensureringen av eksamen, slik den kommer til uttrykk i de foreløpige karakterene, er svært varierende. Mens sensorreliabiliteten i for eksempel matematikk generelt var meget god, var den svakere i de andre fagene, spesielt i norsk og andre språkfag, samt i samfunnsfagene. Analysene viser at vi ikke kan utelukke at eksamensresultatet kunne vært et helt annet hvis eleven var blitt sensurert av et annet sensorpar enn hva som var tilfellet, i alle fall i noen fag. Her må vi også huske at disse analysene er gjort med utgangspunkt i karakterforslag fra sensor 1 og sensor 2, og at den endelige karakteren ikke bestemmes direkte fra deres forslag, men som resultatet av en «forhandling» mellom dem. Dette siste er spesielt viktig i fag som baseres på skriving av lengre tekster hvor også sensorenes skjønn og vurdering spiler en stor rolle, i motsetning til eksamener hvor man bygger på enklere flervalgsoppgaver med forhåndsgitte riktige og mer konkrete svar.

Videre viste analysen at sensorreliabiliteten i norsk, språk- og samfunnsfagene på den ene siden og i matematikk og realfagene på den andre var henholdsvis stabilt lav og høy over tid, mens andre fag viste noen svingninger gjennom disse årene. En skal dog være oppmerksom på at det ikke er sikkert at disse kan sammenliknes direkte, ettersom fagene er meget forskjellig bygget opp, med ulike typer eksamensoppgaver og sannsynligvis store forskjeller i sensureringsmetoder, sensorveiledninger og

sensorskolering. Det er også godt mulig at forskjellig innhold, dvs. fagene i seg selv, legger noen begrensninger som fører til forskjellig sensorreliabilitet. Dette må analyseres videre.

Kasusstudiene som ble gjennomført som MFRM-analyser, viste blant annet at det mulige tolkningsfellesskapet var tilsynelatende lite i mange fag. Det var få overlapp mellom sensorer, noe som gjør det vanskelig å sammenligne kandidater og sensorer på en god måte. Analysen viste også at eksamen generelt sett var bedre på å skille mellom sensorers strenghet enn kandidaters kompetanse. Videre kunne vi i analysen av delutvalgene notere at det ikke fantes statistisk grunnlag for å skille mellom seks nivåer av kompetanse. I gjennomsnitt klarte eksamen å utskille tre nivåer av kompetanse presist nok, ifølge MFRM-analysen. Dette er spesielt interessant i lys av det at mange sensorer har beskrevet at det kan være vanskelig å skille mellom f.eks. 3 og 4, mens det er lettere å bestemme både meget svake og meget gode kandidater, dvs. ytterkantene av skalaen.

Et annet interessant funn var følgende: Hvis vi utgår fra karakterforslag, ville sensorreliabiliteten i flere fag vært over 0,80. Som figur 6 viser, ville kun et fåtall fag da ha en svært lav sensorreliabilitet (dvs. under 0,7).

Konklusjon og veien videre

En generell konklusjon på alle disse analysene er at det finnes store variasjoner i sensorreliabilitet på norske eksamener når man analyserer dette ut ifra de foreløpige karakterene. I noen fag er denne reliabiliteten så lav at vi ikke kan utelukke at eksamenskarakteren ikke bare gjenspeiler den kompetansen kandidatene har, men også vel så mye hvilke sensorer som har vurdert besvarelsen. Dette må utforskes nærmere, slik at passende tiltak kan iverksettes for fag med lav sensorreliabilitet. Bruken av karakterskalaen fra 1 til 6 må også granskes nærmere ettersom MFRM-analysen kan tyde på at den ikke brukes fullt ut i alle fag. De analysene som ble presentert her må derfor brukes som et grunnlag for videre granskning av både egenskapene til eksamenene i sin helhet, en granskning av oppgaver og oppgaveformater, samt analyse av samsvaret mellom læreplaner, undervisning og oppgaver på eksamen. Videre må det gjøres en bedre og grundigere analyse av sensorenes arbeid. Alt dette kan deretter brukes som grunnlag for forbedret sensurering, og dermed mer rettferdige eksamener som leverer konsistente resultater, og som på en god og valid måte reflekterer elevenes kompetanse.

