

6 / 2025

# RAPPORT

2025

**Hvordan har det gått? Analyse av  
eksamenskvalitet og studentenes  
resultater på nasjonal eksamen i  
rettsanvendelse 2022–2024.**



NOKUT – Nasjonalt organ for kvalitet i utdanningen – er et statlig forvaltningsorgan under Kunnskapsdepartementet.



NOKUT har eget styre og er faglig uavhengig i oppgavene som er definert i universitets- og høyskoleloven. I tillegg utfører vi forvaltningsoppgaver som er delegert fra Kunnskapsdepartementet.



Formålet med NOKUTs virksomhet er å føre tilsyn med kvaliteten i høyere utdanning og høyere yrkesfaglig utdanning og å stimulere til kvalitetsutvikling som sikrer et høyt internasjonalt nivå i utdanningstilbudene ved institusjonene.



NOKUT skal bidra til at samfunnet har tillit til kvaliteten i norsk høyere utdanning og høyere yrkesfaglig utdanning. Gjennom arbeidet vårt skal NOKUT søke å bistå institusjonene i deres kvalitetsarbeid.



NOKUT bruker sakkyndige i akkrediteringer, tilsyn, evalueringer og prosjekter.

Du kan lese mer om arbeidet vårt på [nokut.no](https://nokut.no).



NOKUT er sertifisert  
som miljøfyrtårn

|                      |                                                                                                                              |
|----------------------|------------------------------------------------------------------------------------------------------------------------------|
| <b>Tittel</b>        | Hvordan har det gått? Analyse av eksamenskvalitet og studentenes resultater på nasjonal eksamen i rettsanvendelse 2022–2024. |
| <b>Forfatter(e)</b>  | Helen Bråten, Nora Torgersen, Maria Karlsen og Mathias Meier Nilsen, Karl Johan Skeidsvoll og Øystein Guttersrud             |
| <b>Dato</b>          | 15.11.2025                                                                                                                   |
| <b>Rapportnummer</b> | 6-2025                                                                                                                       |
| <b>ISSN-nr</b>       | 1892-1604                                                                                                                    |

© NOKUT Oppgi NOKUT som opphav ved bruk av materiale.

## Forord

Nye masterutdanninger i barnevernsfaglig arbeid og en nasjonal eksamen i rettsanvendelse i barnevernet ble innført i 2022 for å heve kompetansen og særlig den juridiske kompetansen i barnevernet.

Med denne rapporten vil vi dokumentere implementeringen av nasjonal eksamen i juridisk metode og rettsanvendelse, hvilke valg som ble tatt i begynnelsen samt analysere studentenes resultater og eksamenskvalitet herunder sensorenes vurderinger av eksamen. Målet er å frembringe og analysere kunnskap fra de første gjennomføringene av nasjonaleksamen i rettsanvendelse i barnevernet de første gjennomføringene, høsten 2022–høsten 2024.

Denne rapporten er en del av en større gjennomgang som NOKUT gjør av implementering og gjennomføring av de fire første nasjonale eksamenene i rettsanvendelse i barnevern.

Gjennomgangen i sin helhet tar for seg studenters, underviseres, sensorers og programlederes erfaringer med nasjonal eksamen i rettsanvendelse 2022 til 2024 samt dyptgående analyser av eksamenssett, resultat- og sensurdata for å undersøke ulike aspekter av vurderingskvaliteten. Hensikten er å kartlegge og utvikle kvalitet i de nasjonale eksamenene og å videreutvikle kunnskapen i utdanningene når det gjelder vurdering og læring generelt og av juridisk metode i barnevernet spesielt.

I tillegg til denne rapporten består gjennomgangen av to andre rapporter:

- «Undervisning i juridisk metode i barnevernsfaglige mastere: Innsikt og erfaringer fra de første nasjonale eksamenene» (Pedersen & Bråten, 2025)
- «Studentstemmen: Om undervisning, emneinnhold og vurdering i rettsanvendelse for barnevernsfaglige masterprogram» (Karlsen & Pedersen, 2025)

Tusen takk til fagmiljøene som på ulikt vis bidrar til nasjonal eksamen og har bidratt til at vi kan skrive disse rapportene. Takk for innspill til rapportene både i samtaler og på fagseminaret juni 2025!

## Innholdsfortegnelse

|                                                                                              |           |
|----------------------------------------------------------------------------------------------|-----------|
| <b>1 Introduksjon .....</b>                                                                  | <b>9</b>  |
| 1.1 Masterutdanning med nasjonal eksamen: svar på kompetansebehov .....                      | 10        |
| <b>Del 1 Implementeringen</b>                                                                |           |
| <b>2 Implementering av nasjonal eksamen i rettsanvendelse.....</b>                           | <b>12</b> |
| 2.1 Felles innledende eksamen i juridisk metode og rettsanvendelse .....                     | 13        |
| 2.2 Endringer etter de første gjennomføringene.....                                          | 15        |
| <b>Del 2 Eksamen og vurderingskvalitet</b>                                                   |           |
| <b>3 Eksamensutvikling, sensur og utvikling av faglige standarder .....</b>                  | <b>16</b> |
| 3.1 Eksamensutvikling av fagmiljøet .....                                                    | 16        |
| 3.2 Oversikt over eksamenstema.....                                                          | 18        |
| 3.3 Nasjonalt sensorkorps.....                                                               | 19        |
| 3.4 Sensur og kalibrering.....                                                               | 20        |
| 3.5 Oppsummering.....                                                                        | 22        |
| <b>4 Testteoretiske analyser av eksamenskvalitet.....</b>                                    | <b>23</b> |
| 4.1 Rasch-modeller .....                                                                     | 23        |
| 4.2 Data-modelltilpasning.....                                                               | 24        |
| 4.3 Rubrikkenes/poengenes vanskegrad og poenggrensen for bestått .....                       | 28        |
| 4.4 Oppsummering av resultater fra IRT-analyser utført over tid .....                        | 29        |
| <b>5 Reliabilitetsanalyser og sensorsamsvar .....</b>                                        | <b>30</b> |
| 5.1 Datagrunnlag.....                                                                        | 30        |
| 5.2 Tematisk inndeling .....                                                                 | 31        |
| 5.3 Analysemetode .....                                                                      | 32        |
| 5.4 Resultater .....                                                                         | 34        |
| 5.5 Oppsummering.....                                                                        | 41        |
| <b>6 Sensorenes vurdering av eksamenene, vurderingsrubrikker og vurderingsprosessen.....</b> | <b>42</b> |
| 6.1 Datagrunnlag.....                                                                        | 42        |
| 6.2 Eksamen .....                                                                            | 45        |
| 6.3 Vurderingsrubrikker og sensorveiledning.....                                             | 46        |
| 6.4 Nytteeffekt for sensorene.....                                                           | 54        |
| 6.5 Oppsummering.....                                                                        | 55        |

## **Del 3 Eksamenene og studentenes resultater**

|                                                                                                  |           |
|--------------------------------------------------------------------------------------------------|-----------|
| <b>7 Kjennetegn ved de første eksamenene .....</b>                                               | <b>56</b> |
| <b>8 Studentens valg av master .....</b>                                                         | <b>60</b> |
| <b>9 Analyse av studentenes resultater og bakgrunn .....</b>                                     | <b>62</b> |
| 9.1 Datagrunnlag .....                                                                           | 62        |
| 9.2 Metodiske betraktninger .....                                                                | 64        |
| 9.3 Karaktersnitt fra videregående skole .....                                                   | 64        |
| 9.4 Bachelorbakgrunn .....                                                                       | 70        |
| 9.5 Masterprogram .....                                                                          | 73        |
| 9.6 Inneværende karaktersnitt på masterprogrammet .....                                          | 77        |
| 9.7 Oppsummering .....                                                                           | 80        |
| <b>10 Eksamenstema og studentenes prestasjoner .....</b>                                         | <b>82</b> |
| 10.1 Studentenes prestasjoner .....                                                              | 84        |
| 10.2 Avslutning: Rubrikker som grunnlag for kvalitetsutvikling .....                             | 85        |
| <b>11 Avslutning og veien videre .....</b>                                                       | <b>86</b> |
| 11.1 Arena for samarbeid og kvalitetsutvikling .....                                             | 86        |
| 11.2 Eksamen med høy kvalitet og reliabilitet .....                                              | 86        |
| 11.3 Institusjonell variasjon .....                                                              | 87        |
| 11.4 Utfordringer og veien videre .....                                                          | 87        |
| <b>12 Litteraturliste .....</b>                                                                  | <b>89</b> |
| <b>13 Vedlegg .....</b>                                                                          | <b>95</b> |
| 13.1 Krippendorffs alpha for hver rubrikk med tilhørende<br>konfidensintervaller .....           | 95        |
| 13.2 Sensorundersøkelse – Endringer av spørsmålsformuleringer på tvers av<br>undersøkelser ..... | 95        |
| 13.3 Sensorundersøkelse høst 2024 .....                                                          | 96        |

## Sammendrag

Nasjonal eksamen i rettsanvendelse for masterstudenter i barnevern og barnevernsfaglig arbeid ble innført i 2022. Denne rapporten dokumenterer implementeringen av eksamen, og analyserer eksamens- og vurderingskvaliteten, samt studentenes resultater, i perioden høsten 2022 til høsten 2024.

### Implementering og utvikling av eksamen

Innføringen av nasjonal eksamen i juridisk metode og rettsanvendelse bidrar til å sikre et enhetlig og tilstrekkelig høyt juridisk kompetansenivå på tvers av studiesteder. Etableringen av eksamen har skjedd i tett samarbeid med fagmiljøet og har skapt en arena for nasjonalt samarbeid i fagmiljøet med søkelys på vurderingskompetanse og juridisk kompetanse i barnevernet.

### Høy kvalitet og reliabilitet i eksamen og sensur

Arbeidet med vurderingskvalitet i den nasjonale eksamenen følger to komplementære spor. Det første sporet omfatter eksamensutvikling og vurderingsarbeid gjennom skriftlige dokumenter og testteoretiske analyser. Dette inkluderer bruk av vurderingsrubrikker, sensorveiledninger, fagfellevurderinger og studentinnspill, som samlet bidrar til gjennomarbeidede eksamener av høy kvalitet. IRT-analyser viser at rubrikkene fungerer godt, har stor variasjon i vanskegrad og skiller tydelig mellom studenter med høy og lav kompetanse. Analysene dokumenterer høy reliabilitet, særlig rundt grensen for bestått/ikke-bestått, og sensorsamsvaret er gjennomgående høyt i ordinær sensur. Ifølge IRT-analysene er eksamenene ikke for vanskelige for kandidatene. Noen sensorer uttrykker imidlertid bekymring for at rubrikkene i noen tilfeller kan bidra til at faglig svake besvarelser oppnår bestått, ifølge sensorundersøkelsene.

### Felles faglige standarder og vurderingsdialoger

Det andre sporet handler om vurderingsdialoger og kalibrering i fagmiljøet. Fagmiljøet samles i nettverk og arbeider kontinuerlig med forbedringer, utvikle felles faglige standarder, dele erfaringer og diskutere god praksis for læring og vurdering i rettsanvendelse i barnevern.

Underviserne og sensorene understreker viktigheten av nasjonal eksamen som faglig utviklingsarena, og sensorene ser at dette bidrar til deres egen faglige og vurderingsfaglige utvikling. Sensorene rapporterer at dette arbeidet fremmer nettverksbygging og inspirasjon til egen undervisningspraksis og understreker verdien av nasjonal eksamen som en plattform for kvalitetsutvikling og kunnskapsdeling. Samarbeidet kan handle om erfaringsdeling, felles undervisning og deling av læringsressurser. Nasjonal eksamen har også gitt tilgang til kunnskap som ellers ikke ville vært tilgjengelig uten en nasjonal eksamen.

### Studentenes resultater

I perioden 2022–2024 har 935 studenter avlagt eksamen ved 13 utdanningsinstitusjoner. I snitt bestod 58 prosent av studentene eksamen, med betydelig variasjon i prestasjoner og beståttandeler mellom utdanningsinstitusjonene.

Bachelorbakgrunn, dvs. om studenter har en bachelorgrad i vernepleie, sosialt arbeid eller barnevern, forklarer i liten grad variasjonen i studentenes eksamensresultater. Om studentene tar master i barnevern eller master i barnevernsarbeid har også liten forklaringskraft på eksamensresultatene. Alder har en liten, men statistisk signifikant negativ effekt på sannsynligheten for å bestå eksamen.

Mens tidligere analyser har vist at videregående karakterer har vært viktig for å forklare studentprestasjoner på andre nasjonale eksamener, viser dette kun en svak sammenheng for masterstudiene i barnevernsfaglig arbeid. Derimot har karakterer i norsk fra videregående skole en tydelig positiv sammenheng med sannsynligheten for å bestå eksamen. Videre er det en sterk positiv og statistisk signifikant sammenheng mellom studentenes karaktersnitt på masterprogrammet og poengsummen deres på nasjonal eksamen.

Utdanningsinstitusjon viser seg som en viktig forklaringsfaktor for studentenes prestasjoner i alle analyser, også når det kontrolleres for bakgrunnsvariabler. Det er betydelige forskjeller i terskelverdier for å bestå mellom institusjonene, noe som indikerer at institusjonsspesifikke faktorer påvirker studentenes resultater.

Analysene av vurderingsrubrikkene viser at studentene generelt behersker å identifisere juridiske problemstillinger og finne relevante rettskilder. Samtidig ser vi at mange opplever det vanskelig å tolke vilkår. Mange studenter har utfordringer med å vurdere, eller utelater vurderinger knyttet til barnets beste og rettsprinsipper som minste inngreps prinsipp og barnets rett til medvirkning.

### **Veien videre**

Det er behov for å videreutvikle både undervisning og eksamen for å sikre studentenes juridiske metodeforståelse. Den politiske utviklingen med oppmyking av krav til master for arbeid i barnevernet gjør videre arbeid med kompetanseheving i juridisk metode enda viktigere. Denne rapporten bidrar med et kunnskapsgrunnlag for å forstå utfordringene og mulighetene fremover.

## **Abstract**

In 2022, Norway introduced a national examination in legal methodology and application of law for two distinct master programmes in child welfare. This report documents the implementation of the exam and analyses exam quality, assessment practices, and student outcomes from 2022 through 2024. The initiative responded to a long-standing need to strengthen the legal competence in the child welfare sector to safeguard the rights of children and families. The purpose of the national examination is to ensure a consistent and sufficiently high level of legal proficiency across institutions, while also creating a platform for national collaboration.

A key design decision was a common 5 ECTS introductory exam in the first semester to establish a shared legal foundation regardless of bachelor background. The examination is a case-based, individual, and originally four hours (extended to five from autumn 2023). It's graded pass/fail.

## **High quality examinations and reliability in the assessments**

NOKUT and the academic community work with quality in examinations and assessment from two complementary tracks. The first, ensuring quality in assessment through written documents and psychometric—uses rubrics, marker guidelines, peer review (including student input), and IRT analyses to develop high quality exams. IRT-analyses documents well-functioning rubrics, a broad difficulty range, good discrimination, and overall high reliability, particularly around the pass/fail threshold; inter-rater agreement is generally high. Some markers alert that the rubrics occasionally “rescue” weaker students and make them pass despite their competency level as shown in the examination. The second track—assessment dialogues and calibration—brings all the assessors together in a national academic community, discussing and developing shared standards, knowledge and sharing of experiences as well as learning material and resources.

From 2022 through 2024, 935 candidates at 13 institutions have undergone the assessment. The examinations have an average pass rate of 58 %, with substantial institutional variation. Upper-secondary grades show limited and inconsistent predictive power. The grade in Norwegian language and students’ current master’s average grades are strong positive predictors of success; age has a small negative effect, and bachelor and master programme type have limited explanatory importance.

The relatively high non-pass rate shows that continued work is needed both to understand and to develop students’ mastery of legal methodology. In a policy context where the requirement to have a master’s degree in order to work in child welfare has to a large extent been removed, except for leaders, sustained efforts to build legal competence are even more critical. This report provides knowledge to guide further development.



# 1 Introduksjon

I 2021 ble det opprettet nye masterutdanninger i barnevern og barnevernsfaglig arbeid med en obligatorisk nasjonal eksamen i juridisk metode og rettsanvendelse.<sup>1</sup> Målet med nye mastere og nasjonal eksamen var å styrke kompetansen til ansatte i barnevernet, og en nasjonal eksamen ble vurdert som et egnet virkemiddel for å sikre tilstrekkelig juridisk kompetanse hos kandidater som gjennomfører utdanningene (Bufdir, 2019). Kunnskapsdepartementet ga NOKUT i oppdrag å utvikle og forvalte den nasjonale eksamenen.

Den første eksamenen ble gjennomført høsten 2022, og siden høsten 2023 har 13 utdanningsinstitusjoner gjennomført nasjonale eksamener. Det arrangeres to eksamener årlig, en ordinær eksamen og en konteeksamen. Etter to runder med full nasjonal deltakelse ønsket NOKUT å samle erfaringer og frembringe kunnskap: Hvordan har implementeringen gått? Holder eksamenen ønsket kvalitet? Og hvordan har studentene prestert på eksamenen?

Et sentralt mål med en nasjonal eksamen er å gi fagmiljøene mulighet til å sammenligne seg med hverandre og måle juridisk kunnskap og kompetanse på en enhetlig måte på tvers av studiesteder. Dette kan gi tillit til utdanningene og trygghet i at kandidater som uteksamineres har ønsket læringsutbytte. For NOKUT er det også viktig at en nasjonal eksamen fungerer som en arena for refleksjon og dialog om vurdering og læring.

Skal eksamen brukes til sammenligning og kvalitetsutvikling, må den holde høy faglig kvalitet. Rapporten presenterer analyser av både eksamenskvalitet og studentenes resultater i perioden høst 2022 til høst 2024.

Rapportens første del gir ulike innblikk i vurderings- og eksamenskvalitet. Kapittel 2 beskriver innføringen av eksamenen og valg som ble gjort underveis. Del to undersøker eksamenskvaliteten nærmere, og arbeidet med å utvikle eksamenen og vurderingskriterier i faglige fellesskap (kapittel 3). Deretter ser vi nærmere på vurderingskvalitet og sensur (kapittel 4 og 5), inkludert funn fra sensorundersøkelser (kapittel 6). Rapportens tredje del går nærmere inn på studentenes resultater.

Rapporten inngår i en større gjennomgang NOKUT gjør av de første nasjonale eksamenene i rettsanvendelse i barnevernet. Gjennomgangen omfatter erfaringer fra studenter, undervisere, sensorer og programledere samt analyser av eksamen, resultater og sensur for å undersøke ulike aspekter av vurderingskvaliteten. Målet er å videreutvikle kvaliteten på eksamenen og styrke kunnskapsgrunnlaget når det gjelder vurdering og læring i programmene generelt og av juridisk metode i barnevernet spesielt. Rapportene kan også fungere som et referansepunkt for videre oppfølging og evaluering av nasjonal eksamen for rettsanvendelse.<sup>2</sup>

---

<sup>1</sup> Kunnskapsdepartementet fastsatte nye forskrifter for master i barnevern og master i barnevernsarbeid 9. april 2021 (Forskrift om masterutdanning i barnevern 2025; Forskrift om masterutdanning i barnevernsarbeid 2025). Disse ga hjemmel for en nasjonal eksamen.

<sup>2</sup> I oppdragsbrevet av 12. mai 2021 fremgår det at «Ordningen skal vurderes etter fire år fra første eksamen avholdes» (Kunnskapsdepartementet, 2021).

## 1.1 Masterutdanning med nasjonal eksamen: svar på kompetansebehov

Bakgrunnen for nye masterutdanninger og nasjonal eksamen var kompetansebehovet i barnevernet, som hadde vært gjenstand for bekymring i flere tiår. Ansatte i barnevernet har en viktig og krevende jobb. Barnevernet skal «sikre at barn og unge som lever under forhold som kan skade deres helse og utvikling, får nødvendig hjelp, omsorg og beskyttelse til rett tid» (Barnevernsloven 2021, §1-1). Barnevernet skal ivareta de mest sårbare i samfunnet og er satt til å løse faglig komplekse saker. De skal hjelpe til rett tid i saker som kan være krevende å tolke, konfliktfylte og fulle av dilemmaer. Barnevernet kan iverksette særdeles inngripende tiltak som å flytte barn ut av hjemmet og frata foreldreansvar. Arbeidet er faglig og følelsesmessig krevende og stiller høye krav til kompetanse og skjønnsutøvelse. For barnevernets tillit i befolkningen er det viktig at barn og familier som trenger hjelp, møtes av kompetente fagmiljøer.

Flere rapporter peker imidlertid på at sentrale årsaker til svikt i hjelpetilbud til barn handler om mangel på kompetanse og kunnskap i barnevernet (Bufdir, 2019; NOU 2017:12; NOU 2023:7; Oslo Economics, 2019). Det formelle utdanningsnivået er lavt, både sammenliknet med andre nordiske land og andre grupper som barnevernet arbeider med som psykologer, lærere og advokater (Prop. 133L 2020-2021). Barne-, ungdoms- og familiedirektoratet (Bufdir, 2019) påpekte også store variasjoner i innhold og kvalitet i utdanningene, og at det mangler krav og tiltak som sikrer at barnevernet har tilstrekkelig kvalifiserte og kompetente ansatte. Kritikken fra tilsynsmyndigheter og domfellelser i Den europeiske menneskerettsdomstolen<sup>3</sup> aktualiserte ytterligere behovet for kompetanseheving (NOU 2023:7, s.24). Særlig ble mangelen på tilstrekkelig analytisk og juridisk kompetanse trukket frem. Som rettssikkerhetsutvalget skrev:

[...] utvalget er opptatt av kompetanse i barnevernet og mener at for å sikre kvalitet og rettssikkerhet, så er det avgjørende at beslutningstakere og behandlere i alle deler av barnevernssystemet har riktig kompetanse og nødvendig utdanning. Her har også utdanningsinstitusjonene et klart ansvar. Faglig god skjønnsutøvelse forutsetter at utdanningene holder høy kvalitet og har forskningsbasert undervisning. Utvalget mener kunnskap, kompetanse og utdanning er helt avgjørende for et rettsikkert barnevern. (NOU 2023:24, s. 30)

Dahlen og Langsrud (2021) identifiserte også betydelige kompetanseutfordringer i barnevernet knyttet til rettsanvendelse. Til tross for at barnevernet opererer innenfor et komplekst lovverk i stadig endring, har profesjonsutdanningene hatt begrenset innhold i juridiske fag og metodelære, påpeker de. Dette har ført til at ansatte i barnevernet ofte mangler nødvendig juridisk kompetanse. Studien viser at mange opplever seg som selvlærte, utvikler «egne» regler og benytter malbaserte løsninger.