Vi mener derfor at en kommende analyse bør inneholde flere forklaringsvariabler (f.eks. type sensorskolering, type vurderingskriterier, type oppgaver, antall oppgaver m.m.) for å kunne utarbeide og gjennomføre forbedringstiltak der det trenges. Det er ikke sikkert at de samme tiltakene må iverksettes i alle fagene som nå har lav sensorreliabilitet. Årsaken til den lave sensorreliabiliteten kan være ulik i forskjellige fag, og videre analyser av hvert enkelt fag er derfor nødvendige. Vi mener også at for å kunne gjennomføre systematiske undersøkelser av sensoratferd og koble disse undersøkelsene mot aspekter som eksempelvis sensorskolering, må det faktiske tolkningsfellesskapet utvides betraktelig. En måte å gjøre det på er å la et lite antall besvarelser være felles for alle sensorer i et fag.

Hovedkonklusjonen fra disse analysene er at sensorreliabiliteten på eksamen i videregående opplæring varierer ganske mye, men antakeligvis av ulike årsaker i forskjellige fag. Vi mangler ennå veldig mange biter i dette puslespillet, og helhetsbildet blir ikke synlig før alle bitene er satt sammen på riktig plass.

Referanser

Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. Psychological Reports, 19, 3-11.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, 70, 213-220.

Fleiss, J.L., Cohen, J., & Everitt, B.S. (1969). Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72, 323-327.

Gamer, M., Lemon, J. & Fellows, I. (2019). Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Huebner, A. & Skar, G. B. (2021). Conditional Standard Error of Measurement: Classical Test Theory, Generalizability Theory and Many-Facet Rasch Measurement with Applications to Writing Assessment. [Submitted to journal.] Department of Applied and Computational Mathematics and Statistics, University of Notre Dame & Department of Teacher Education, NTNU.

Kendall, M.G. (1948). Rank correlation methods. London: Griffin.

Landis, J. R. and G. G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." Biometrics 33(1): 159-174.

LeBreton, J. M. and J. L. Senter (2007). "Answers to 20 Questions About Interrater Reliability and Interrater Agreement." Organizational research methods 11(4): 815-852.

Linacre, J. M. (2018). *A user's guide to FACETS. Rasch-model computer programs. Program manual 3.80.4*. Winsteps.com.

Linacre, J. M. (2020) *Facets computer program for many-facet Rasch measurement, version 3.83.4*. Beaverton, Oregon: Winsteps.com

McGraw, K.O., & Wong, S.P. (1996), Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1, 30-46.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. The University of Chicago Press.

Shrout, P.E., & Fleiss, J.L. (1979), Intraclass correlation: uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

StataCorp. (2017). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.

IBM (2019). *IBM SPSS Statistics for Windows, Version 26.0*. IBM Corp.

H.E.A. Tinsley, D.J. Weiss, Interrater reliability and agreement, in Handbook of Applied Multivariate Statistics and Mathematical Modelling, ed. by H.E.A. Tinsley, S.D. Brown (Academic Press, San Diego, 2000), pp. 95–124.

Package IRR i R:

Version: 0.84.1
Date: 2012-01-22
Title: Various Coefficients of Interrater Reliability and Agreement
Author: Matthias Gamer <m.gamer@uke.uni-hamburg.de>, Jim Lemon <jim@bitwrit.com.au>, Ian Fellows <ifellows@uscd.edu> Puspendra Singh <puspendra.pusp22@gmail.com>
Maintainer: Matthias Gamer <m.gamer@uke.uni-hamburg.de>
Depends: R (>= 2.10), lpSolve
Description: Coefficients of Interrater Reliability and Agreement for quantitative, ordinal and nominal data: ICC, Finn-Coefficient, Robinson's A, Kendall's W, Cohen's Kappa, ...
License: GPL (>= 2)
URL: https://www.r-project.org
Packaged: 2019-01-26 16:15:29 UTC; hornik
Repository: CRAN
Date/Publication: 2019-01-26 17:07:15 UTC
NeedsCompilation: no
Built: R 3.6.2; ; 2020-01-28 05:57:22 UTC; windows