---

<sup>3</sup> Rettssikkerhetsutvalget (NOU 2023:7, s.38) skriver at Helsetilsynet og Riksrevisjonen har avdekket feil og mangler i barnevernstjenestens dokumentasjon, involvering av barn og foreldre samt i arbeidet med undersøkelser (Helsetilsynet 2022 og Riksrevisjonen 2022). Liknende feil og mangler er påpekt av Den europeiske menneskerettsdomstolen (EMD), som har ført til avgjørelser mot Norge. Mellom 2016 og 2023 ble det dokumentert at Norge har gjort feil i 149 barnevernssaker (NOU 2023:7, s.39). Feilene peker særlig på svakheter i avgjørelsesgrunnlag, begrunnelser og avveininger ved beslutninger når det gjelder inngrep overfor foreldre og barn. Rettssikkerhetsutvalget (NOU 2023:7, s. 44) skriver at «Norge er som trolig det landet som har flest barnevernssaker som er behandlet i EMD i hele EMDs historie».

Kompetanseheving var derfor sett på som avgjørende for å sikre kvalitet i barnevernets arbeid og adressere tidligere kritikk (BFD 2021; Prop. 133L 2020-2021). Det ble derfor innført krav om barnevernsfaglig mastergrad for ledere og stedfortredere i barnevernet samt til kommunalt ansatte i barnevernstjenesten som skulle utføre kjerneoppgaver.<sup>4</sup> De barnevernsfaglige masterne skulle inneholde en juridisk nasjonal eksamen for å legge vekt på jusskompetanse.

---

<sup>4</sup> Kompetansekravet skulle gjelde fra 2031, men med overgangsordninger.

# Del 1 Implementeringen

## 2 Implementering av nasjonal eksamen i rettsanvendelse

Da Kunnskapsdepartementet fastsatte nye forskrifter 9. april 2021 (Forskrift om masterutdanning i barnevern 2025 og forskrift om masterutdanning i barnevernarbeid 2025), startet NOKUT arbeidet med å utvikle en nasjonal eksamen i juridisk metode. På dette tidspunktet var flere utdanningsinstitusjoner allerede godt i gang med å utforme masterprogrammene sine i tråd med de nye forskriftene.

Det ble tidlig klart at institusjonene befant seg på ulike stadier i utviklingsprosessen. Noen, særlig de som skulle starte opp høsten 2022, hadde kommet langt og hadde ikke tatt høyde for en nasjonal eksamen eller behovet for koordinering. Andre institusjoner hadde ennå ikke startet arbeidet. For å ferdigstille programplaner, studieplaner og informasjon til søkere i tide før opptak høsten 2022, opplevde flere aktører – inkludert NOKUT – at tidsrammen var stram. Det var mange forhold som måtte avklares, og institusjonene hadde ulike rutiner og frister for godkjenning og akkreditering av studieplaner.

NOKUT utviklet rammene for den nasjonale eksamenen i tett samarbeid med fagmiljøene. Selv om NOKUT hadde erfaring med nasjonale eksamener på andre fagområder siden 2015, var det viktig å bli kjent med de nye fagmiljøene. Gjennom innspillmøter og dialog med fagpersoner, myndigheter og forskere ble sentrale spørsmål diskutert: læringsutbyttebeskrivelser, organisering, vurderingsform og om eksamen skulle være lik for begge masterutdanningene.<sup>5</sup>

NOKUT kartla omfang og innhold av juridiske emner i de ulike bachelorløpene og innhentet informasjon om etterutdanningene i juss for barnevernansatte ved Universitetet i Stavanger og OsloMet samt tilsvarende emner i rettsvitenskap. Det ble også innhentet informasjon om vurderingspraksiser og eksamensvarianter med sensorveiledninger for inspirasjon og kunnskapsgrunnlag. OsloMet hadde benyttet vurderingsrubrikker tidligere, men bruk av vurderingsrubrikker var ikke en utbredt praksis i fagmiljøet. Noen institusjoner benyttet noe tilnærmet mønsterbeskrivelser. Dette ble benyttet som inspirasjon for vurderingsrubrikker og sensorveiledninger.

Kartleggingen viste at det var store forskjeller i hvordan ulike institusjoner og bachelorprogrammer vektla juridisk metode og kompetanse. Det var også store

---

<sup>5</sup> Samtaler og møter ble gjennomført med blant annet forskere, Universitets- og høyskolerådet (UHR), Kunnskapsdepartementet, Barne- og familiedepartementet (BFD) og Barne-, ungdoms- og familiedirektoratet (Bufdir), studenter fra etter- og videreutdanninger som var ansatte i barnevernstjenestene og ansatte ved høyere utdanningsinstitusjoner. Det ble avholdt flest møter med programledere og undervisere hvor rammene for eksamen ble utviklet i fellesskap med utdanningsinstitusjonene. Mer om dette siste arbeidet [her](#) og [her](#).

institusjonelle forskjeller mellom samme bachelorprogram, hvor enkelte program hadde egne emner i forvaltningsrett og materiell rett knyttet til barnevernet, mens andre kun introduserte juridiske perspektiver på et mer overordnet nivå. Noen studieprogramledere og undervisere uttrykte en bekymring for at noen bachelorstudenter skulle møte for høye krav til juridisk kompetanse når de begynte på masterprogrammene, og at det var et gap mellom bachelor- og masternivå. Dette gjaldt særlig studenter til master i barnevernsarbeid som hadde bakgrunn som vernepleier eller som ikke hadde en bachelorutdanning med juss. Noen tok dermed til orde for at en nasjonal eksamen i rettsanvendelse heller burde innføres på bachelornivå. Andre undervisere fremhevet at poenget med de nye masterprogrammene og en nasjonal eksamen var nettopp å heve den juridiske kompetansen. De argumenterte for at tiltak som forkurs heller måtte løse variasjonen i forkunnskap fra bachelornivå på daværende tidspunkt, og at denne eksamenen kunne medføre endringer på bachelornivået i fremtiden.

NOKUT så allerede under implementeringen av eksamen at jusskompetanse på bachelornivå fikk mer oppmerksomhet. Nye emner ble utviklet og eksisterende emner ble justert til å behandle juridiske problemstillinger og tematikker grundigere på bachelornivå.

## 2.1 Felles innledende eksamen i juridisk metode og rettsanvendelse

Ettersom kandidatene skulle ut i samme yrke, valgte fagmiljøene og NOKUT at begge mastere skulle ha en felles eksamen som sikret et likt læringsutbytte. Med en felles eksamen er det verdt å merke seg at masterprogrammene har ulike forskrifter, ulikt opptaksgrunnlag og at læringsutbyttene er noe ulikt formulert og dermed har ulikt omfang. Master i barnevern er en utdanning for barnevernspedagoger, mens master i barnevernsarbeid bygger på bachelor i sosialt arbeid, vernepleie og barnevern. Master i barnevernsarbeid er også åpen for kandidater med annen utdanningsbakgrunn, som for eksempel lærerutdanning og andre pedagogiske utdanninger. Opptakskrav til denne masteren varierer mellom institusjonene. Selv om læringsutbyttene i de to masterforskriftene var ulikt formulert, ble de likevel vurdert av fagmiljøene og NOKUT som substansielt like for en felles nasjonal eksamen.

Det ble vurdert to modeller for eksamen: en mindre inngangseksamen i juridisk metode og en større eksamen på slutten av studieløpet som vurderte alle juridiske læringsutbytter. Flere av de som underviste i juss, ønsket en eksamen på slutten av masterløpet som vurderte hvorvidt studentene hadde oppnådd læringsutbyttene i forskriftene knyttet til rettsanvendelse i barnevernet.

Andre undervisere og særlig noen programledere var bekymret for at juss skulle ta for stor plass i programmene. De ønsket heller en tidlig inngangseksamen som skulle sikre en juridisk grunnmur tidlig i studieløpet, og som deretter kunne bygges på videre med ulike studiemodeller. Flere var opptatt av at en inngangseksamen ville gi institusjonene størst mulig frihet i hvordan de ville organisere, designe og prioritere undervisningen i videre programinnretning.

Valget falt ned på en inngangseksamen med fem studiepoeng i juridisk metode plassert i første semester for å sikre nødvendig forståelse i juridisk metode. Intensjonen var å sikre en felles forståelse av juridisk metode som er avgjørende for barnevernsarbeidet. Noen hevdet enkelte program allerede hadde dette dekket på bachelornivå, mens andre ønsket å sikre at alle masterstudenter uavhengig av bachelorbakgrunn hadde denne kompetansen. Eksamen

skulle gjennomføres to ganger i året, i april og desember, hvor våreksamen derfor primært ble en konteeksamen.

Fagmiljøene og NOKUT utarbeidet læringsutbyttebeskrivelser for emnet denne eksamenen skulle inngå i og ulike tematikker eksamenen kunne knyttes til. Det var enighet om at eksamen i rettsanvendelse i barnevernsfaglig arbeid skulle knyttes til juridisk metode og hvordan man resonerer når man tar stilling til rettslige spørsmål. Eksamenen skulle bygge på læringsutbyttebeskrivelser som sier at kandidatene har inngående kunnskap i juridisk metodelære (rettskildelære) og at kandidaten selvstendig kan identifisere rettslige problemstillinger og anvende juridisk metode for å løse disse. Dette var basert på et konkret forslag utarbeidet av en gruppe jurister fra de ulike studiestedene, ledet av Merete Havre (OsloMet). De andre som deltok i arbeidet var Camilla Bennin (Høgskolen i Innlandet), Karsten Brynildsrud (Høgskolen i Østfold), har Pia M. Hellevik (Universitetet i Stavanger), Elizabeth Langsrud (NTNU), Carola Lingaas (VID), Rikke Rindahl (Universitetet i Tromsø), Randi Sigurdson (Universitetet i Sørøst-Norge), Egil Arne Standal (Høgskulen i Volda) og Ove Veum (Universitetet i Bergen).

Læringsutbyttebeskrivelsene og oppsummeringen av arbeidet høsten 2021 ble sendt til alle institusjoner og dannet således et grunnlag for emnebeskrivelser. Læringsutbytterne oppfyller delvis forskriftene for utdanningene, siden nasjonal eksamen endte med å bli plassert innledningsvis i utdanningen. Det er først på slutten av masterløpet at studentene har fått det totale kunnskapsnivået forskriftene angir for disse utdanningene.

Eksamenen ble satt til fire timer i likhet med tilsvarende emner. Eksamensformen ble en individuell, skriftlig skoleeksamen. Gitt at studentene hadde ulike forkunnskaper, ble det ansett som viktig at studentene skulle ha tilgang til Lovdata Pro og alle rettskilder og hjelpemidler. Fagmiljøene visste også at det kom ny lov 1. januar 2023. Derfor ble det bestemt at studentene kunne velge om de ville benytte den gamle barnevernsloven (eksisterende lov) eller den nye.

Eksamensformen ble valgt til å være casebasert, og at domsanalyser kunne inngå.<sup>6</sup> Teorioppgaver ble også vurdert. En innledende diskusjon var hvorvidt man kunne kreve at studentene selvstendig skulle identifisere de juridiske problemstillingene i et case, eller om de skulle få oppgitt problemstillingene. Gitt læringsutbytte og at dette var på masternivå, ble det konkludert at studentene måtte identifisere disse selv.<sup>7</sup>

Innledningsvis var det også diskusjoner knyttet til felles pensum. Flere emneansvarlige og undervisere trakk frem at det var tidkrevende og utfordrende å finne godt og relevant pensum høsten 2022. Masterne var nye, og emnene knyttet til juridisk metode og rettsanvendelse var nybrotsarbeid. Gitt at emnene er ulikt innrettet, og av ulikt omfang, ble det fastslått at det er opp til hvert studieprogram å velge relevant pensumlitteratur. Flere av underviserne utarbeidet imidlertid læringsressurser og pensumbøker til neste studieår.

Vurderingsuttrykket ble valgt til bestått/ikke-bestått. Bakgrunnen for dette var at eksamen var en inngangseksamen. Studentene møter denne eksamenen første semester etter å ha

---

<sup>6</sup> Etter første gjennomføring ble det presisert at eksamensformen var case-basert hvor studentene skulle identifisere juridiske problemstillinger. Altså at studentene ikke ville få domsanalyser. Vurderingen var at teorioppgaver og domsanalyser ville være vanskeligere for studentene og at caseoppgaver var mer praksisrelevant. Alle eksamensoppgaver, vurderingskriterier og sensorveiledninger ble også offentliggjort og tilgjengelig for studentene.

<sup>7</sup> I senere analyser av eksamensresultat har vi sett at dette ikke var en utfordring for studentene.

mottatt begrenset undervisning på masterstudiet. En tidlig eksamen innebærer derfor også en vurdering av studentenes forkunnskaper ved studiestart.

Vurderingsuttrykket bestått/ikke-bestått kan redusere studentenes stress, øke samarbeid mellom studentene og motivere studentenes faglige forståelse. Det kan også redusere undervisernes ressurser knyttet til vurdering og begrunnelser som heller kan benyttes til studentenes læring (Jørgensen & Bråten, 2019). I tillegg anses vurderingsuttrykket bestått/ikke-bestått som et godt vurderingsverktøy for helhetlig vurdering med utgangspunkt i kvalitative kriterier og profesjonelt skjønn, ifølge Universitets- og høgskolerådet (UHR, 2015a).

## **2.2 Endringer etter de første gjennomføringene**

Etter eksamensgjennomføringene høsten 2022 og våren 2023 ble det gjennomført en foreløpig evaluering med undersøkelser til sensorene og samtaler med alle programledere og undervisere. Her kom det frem at studentene som gikk opp til nasjonal eksamen, var svært nervøse, og de opplevde fire timer som knapt med tid. På bakgrunn av evalueringen ble det gjort noen justeringer. I hovedsak bestod justeringene av tre tiltak: utvidelse av eksamenstiden, kortere eksamenstekster og begrensnig av hjelpemidler, blant annet tilganger i Lovdata Pro.

Den viktigste endringen var utvidelse av eksamenstiden. Eksamenstiden ble utvidet med én time, til totalt fem timer uten at omfang eller andre innholdsmessige justeringer ble gjort for å gi studentene mer ro og tid til å vise hva de kunne. Eksamenssettet og tekstene som studentene får på eksamensdagen ble også forsøkt kortet ned. I tillegg ble det gjort justeringer i hjelpemidlene. Fra flere emneansvarlige ble det påpekt at studentene kun trengte lovtekstene for å svare godt på oppgavene. Derimot var det mange studenter som hadde benyttet unødvendig mye tid på å orientere seg blant ulike kilder. Det ble derfor besluttet å redusere hjelpemidler til kun Lovdata Pro og utvalgte lover og rettspraksis. I tillegg ble det presisert i informasjonen til studentene fra og med høsten 2023 at eksamenen ville bestå av et case. Endringene ble innført til neste ordinære gjennomføring, høsten 2023.

## Del 2

# Eksamen og vurderingskvalitet

### 3 Eksamensutvikling, sensur og utvikling av faglige standarder

NOKUT arbeider langs to hovedspor med nasjonal eksamen for å sikre og styrke kvaliteten i selve eksamenen og vurderingsarbeidet. Det første sporet er analytisk og statistisk orientert, med vekt på psykometriske analyser, vurderingskriterier og sensorpålitelighet. Dette innebærer at det legges ned et stort arbeid for å utvikle gode eksamensoppgaver, tydelig informasjon til studenter, kriterier, vurderingsrubrikker samt presise sensorveiledninger. Språklig tydelighet og presisjon er sentralt i dette arbeidet.

Det andre sporet retter seg mot fagmiljøenes arbeid med vurderingskvalitet for å fremme felles forståelse, felles faglige standarder og kalibrering. En kjerneidé er at arbeidet skjer i fagnettverk preget av fagfellekultur, for å styrke eierskap, legitimitet og kvalitetsutvikling. Vi håper altså å stimulere til såkalte praksisfellesskap, *communities of practice* (Wenger, 1998). Derfor prioriteres dialog og samarbeid i ulike fagnettverk og grupper, med felles møtepunkter. Et grunnleggende prinsipp for nasjonale eksamener er at fagmiljøene selv – de som underviser – skal ha eierskap til eksamenen og ansvar for kvalitetsvurderingen. Fagmiljøene utvikler i fellesskap faglige standarder for hva studentene skal og bør kunne.

Arbeidet på begge spor organiseres på ulike arenaer: i fagmiljøet, i en eksamensgruppe og en referansegruppe samt i det nasjonale sensorkorpset. I tillegg inngår test-teoretiske analyser og sensuranalyser som grunnlag for kontinuerlig forbedring av vurderingskvaliteten. Dette utdypes i videre i del 2.

#### 3.1 Eksamensutvikling av fagmiljøet

For å utvikle den nasjonale eksamen oppnevner NOKUT en faggruppe med sakkyndige, videre kalt en eksamensgruppe. Til eksamensgruppen rekrutterer NOKUT fagpersoner med lang erfaring og ekspertise innen undervisning i rettsanvendelse i barnevern, som faglærere og emneansvarlige fra institusjonene som tilbyr masterutdanningene, til å utarbeide eksamen. Eksamensgruppen består totalt av 4–6 vitenskapelig ansatte fra ulike institusjoner for å sikre faglig forankring fra ulike fagmiljøer ved forskjellige institusjoner og ulike geografiske områder. Personene skal ha kunnskap om undervisning og læring i barnevern generelt og rettsanvendelse og juridisk metode spesielt. Gruppen skal sikre kompetanse og representasjon fra ulike fagområder som rettsvitenskap, barnevern, sosialt arbeid og vernepleie. Som helhet skal gruppen inneha komplementerende kompetanse.



Gruppens sammensetning skal variere over tid og slik sikre spredning av kompetanse og forankring. Et premiss for den første eksamensgruppen var også at den skulle bestå både av ansatte fra institusjoner som startet opp mastere høsten 2022 og de som startet opp høsten 2023.

Eksamensgruppens mandat er å utvikle to nasjonale eksamener årlig med tilhørende vurderingskriterier, rubrikker og sensorveiledning samt å bistå i utviklingen av eksamen og informasjon til studenter.

Fra høsten 2022 til ut året 2023, bestod gruppen av:

- Professor Randi Sigurdson, Universitetet i Sørøst-Norge (leder)
- Førsteamanuensis Kasten Brynildsrud, Høgskolen i Østfold
- Dosent Aina Aune Kane, UiT
- Førsteamanuensis Elizabeth Langsrud, NTNU
- Høgskolelektor Karen Reimers, Høgskulen på Vestlandet
- Førsteamanuensis Egil Arne Standal, Høgskulen i Volda

Fra 1. januar 2024 og ut året 2025, bestod gruppen av:

- Professor Randi Sigurdson, Universitetet i Sørøst-Norge (leder)
- Universitetslektor Pia Moum Hellevik, Universitetet i Stavanger
- Universitetslektor Alexander Zimmermann Alsaker, Universitetet i Agder
- Førsteamanuensis Egil Arne Standal, Høgskulen i Volda

Denne første gruppen hadde en særlig viktig rolle med å utvikle eksamen. I og med at dette var en ny eksamen, for nye masterutdanninger, var dette nybrottsarbeid. Det ble derfor utarbeidet en prøveeksamen. Prøveeksamenen skulle vise eksamenstype, omfang og illustrere nivå. Tanken var at den kunne brukes i undervisning og læringsaktiviteter slik at studenter og undervisere ble litt kjent med eksamenen og formatet.<sup>8</sup> Flertallet av utdanningsinstitusjonene benyttet denne aktivt i undervisningen.

Eksamensgruppen utviklet videre eksamen for høsten 2022. Den bestod av en caseoppgave, en sensorveiledning og detaljerte vurderingsrubrikker.<sup>9</sup> Arbeidet med rubrikker var nybrottsarbeid i fagmiljøet, og diskusjoner om faglige standarder og kriterier var sentralt.

I tillegg til eksamensgruppen oppnevnes det også en referansegruppe. Referansegruppen skal gi innspill til eksamensgruppen og kvalitetssikre eksamensoppgaver, sensorveiledninger, vurderingsrubrikker og andre eksamensdokumenter. De skal også gi råd om videreutvikling av eksamen. De samme utvelgelseskriteriene gjelder for denne gruppen, i tillegg til at det skal være en studentrepresentant, og det er ønskelig med representasjon fra praksisfeltet. Studenten skal utfra et studentperspektiv gi tilbakemeldinger på eksamen. Dette er blant annet for å bidra til at eksamensoppgaven forstås av studentene i henhold til intensjonen, at arbeidsmengden er tilpasset eksamenstiden og at vanskegraden er rimelig.

I den første referansegruppen ble det valgt en studentrepresentant som hadde gjennomført etter- og videreutdanning i juss da nye mastere etter nye forskrifter akkurat var opprettet. I tillegg var personen ansatt i barnevernstjenesten, mens en annen var nemndleder for Barneverns- og helsenemda. Disse hadde et mer praksisrettet blikk på

---

<sup>8</sup> Prøveeksamen kan finnes [her](#).

<sup>9</sup> Fra og med høsten 2004 ble sensorveiledningen og vurderingsrubrikkene slått sammen til et dokument.

utformingen. Referansegruppen skal sikre profesjonstilknytning og relevans for yrket som barnevernsstudentene skal møte.