Vedlegg 1. Antall elever i alle fag delt på år og årstid

		ÅR									Total
		H2015	H2016	H2017	H2018	H2019	V2016	V2017	V2018	V2019	
Fagkode	Anleggsmaskinførerfaget, skriftlig	0	295	282	242	246	299	346	299	252	2261
	Tverrfaglig eksamen, automatiseringsfaget"	0	72	50	51	69	495	541	527	502	2307
	Barne- og ungdomsarbeiderfaget, skriftlig	0	1079	1244	1111	883	2081	1955	2049	1648	12050
	Elektrikerfaget, skriftlig	0	615	646	538	577	1404	1535	1485	1644	8444
	Engelsk, Vg1 studieforberedende utdanningsprogram	1076	1201	1278	1281	1304	4186	4774	4746	4547	24393
	Helsearbeiderfaget, skriftlig	0	1236	1472	1270	1389	1782	2006	1878	2048	13081
	Helsefremmende arbeid	0	317	348	328	320	196	226	221	220	2176
	Yrkesutøvelse	0	296	339	307	296	171	210	216	220	2055
	Treningslære 1	0	0	79	104	117	0	1302	1488	1836	4926
	Treningslære 2	0	0	0	95	113	0	0	3547	3697	7452
	Logistikkfaget, skriftlig	0	520	414	428	389	515	512	472	496	3746
	Matematikk 1P-Y	0	591	653	753	836	1492	789	1347	1042	7503
	Bygg- og anleggsteknikk	0	20	23	60	39	0	160	243	267	812
	Elektrofag	0	9	12	13	16	0	221	320	308	899
	Helse- og oppvekstfag	0	41	118	102	99	0	321	444	645	1770

	Teknikk og industriell produksjon	0	41	54	71	46	0	199	404	385	1200
	Matematikk 2P-Y	1686	1623	1732	1784	1851	6480	6993	6803	6463	35415
	Matematikk 1P	3800	4054	4432	4350	4671	4112	4226	3705	4123	37473
	Matematikk 2P	1996	1816	1978	1893	2111	6842	6804	7396	7377	38213
	Musikk i perspektiv 2	0	5	9	6	0	7	86	269	309	691
	Norsk, Vg2 yrkesfaglige utdanningsprogram	0	220	236	229	204	1457	1944	2170	2345	8805
	Norsk hovedmål, Vg3 studieforberedende utdanningsprogram, skriftlig	2262	2199	1981	1919	1988	34874	35389	35746	38126	154484
	Norsk sidemål, Vg3 studieforberedende utdanningsprogram, skriftlig	1462	1439	1436	1401	1524	17379	17502	18151	19233	79527
	Norsk hovedmål, Vg3 påbygging til generell studiekompetanse, skriftlig	1054	1017	902	859	725	12419	12245	12255	10175	51651
	Norsk sidemål, Vg3 påbygging til generell studiekompetanse, skriftlig	621	635	548	621	560	5252	5347	5430	4728	23742
	Biologi 2	279	297	306	256	261	2917	2980	2900	2843	13039
	Kjemi 2	412	405	501	373	380	3857	3887	4052	3961	17828
	Matematikk R1	1823	1526	1486	1702	1636	4238	4260	4223	4099	24993
	Matematikk R2	634	650	608	631	655	5168	5001	5103	5356	23806
	Matematikk S1	1156	1086	1088	1158	1187	2903	2989	2987	3176	17730
	Matematikk S2	556	439	388	468	395	3551	4102	3929	4196	18024

	Renholdsoperatørfaget, skriftlig	0	435	510	404	518	520	652	479	579	4097
	Sosialkunnskap	409	411	344	354	337	2999	3002	3035	3322	14213
	Politikk og menneskerettigheter	177	182	128	136	156	2654	2624	2746	2548	11351
	Rettslære 2	165	192	151	126	116	3084	3014	2961	2877	12686
	Psykologi 2	256	254	338	407	363	3304	3786	4262	4789	17759
	Salgsfaget, skriftlig	0	328	500	364	515	646	474	435	554	3816
	Internasjonal engelsk, skriftlig	412	382	324	312	313	3542	3477	3295	3036	15093
	Tømrerfaget, skriftlig	0	387	334	437	430	425	488	493	548	3542
	Yrkessjåførfaget, skriftlig	0	338	419	446	439	380	487	467	411	3387
Total		20236	26653	27691	27390	28074	141631	146856	152978	154931	726440