Referansegruppen fra høsten 2022 og til desember 2023:

- Universitetslektor og nemndleder Ove Kurt Veum, Universitetet i Bergen og Barneverns- og helsenemnda
- Førstelektor Inger Therese Øvrur, VID vitenskapelige høyskole
- Universitetslektor Pia Moum Hellevik, Universitetet i Stavanger
- Student Hege Håheim med fullført etter- og videreutdanning i juss og ansatt i barnevernstjenesten i Stavanger kommune
- Universitetslektor Alexander Zimmermann Alsaker, Universitetet i Agder<sup>10</sup>

Referansegruppe fra 1. januar 2024 og ut året 2025:

- Universitetslektor og nemndleder Ove Kurt Veum, Universitetet i Bergen og Barneverns- og helsenemnda
- Førstelektor Inger Therese Øvrur, VID vitenskapelige høyskole
- Dosent Aina Aune Kane, UiT – Norges arktiske universitet
- Førstelektor Håkon Glommen Eriksen, Universitetet i Innlandet
- Ingrid Ekeland Iversen (studentrepresentant), Høgskulen på Vestlandet (2024)
- Ingrid Stevning (studentrepresentant), OsloMet (2025)

I tillegg til disse gruppene har NOKUT knyttet til seg eksperter innen psykometri og testteori. Disse gir innspill på eksamenssett, oppgavetyper og spørsmål, vurderingskriterier og sensorveiledninger. I tillegg gjør de analyser av sensur og eksamenskvaliteten etter eksamensgjennomføringen. Hensikten er å forbedre og utbedre eksamenskvaliteten gjennom å identifisere hva som har fungert bra og hva som ikke gikk som intendert. Sammen med NOKUTs ansatte med kompetanse i vurdering gir det gode faglige rammer for å utvikle eksamen i et faglig fellesskap.

## 3.2 Oversikt over eksamenstema

Studentene skal identifisere juridiske problemstillinger i en caseoppgave. I tabell 1 under er en oversikt over tema og problemstillinger som er gitt ved de ulike gjennomføringene.

Tabell 1. Oversikt over eksamenstema og eksamensdokumenter.

| Gjennomføring          | Tema                                                                                | Eksamensdokumenter                                                                                                                                                         |
|------------------------|-------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Prøveeksamen høst 2022 | 1. Partsrettigheter<br>2. Taushetsplikt – meldeplikt<br>3. Innsyn                   | <ul style="list-style-type: none"><li>• <a href="#">Eksamensoppgave</a></li></ul>                                                                                          |
| Høst 2022              | 1. Melding<br>2. Undersøkelse<br>3. Taushetsplikt<br>4. Pålegg om å gi opplysninger | <ul style="list-style-type: none"><li>• <a href="#">Eksamensoppgave</a></li><li>• <a href="#">Sensorveiledning</a></li><li>• <a href="#">Vurderingsrubrikker</a></li></ul> |
| Vår 2023               | 1. Tilbakeføring etter omsorgsovertakelse<br>2. Samvær, omfang                      | <ul style="list-style-type: none"><li>• <a href="#">Eksamensoppgave</a></li><li>• <a href="#">Sensorveiledning</a></li><li>• <a href="#">Vurderingsrubrikker</a></li></ul> |

<sup>10</sup> Fra og med høsten 2023 til desember 2023.

|           |                                                                                                                                                            |                                                                                                                                                                                |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|           | 3. Partsrettigheter                                                                                                                                        |                                                                                                                                                                                |
| Høst 2023 | <ol style="list-style-type: none"> <li>1. Akuttvedtak adferdssak</li> <li>2. Habilitet</li> <li>3. Nemd pålegge foreldre-støttende hjelpetiltak</li> </ol> | <ul style="list-style-type: none"> <li>• <a href="#">Eksamensoppgave</a></li> <li>• <a href="#">Sensorveiledning</a></li> <li>• <a href="#">Vurderingsrubrikker</a></li> </ul> |
| Vår 2024  | <ol style="list-style-type: none"> <li>1. Kroppsvisitasjon institusjon</li> <li>2. Mild form for fysisk makt</li> <li>3. Klagerett</li> </ol>              | <ul style="list-style-type: none"> <li>• <a href="#">Eksamensoppgave</a></li> <li>• <a href="#">Sensorveiledning</a></li> <li>• <a href="#">Vurderingsrubrikker</a></li> </ul> |
| Høst 2024 | <ol style="list-style-type: none"> <li>1. Partsrettigheter</li> <li>2. Hjemmebesøk</li> <li>3. Innsynsrett</li> <li>4. Omsorgsovertakelse</li> </ol>       | <ul style="list-style-type: none"> <li>• <a href="#">Eksamensoppgave</a></li> <li>• <a href="#">Sensorveiledning</a></li> </ul>                                                |

### 3.3 Nasjonalt sensorkorps

Eksamensbesvarelsene vurderes av et nasjonalt sensorkorps som rekrutteres av NOKUT etter bestemte kriterier.<sup>11</sup> Sensorene må ha mastergrad eller høyere grads embetseksamen i rettsanvendelse eller tilsvarende relevant juridisk kompetanse knyttet til barnevern. Sosionomer, vernepleiere og barnevernspedagoger må ha utdanning i rettsvitenskap eller juridiske tematikker ut over egen grunnutdanning. Lengre erfaring fra undervisning i rettsanvendelse i barnevern kan kompensere for formell utdanning i faget. Videre er det ønskelig at sensor har erfaring fra undervisning i rettsanvendelse i barnevernsutdanning og/eller erfaring som sensor i liknende fag.

Sensorene representerer alle institusjonene som tilbyr master i barnevern og barnevernsfaglig arbeid i Norge. Dette er viktig for å sikre forankring i fagmiljøene som underviser i emnet og at læring og vurdering henger sammen. I tillegg er det sensorer som arbeider som nemndledere for Barneverns- og helsenemnda. Slik sikres en tilknytning til både barnevernsfagfeltet og yrkesutøvelsen.

Både eksamensgruppen og referansegruppen inngår i det nasjonale sensorkorpsset. I tillegg til å være undervisere er det viktig at disse gruppene ser hvordan oppgavene og vurderingskriteriene fungerer i praksis og deltar aktivt i vurderingsdialoger med andre sensorer og fagpersoner.

Totalt består sensorkorpsset av rundt 30 fagpersoner og skal, i likhet med gruppene, være tett tilknyttet undervisning- og emneutvikling for rettsanvendelse og juridisk metode. NOKUT har rekruttert alle som underviser rettsanvendelse og juridisk metode ved begge masterprogrammene til å være sensorer.

I oppstarten var det få ansatte som arbeidet med rettsanvendelse i barnevern hos de ulike institusjonene. Dette gjorde det vanskelig å rekruttere tilstrekkelig med sensorer. Mange av de som underviste rettsanvendelse i barnevern utrykte at de arbeidet alene, og flere institusjoner var i rekrutteringsprosesser internt. I tillegg til dette var det lite samarbeid på tvers av institusjonene for dette fagmiljøet, og det var få arenaer for erfaringsdeling (jf. Pedersen & Bråten, 2025).

NOKUT arbeidet derfor aktivt med å rekruttere nok sensorer til denne eksamenen. Det har vært utfordrende, men underviserne har fått økt kjennskap til hverandre gjennom arbeidet med eksamens- og referansegruppen og sensorkorpsset. Disse arenaene har blitt faglige fellesskap og nasjonale nettverk for erfaringsdeling og samarbeid. Særlig for små fagmiljøer

<sup>11</sup> Kriteriene er utviklet av den første eksamensgruppen basert på innspillsrundene.

har dette vært nyttig. Arbeidet med vurdering og nasjonal eksamen har gitt fagpersoner bedre kjennskap til hverandre, etablert et faglig nettverk og styrket samarbeidet. Dette gjør at sensorer og medlemmer i faggruppene ønsker å fortsette.

### 3.4 Sensur og kalibrering

Sensurarbeidet baseres på skriftlige dokumenter og retningslinjer som gir sensorene informasjon om sensurprosessen og sensuren, som sensorinstruks, vurderingsrubrikker og sensorveiledning. Eksamensgruppen, sammen med referansegruppen, har utarbeidet slike retningslinjer, med hjelpetekster, for å sikre en enhetlig sensur. Det legges ned stor arbeidsinnsats i forberedelse av eksamenen for at disse dokumentene og verktøyene skal ha høy kvalitet, med kvalitetssikring i mange ledd.

Slike skriftlige standarder er viktige verktøy for å sikre en rettferdig og reliabel sensur. Ecclestone (2001, s. 302 sitert i Bjølseth et al., 2011) poengterer at slike skriftlige dokumenter som kriterier og læringsmål, gir grunnlag for mer moderering, kalibrering og standardisering mellom sensorer, noe som bidrar til å styrke reliabiliteten og vurderingskvaliteten. Samtidig skapes ikke høy vurderingskvalitet kun gjennom skriftlige dokumenter, som trekkes frem blant annet av Bloxham, Reimann & Rust (2018, s. 2):

Repeated studies over many years demonstrate considerable inconsistency in academics' judgements about student performance and variation in the meaning they accord to written standards.

Slike funn finner vi igjen i Norge (Bjølseth et al., 2011; Gynnild, 2013; UHR, 2015b; UHR, 2016). Å utvikle og definere klare vurderingskriterier betyr ikke nødvendigvis at sensorene benytter dem i praksis eller benytter dem på samme måte. Skriftlige dokumenter krever tolkning og operasjonalisering for å bli praksis. Ulike personer kan legge ulik mening i ord, og vurderingskriteriene og veiledningene må knyttes til spesifikke caser, tema og situasjoner som ikke forstås og tolkes likt. I likhet med læringssituasjoner er også vurderingssituasjoner kontekstspesifikke.

Ved nasjonal eksamen arbeides det derfor på ulike måter for å fremme enhetlige praksiser og felles vurderingsstandarder. Det arbeides med en type «opplæring» og refleksjon knyttet til vurderingskvalitet gjennom å diskutere hva som er god eksamenskvalitet, rettferdige oppgaver og gode sensorveiledninger blant annet i lys av test-teori. I tillegg legges det stor vekt på at fagmiljøet sammen definerer og finner kjennetegn på kvalitet og faglige standarder. Mange av disse elementene handler om å sette ord på taus kunnskap og ha kollegiale diskusjoner om vurdering, sensur og faglig kvalitet. Dette kaller vi *vurderingsdialoger*, basert på forskning som vektlegger disse kollegiale fagfellesskap (Bjølseth et al., 2011; Bloxham et al., 2018; Price, 2005). Ulik tolkning og praksis knyttet til vurderingsstandarder, som Bloxham med kolleger (2018) beskrev, viser viktigheten av vurderingsdialoger i fagmiljøet. Målet med vurderingsdialogene er å ha en reliabel og enhetlig sensur, samtidig som vi utvikler nasjonale faglige standarder.

En måte NOKUT arbeider konkret med vurderingsdialog på er ved å vektlegge kalibrering i selve sensurarbeidet, hvor sensorene arbeider sammen i par og i grupper samt i felles nasjonale kalibreringsmøter for hele sensorkorpset. Hver besvarelse vurderes av to sensorer. Sensorene kjenner ikke studentenes identitet eller institusjonstilknytning. Sensorene sensurerer besvarelser på tvers av institusjoner. Alle sensorene sensurerer derfor i par. De foretar først en individuell vurdering av hver besvarelse i henhold til vurderingsrubrikker og sensorveiledning. De gjør en analytisk vurdering basert på

vurderingskriterier (rubrikker) og ikke en holistisk skjønnsmessig vurdering. Deretter skal sensorene samsensurere ved å diskutere seg frem til en enighet på poeng på hvert vurderingskriterium og endelig karakter. For studenten er det samsensuren som utgjør begrunnelsen<sup>12</sup> og resultatet hen får. Slik skjer det en kalibrering og diskusjon av faglige standarder i par.

I tillegg til sensorparene settes også sensorene sammen i grupper på fire–fem personer. Hver sensor inngår i tre, fire eller fem ulike sensorpar, som til sammen utgjør en gruppe. Sensorgruppene settes sammen utfra institusjonstilhørighet, profesjon og sensorerfaring. Eksempelvis settes en ny sensor fra VID med juristbakgrunn sammen med en eldre sensor fra HVL med sosionombakgrunn. Medlemmer i eksamens- og referansegruppen spres i ulike par og grupper så langt det lar seg gjøre. Dette betyr at sensorgruppene dekker ulike kompetanse og institusjonstilknytning. Slik skal sensurarbeidet bidra til kalibrering og diskusjon av faglige standarder på tvers av institusjoner og grupper.

I tabell 2 ser vi fordelingen av antall besvarelser på tvers av gjennomføringer. Dette kan si oss noe om sensorenes kalibreringsgrunnlag for hver gjennomføring.

Tabell 2. Fordeling av besvarelser per gjennomføring.

| Gjennomføring | Antall kandidater | Antall besvarelser per sensor | Antall sensorpar | Antall sensorer |
|---------------|-------------------|-------------------------------|------------------|-----------------|
| Høst 2022     | 165               | ~20                           | 25               | 32              |
| Vår 2023      | 70                | ~12                           | 18               | 12              |
| Høst 2023     | 300               | ~18                           | 48               | 31              |
| Vår 2024      | 133               | ~17                           | 24               | 16              |
| Høst 2024     | 267               | ~16                           | 42               | 30              |

I tillegg til sammensetning og samarbeid i sensorpar og sensorgrupper arbeider sensorkorpset samlet i nasjonale kalibreringsmøter. Først et innledende møte etter sensorene har sensurert et lite utvalg besvarelser. Dette er for å avklare uklarheter og nyanser i selve oppgaven og vurderingskriteriene og for å bli enige om hva som kreves av nivå på ulike kriterier. Deretter avholdes et kalibreringsmøte hvor sensorene går mer detaljert inn i eksamenenes vanskegrad, læringsutbytter og setter karaktergrenser for bestått og ikke bestått. Begge møtene benyttes for å moderere, kalibrere og diskutere hva som er faglig kvalitet.

Det arbeides også med kalibrering og felles standarder i fagnettverk knyttet til nasjonal eksamen, for eksempel ved hjelp av mønsterbesvarelser som deles med alle som underviser i emnet. Dette betyr at i tillegg til å fokusere på skriftlige dokumenter arbeides det også med hva Bloxham et al. (2018) kaller *uformell kalibrering*, en prosess hvor akademiske standarder deles over tid, muntlig, gjennom kontakt mellom utdanningsinstitusjoner og i nasjonale fagnettverk. Allerede i 1997 ble slike tiltak pekt på som viktige for god vurderingskvalitet og med økende viktighet i kvalitetssikring nasjonalt for å opprettholde faglige standarder gitt økende student-tall i høyere utdanning.

Consistent assessment decisions among assessors are the product of interactions over time, the internalisation of exemplars, and of inclusive networks. Written instructions, mark schemes and criteria, even when used with scrupulous care, cannot substitute for these. Higher Education Quality Council (HEQC), 1997.

<sup>12</sup> I de første gjennomføringene også sammen med sensorveiledningen.

### **3.4.1 Klagesensur**

Dersom studentene klager, vil besvarelsen sensureres av et nytt sensorpar gjennom samme vurderingsprosess som i ordinær sensur. Selv om prosessen er lik, er riktignok konteksten for klagesensuren ulik. For det første foretas klagesensur etter sensurfrist og naturlig nok ordinær eksamen. Det betyr at sensorene har et større kunnskapsgrunnlag fra tidligere sensur med flere sensorer og kalibreringsmøter og diskusjoner når de starter på klagesensuren. Dette er ikke alltid tilfelle ved lokalt gitte eksamener. Det er viktig at sensorene har med seg nivået fra tidligere kandidater og diskusjoner i sensorkorpset som helhet. For det andre er det rimelig å anta at sensorene vet at kandidatene som klager, ikke har bestått, i og med at eksamen vurderes med bestått/ikke-bestått. Sensorene har i så måte kjennskap til opprinnelig vurdering. Samtidig kjenner ikke sensorene i klagesensuren til hvor langt unna beståttgrensen besvarelsen var vurdert tidligere og hvilke vurderingsrubrikker de tidligere sensorene har gitt høy eller lav uttelling.

## **3.5 Oppsummering**

For å oppsummere kan vi derfor si at NOKUT arbeider med eksamens- og vurderingskvalitet i nasjonale eksamener gjennom arbeid med gode skriftlige dokumenter som eksamensoppgaver, veiledninger og informasjon til studenter, sensorveiledninger og vurderingskriterier. Dette gjøres gjennom å involvere bredt og store deler av det nasjonale fagmiljøet, blant de som underviser i faget og som deltar i eksamensgruppen, referansegruppen og i sensorkorpset. Det skaper felles kunnskap og refleksjon knyttet til vurdering og gode eksamensoppgaver. Dette involverer arbeid med mønsterbesvarelser. I tillegg jobbes det med kalibrering, felles nasjonale faglige standarder og et felles språk og operasjonalisering gjennom samarbeid og dialog, blant annet i nasjonale sensormøter og i fagnettverk. Eksamensarbeidet involverer hele fagmiljøet som underviser i juridisk metode og rettsanvendelse. Studenter og praksisfeltet involveres også i dette arbeidet.

I tillegg til å fokusere på involvering fra fagmiljøer samarbeid og felles standarder evaluerer NOKUT og fagmiljøet kvaliteten i eksamen og arbeidet. Dette gjøres blant annet gjennom testteoretiske analyser, sensuranalyser og sensorundersøkelser. Dette vil vi nå gi innblikk i.

## 4 Testteoretiske analyser av eksamenskvalitet

Testteori gir et matematisk rammeverk for å beskrive hvordan en eksamen fungerer. Klassisk testteori er basert på korrelasjoner og gir informasjon om en eksamen på et helhetlig nivå fremfor informasjon om eksamenens enkeltoppgaver. Moderne testteori, og mer spesifikt *Item response theory* (IRT), er basert på sannsynlighet og vektlegger derimot hvordan enkeltoppgavene fungerte. IRT-analyser hjelper oss dermed til å analysere og beskrive eksamensoppgavenes kvalitet. NOKUT<sup>13</sup> har over tid brukt en type IRT-modeller kalt Rasch-modeller (Rasch, 1960). Analysene brukes av eksamensgruppene i oppgaveutvikling og utforming av vurderingsrubrikker. De legger videre grunnlag for diskusjoner i fagmiljøene om kjennetegn på gode oppgaver, rettfærdig sensur, og hva som er tilstrekkelige kunnskaper og ferdigheter hos studentene.

NOKUT gjennomførte IRT-analyser av data fra eksamen høsten 2023, våren 2024 og høsten 2024. I dette kapittelet beskriver vi kort hva IRT-analyser er, og vi illustrerer med eksempler fra eksamen høsten 2024. Vi vektlegger analyser som gir grunnlag for å diskutere hvordan eksamen kan videreutvikles, og som gir innblikk i hvordan NOKUT og eksamensgruppen bruker testteori i arbeidet med å utvikle eksamen.

### 4.1 Rasch-modeller

IRT-modeller undersøker sammenhengen mellom kandidatens underliggende (latente) egenskaper – som den ikke-direkte målbare variabelen «dyktighet i rettsanvendelse» – og kandidatens observerte prestasjoner på de individuelle oppgavene. I vårt tilfelle uttrykkes observerte prestasjoner om kandidaten oppnådde vurderingskriteriene gitt ved de ulike vurderingsrubrikkene.

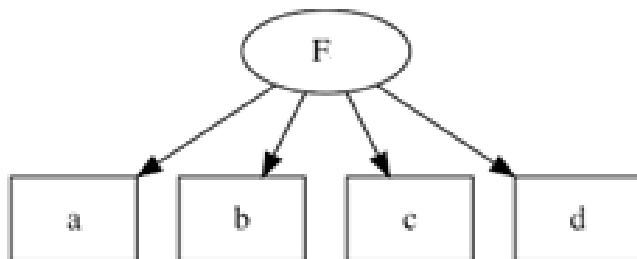
Figur 1 illustrerer at den underliggende uobserverte faktoren ( $F$ ) «dyktighet i rettsanvendelse» styrer kandidatens svar og dermed om vurderingskriteriene a–d er oppfylt, og hvordan de tilhørende rubrikkene er poengsatt (observerte variabler). Kriteriet som ligger til grunn for riktig svar, kan være vanskelig eller lett å oppnå. Om kandidaten oppnår 0 eller 1 poeng på en rubrikk avhenger av forholdet mellom rubrikkens vanskegrad (estimert som  $b$ ) og dyktigheten til kandidaten (estimert som  $\theta$ ). Estimaten for kandidatens dyktighet  $\theta$  gjenspeiler kandidatens oppnådde poengsum på eksamen, og estimaten for vanskegraden  $b$  til en gitt rubrikk gjenspeiler hvor mange kandidater som oppnådde vurderingskriteriet.<sup>14</sup>

<sup>13</sup> Det er Øystein Guttersrud, førsteamanuensis OsloMet/Naturfagsenteret som har gjort disse analysene for NOKUT.

<sup>14</sup> Teknikken «maximum likelihood estimation» ligger til grunn for estimeringene, og statistikkpakkene «mirt» (Chalmers, 2012) og «RUMM2030 Plus» (Andrich and Sheridan, 2019) ble brukt.



Figur 1. Illustrasjon av at ett latent trekk eller én faktor  $F$  (endimensjonal modell) styrer de observerte verdiene på variablene  $a$ – $d$ .



#### 4.1.1 Logistisk regresjon med logit som linkfunksjon

Siden  $\theta$  og  $b$  er kontinuerlige og kan variere mellom  $-\infty$  og  $+\infty$ , mens variabelen for riktig/feil svar er kategorisk, innføres sannsynligheten  $p$  for riktig svar fordi «den naturlige logaritmen til oddsen»  $\ln \frac{p}{1-p}$  varierer mellom  $-\infty$  og  $+\infty$ . Vi kan dermed kople eller «linke» det hele sammen ved å bruke «logaritmen til oddsen» (logit) som linkfunksjon:

$$\ln \left( \frac{p}{1-p} \right) = \theta - b$$

Oddsene for riktig svar øker dermed med økende dyktighet  $\theta$ , og den avtar med økende vanskegrad  $b$ . Uttrykket representerer Rasch-modellen for dikotome variabler (Rasch, 1960). Uttrykket kan omskrives til  $p = \frac{e^{(\theta-b)}}{1+e^{(\theta-b)}}$ , som angir sannsynligheten  $p$  for riktig svar som funksjon av kandidatens dyktighet og oppgavens vanskegrad.

Den heltrukne kurven i figur 2 er den grafiske representasjonen av Rasch-modellen for dikotome vurderingsrubrikker (0 eller 1 poeng).<sup>15</sup>

## 4.2 Data-modelltilpasning

For at oppgaver skal være rettferdige, bør dyktige studenter ha høyere sannsynlighet for å svare riktig enn faglige svakere studenter. Det er altså en sammenheng mellom målt dyktighet, sannsynligheten for å svare riktig og dermed å få poeng på rubrikken. Ved å vurdere ulike rubrikkers tilpasning til modellen, altså om de observerte verdiene passer godt til de forventede verdiene predikert av Rasch-modellen, kan vi vurdere om vurderingskriteriet fungerte bra og var rettferdig. Vi gir nedenfor eksempler på en rubrikk som passet godt til modellen og en som kunne fungert bedre.

Kurven i figur 2 fremstiller sannsynligheten  $p$  for riktig svar (y-aksen) som funksjon av dyktigheten (x-aksen) for en rubrikk med vanskegraden  $b = 0,52$ . Det betyr at kandidater med dyktigheten  $\theta = 0,52$  hadde 50 prosent sannsynlighet for 1 poeng:

<sup>15</sup> Andre IRT-modeller, som 2PL (Hambleton, Swaminathan, og Rogers, 1991, s. 15), innfører diskrimineringsparameteren  $a$  slik at  $\ln \left( \frac{p}{1-p} \right) = a(\theta - b)$  og  $p = \frac{e^{a(\theta-b)}}{1+e^{a(\theta-b)}}$ . Parameteren  $a$  endrer stigningstallet til kurven i vendepunktet slik at kurven blir brattere eller slakere.

Hensikten er å forbedre tilpasningen mellom data og modell. NOKUT bruker Rasch-modeller fordi innføring av parameteren  $a$  medfører at modellen mister fundamentale egenskaper, inkludert «spesifikk objektivitet» (Reise og Revicki, 2015, s. 291).

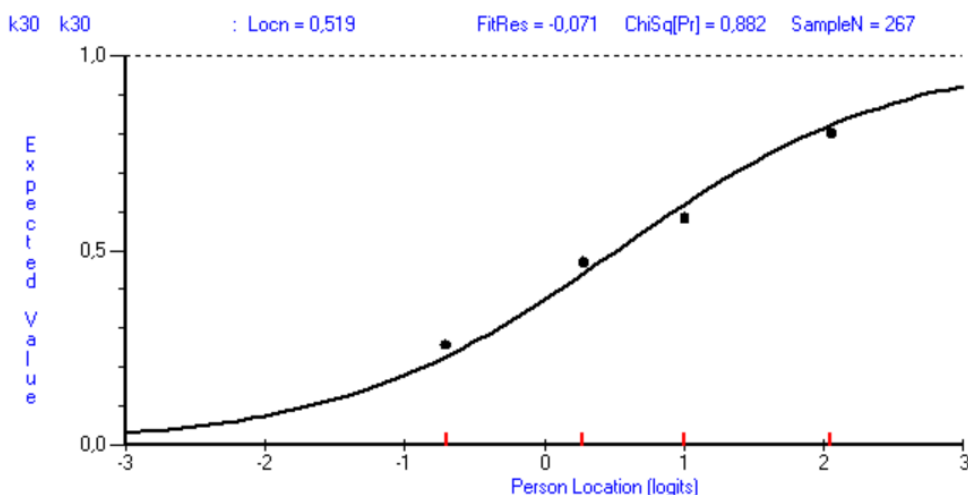


$$p = \frac{e^{(\theta-b)}}{1+e^{(\theta-b)}} = \frac{e^{(0,52-0,52)}}{1+e^{(0,52-0,52)}} = \frac{e^0}{1+e^0} = \frac{1}{1+1} = 0,50 = 50 \%$$

De fire målepunktene angir andelen riktige svar i fire «dyktighetsintervaller» på x-aksen (hvert intervall inneholder ca. 25 prosent av kandidatene). Punktet lengst mot venstre med koordinatene (-0,71, 0,26) viser at de 25 prosent faglig svakeste kandidatene hadde gjennomsnittlig dyktighet -0,71, og at 0,26 eller 26 prosent av dem fikk 1 poeng. Punktet lengst mot høyre viser at blant de 25 prosent faglig dyktigste fikk 80 prosent 1 poeng. Den aktuelle rubrikken ga data med god tilpasning til Rasch-modellen. God tilpasning betyr at vurderingskriteriet fungerte bra. Kriteriet var «rettferdig» fordi det «diskriminerte» godt ved å gi poeng til mange dyktige kandidater, samtidig som det ga poeng til få faglig svake kandidater.

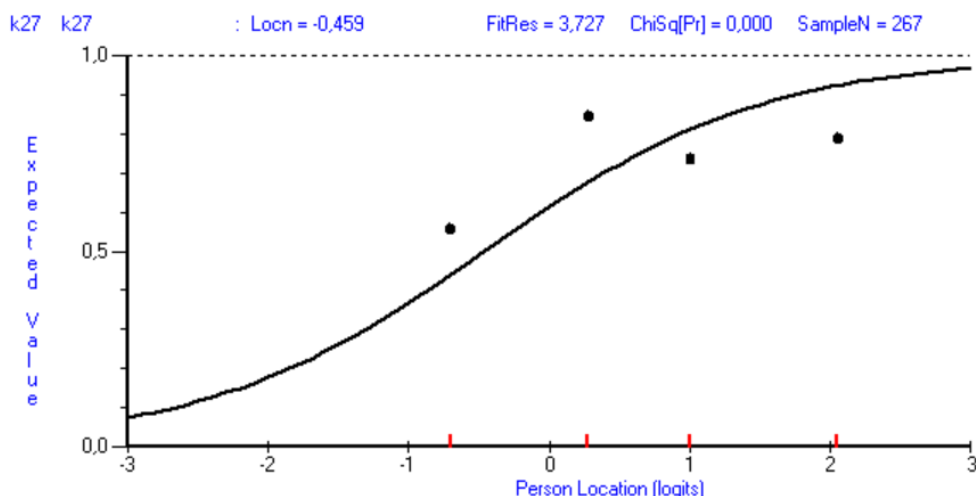
Figur 3 viser tilpasningen mellom data og modell for en rubrikk som *ikke* fungerte tilstrekkelig godt. Som et ledd i videreutviklingen av kvaliteten på eksamen får eksamensgruppen tilbakemelding om hvilke vurderingskriterier/rubrikker som har svak tilpasning til Rasch-modellen og dermed kunne fungert bedre.

Figur 2. Sannsynligheten for riktig svar (y-aksen) på rubrikk 30 høst 2024<sup>16</sup> som funksjon av dyktighet (x-aksen). Målepunktene viser faktisk andel kandidater som fikk poeng i hvert av fire «dyktighetsintervallene» fra lav til høy dyktighet. De observerte verdiene (punktene) passer godt til de teoretisk forventede verdiene (kurven).



<sup>16</sup> Rubrikk 30 var: «Stiller spørsmål ved om det finnes bestemmelse som gir Karl rett i å få innsyn i hvilke faktiske opplysninger den sakkyndige har lagt til grunn.»

Figur 3. Sannsynligheten for riktig svar (y-aksen) på rubrikk 27 høst 2024<sup>17</sup> som funksjon av dyktighet (x-aksen). Målepunktene viser faktisk andel kandidater som fikk poeng i hvert av fire «dyktighetsintervallene» fra lav til høy dyktighet. De observerte verdiene (punktene) passer ikke godt til de teoretisk forventede verdiene (kurven).



#### 4.2.1 Endimensjonalitet

Endimensjonale Rasch-modeller krever at dataene er endimensjonale og «lokalt uavhengige». Høsten 2024 rangerte poengsummen basert på rubrikkene 14–35 (delskala 1) «for mange» kandidater på «for forskjellig måte» enn poengsummen basert på rubrikkene 1–13 og 36–58 (delskala 2). Det betyr at disse to delene av eksamenen syntest å måle noe forskjellige kompetanser, og at eksamenen dermed ikke var endimensjonal. Vi fremhever at når eksamen utformes som en sammensetning av ulike delskalaer med hensikt å øke validiteten, er brudd på endimensjonalitet uunngåelig. NOKUT og eksamensgruppen vurderer det derfor dithen at noe multidimensjonalitet er en positiv egenskap ved denne nasjonale eksamenen.

#### 4.2.2 Lokal uavhengighet

Lokal uavhengighet betyr at gitt én dominerende faktor eller latent egenskap (dyktighet i rettsanvendelse), skal poengsettingen av en rubrikk være statistisk uavhengig av poengsettingen på andre rubrikker. Lokal uavhengighet kan brytes ved at a) andre faktorer enn den ene dominerende faktoren har sammenheng med poengsettingen, eller at b) noen rubrikker er «for like» og har mer til felles med hverandre enn den ene dominerende faktoren. Fenomenene kalles henholdsvis *differential item functioning* (DIF) og *svaravhengighet*, hvor det sistnevnte undersøkes ved å korrelere avvikene fra Rasch-modellen (residualene) parvis på tvers av rubrikkene.

Et eksempel på en «annen faktor» enn dyktighet, kan være kandidatens kjønn ved at en eller flere rubrikker «favoriserer» kvinnelige kandidater. DIF-analyser har ikke blitt utført da bakgrunnsvariabler, som kandidatens kjønn, ikke var tilgjengelige på analysetidspunktene.

<sup>17</sup> Rubrikk 27 høsten 2024 inneholder to elementer: «Viser til bvl. § 12-6 første ledd og klargjør at en part har rett til å gjøre seg kjent med sakens dokumenter etter regler i forvaltningsloven.»

I likhet med tidligere eksamener, var det høsten 2024 flere par av statistisk avhengige rubrikker – altså rubrikker som var «for like» hverandre. Et eksempel på svært sterk svaravhengighet finner vi mellom rubrikkene 32 og 35 (se tabell 3).

Tabell 3. Eksempel på rubrikker med sterk svaravhengighet på eksamen høsten 2024.

|     |                                                                                                                                                                                                                                                                                    |                                        |                               |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|-------------------------------|
| 32. | Finner relevant lovbestemmelse i fvl. § 18 c «Selv om dokumentet eller deler av det er unntatt etter reglene i §§ 18a og 18b, har parten rett til å gjøre seg kjent med de deler av det som inneholder faktiske opplysninger eller sammendrag eller annen bearbeidelse av faktum.» | Blir ikke gjort<br>0 poeng             | Blir gjort<br>1 poeng         |
| 35. | Konkluderer med at Karl etter fvl. § 18 c har rett til innsyn i faktiske opplysninger                                                                                                                                                                                              | Ingen eller feil konklusjon<br>0 poeng | Korrekt konklusjon<br>1 poeng |

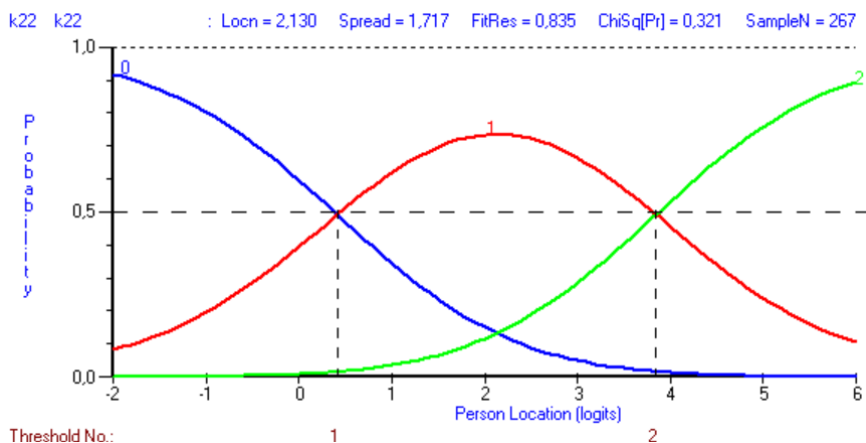
Denne typen analyser har blitt utført siden høsten 2023, og eksamensgruppen jobber kontinuerlig med å begrense svaravhengighet og forklare svaravhengighet i lys av faglig innsikt. I juridisk metode følges prosedyrer med flere steg. Studentene kan dermed ikke konkludere på et forsvarlig grunnlag uten å ha funnet relevant lovbestemmelse, og dermed observeres en viss avhengighet. En mulighet er å slå sammen «avhengige» dikotome rubrikker og skåre den resulterende rubrikken polytomt.

### 4.2.3 Polytome rubrikker

Ideene beskrevet ovenfor gjelder også for polytome rubrikker (ingen uttelling (0 poeng), delvis uttelling (1) og full uttelling (2)). Polytome rubrikker med graderingen 0–2 har to terskler ( $\tau$ ) i stedet for én vanskegradparameter ( $b$ ), hvor tersklene tilsvarer vanskegraden til poengene 1 og 2. NOKUT bruker den polytome Rasch-modellen kalt *partial credit model* (PCM; Masters, 1982), som også håndterer rubrikker med graderingen 0–1.

For polytome rubrikker med graderingen 0–2 bør det være et område på det latente trekket hvor 1 poeng er den mest sannsynlige. Figur 4 illustrerer at kandidater med dyktighet mellom  $\tau_1 = 0,4$  og  $\tau_2 = 3,8$  hadde størst sannsynlighet for 1 poeng. Bare svært dyktige kandidater  $\theta > 3,8$  hadde størst sannsynlighet for 2 poeng. Analysene viser at alle polytome rubrikker med graderingen 0–2 har fungert godt på alle eksamenssettene.

Figur 4. Kurvene beskriver sannsynligheten for 0, 1 og 2 poeng (y-aksen) på rubrikk 22 høst 2024<sup>18</sup> som funksjon av dyktighet (x-aksen). Poengene fungerte godt, fordi noen dyktigheter har 0, 1 eller 2 poeng som sin mest sannsynlige poengverdi. De vertikale stiplede strekene angir vanskegraden til poeng 1 (0,4) og poeng 2 (3,9). På rubrikk 22 fikk 52 % av kandidatene 1 poeng og 5 % fikk 2 poeng.



### 4.3 Rubrikkenes/poengenes vanskegrad og poenggrensen for bestått

Ved sensureringen av eksamen høsten 2024 ble det brukt 58 ulike rubrikker, hvorav 13 var polytome med graderingen 0–2. Maksimal oppnåelig poengsum var dermed 71 poeng, og poenggrensen for bestått ble satt til 37 poeng. Kandidater med 37 poeng hadde dyktigheten  $\theta = 0,46$ , som betyr at de «typisk» ikke oppfylte vurderingskriteriene til dikotome rubrikker med vanskegraden  $b > 0,46$  eller «kunne» det som krevdes for å oppnå 1 av 2 poeng på polytome rubrikker med  $\tau_1 > 0,46$ . Tabell 4 gir eksempler på slike rubrikker.

Tabell 4. Eksempel på rubrikker fra høst 2024 med vanskegrad høyere enn dyktigheten til kandidater som ikke bestod eksamen (37 poeng, dyktighet = 0,46). For rubrikk 19 (polytom) har vi gjengitt kriteriet for poeng 1.

|     |                                                                                                                                                                                                                                                                                      |                 |            |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|------------|
| 19. | Drøfter momentene                                                                                                                                                                                                                                                                    | Blir ikke gjort | Blir gjort |
|     | Drøftelsen skal gjøres med tydelig utgangspunkt i Emils situasjon.                                                                                                                                                                                                                   | 0 poeng         | 1 poeng    |
| 49. | Bruker relevant faktum i drøftelsen                                                                                                                                                                                                                                                  | Blir ikke gjort | Blir gjort |
|     | Opplysninger i faktum må brukes til å underbygge vurderingen av om relevante vilkår er oppfylt. Det gis ikke poeng for ren oppsummering av faktum. Siden samme faktum er relevant i flere sammenhenger, kan kandidatene vise til tidligere drøftelse. Det skal ikke trekkes for det. | 0 poeng         | 1 poeng    |
| 50. | Tolker «barnets beste» og «kan»                                                                                                                                                                                                                                                      | Blir ikke gjort | Blir gjort |
|     | Kandidatene må få frem at «barnets beste» viser til det konkrete barnet. «Kan» viser til at det skal foretas en konkret skjønnsutøvelse                                                                                                                                              | 0 poeng         | 1 poeng    |

<sup>18</sup> «Bruker relevant faktum i drøftelsen av om det vil være skånsomt å gjennomføre hjemmebesøk.»

Eksamensgruppen mottok informasjon om hvilke vurderingskriterier kandidater som akkurat bestod «typisk ikke oppfylte», slik at eksamensgruppen kunne vurdere om beståttgrensen var satt tilstrekkelig høyt. Denne kunnskapen gir grunnlag for å sette riktige poenggrenser på eksamenene.

## **4.4 Oppsummering av resultater fra IRT-analyser utført over tid**

Samlet viser IRT-analysene for høsten 2023, våren 2024 og høsten 2024 at vurderingsrubrikkene fungerer godt. De tilhørende vurderingskriteriene, og dermed også poengene som deles ut, har stor variasjon i vanskegrad. Eksamenen gir god informasjon både om hva faglig dyktige kandidater typisk «kan» og hva faglig svakere kandidater typisk «ikke kan». Det høye antallet vurderingsrubrikker resulterer i at mange poeng kan deles ut. Dette gir mye informasjon om kandidatenes kunnskapsnivå og dermed lavere standard målefeil. Eksamenen har høy reliabilitet og gir pålitelig informasjon om kandidater nær området for grensen mellom ikke-bestått og bestått.

Eksamenssettene fremstår statistisk sett ikke som vanskelige relativt til kandidatenes dyktighet. Vi ser samtidig at flere rubrikker er statistisk avhengige. Hvis kandidaten oppnår poeng på en rubrikk, har vedkommende større sannsynlighet for å oppnå poeng på en annen og avhengig rubrikk. Dette kan fremstå som overflødige rubrikker i test-teorien, men må ses i en faglig kontekst der juridisk metode følger en bestemt prosess med flere steg.

## 5 Reliabilitetsanalyser og sensorsamsvar

For at både studentene på barnevernsfaglige masterprogrammer og fremtidige arbeidsgivere skal ha tillit til resultatet på den nasjonale eksamen, er det viktig at sensuren er til å stole på eller sagt med et annet ord: *reliabel*. En måte å undersøke sensurens reliabilitet på er ved å se på hvorvidt sensorene er enige i poenggivningen av en kandidats besvarelse.

For å undersøke dette har vi gjort såkalt *interrater reliability analysis*, som kan oversettes til *reliabilitetsanalyser mellom sensorer*. Her ser vi på hvor sannsynlig det er at to eller flere sensorer er samstemte i poenggivningen når de vurderer samme kandidat. Vi regner ut en reliabilitetskoeffisient som kan si noe om graden av samsvar mellom to sensorer. En høy reliabilitetskoeffisient indikerer en stabil sensurprosess og at poengsettingen er riktig (reliabel). I mer tekniske termer betyr høy reliabilitet at målingen er reproduserbar over tid, i ulike situasjoner og av ulike sensorer (Zapf et al. 2016).

Til våre analyser benytter vi oss av Krippendorffs alpha ( $\alpha$ ) utviklet av den tyske samfunnsvitenskapelige metodologen Klaus Krippendorff. Krippendorffs alpha tar først og fremst høyde for sannsynligheten for tilfeldig samsvar mellom sensorene – en viktig forutsetning for en god reliabilitetsmodell. Høyt prosentvist samsvar indikerer ikke nødvendigvis høy reliabilitet, siden mye av samsvaret kan skyldes tilfeldigheter, særlig hvis sensorene har få kategorier å velge mellom. I tillegg regnes Krippendorffs alpha for å være svært fleksibel på grunn av dens evne til å håndtere komplekse datastrukturer med ulike datakilder, herunder manglende verdier, flere enn to sensorer og ulike poengsettinger (Krippendorff 2013; Zapf et al. 2016; Shabankhani et al. 2020; Marzi et al. 2023).

### 5.1 Datagrunnlag

Datagrunnlaget består av poengforslag fra begge sensorer på totalt 911 besvarelser i ordinær sensur fordelt på fem gjennomføringer (se tabell 5). Alle besvarelser rettes av to sensorer satt sammen fra et sensorkorps bestående av flere sensorer, hvor antallet avhenger av antall besvarelser. På vårsemesteret er det, som nevnt, konteeksamen for nesten alle institusjoner, og antall kandidater er derfor langt lavere enn ved høstsemesteret. Første gjennomføring, høst 2022, var rundt halvparten av institusjonene med, noe som forklarer det lave antallet besvarelser her. Høst 2023 og høst 2024 er gjennomføringene med størst datagrunnlag.

Poengforslagene følger vurderingsrubrikkene som er utviklet til hver gjennomføring. Antall rubrikker avhenger av faggruppens vurdering av eksamensoppgavens omfang og kompleksitet, og det varierer fra 48 rubrikker høst 2022 til 58 rubrikker de to siste gjennomføringene. Rubrikkene har forhåndsdefinert poengskala utfra kompetansenivået i rubrikken. De fleste rubrikkene på tvers av gjennomføringer bruker nominell (dikotom) poengsetting (0 og 1 poeng), mens noen rubrikker bruker ordinal (rangert) poengsetting (0, 1 eller 2 poeng).

Tabell 5. Oversikt over datamaterialet.

| Gjennomføring | Besvarelser i ordinær sensur | Sensorer i ordinær sensur | Antall rubrikker | Antall poeng |
|---------------|------------------------------|---------------------------|------------------|--------------|
| Høst 2022     | 141*                         | 16                        | 48               | 60           |
| Vår 2023      | 70                           | 12                        | 53               | 65           |
| Høst 2023     | 300                          | 31                        | 55               | 62           |
| Vår 2024      | 133                          | 16                        | 58               | 70           |
| Høst 2024     | 267                          | 30                        | 58               | 71           |
| Totalt**      | 911                          | 105                       | 272              | 328          |

\*165 besvarelser totalt, men 17 besvarelser utgår på grunn av manglende poengforslag fra en av sensorene.

\*\*Totalsum tar ikke høyde for at kandidatene og sensorene kan være de samme personene på tvers av gjennomføringer, men har til hensikt å gi oversikt over størrelsen på datastrukturen sett under ett.

I tabell 6 ser vi hvordan datastrukturen ser ut for hver vurderingsrubrikk, i dette tilfellet for en ordinal rubrikk, for hver gjennomføring. Kolonnene representerer hver sensor, og radene representerer hver kandidat. Siden det kun er to sensorer som retter en kandidat, vil mange celler mangle verdier (NA = not applicable/assessed). I tillegg har vi analysert de summerte poengforslagene per kandidat (total poengsum), og datastrukturen vil dermed minne om tabellen under bare at verdiene består av eksempelvis 32 og 34 istedenfor. Disse summerte poengforslagene (her kalt total poengsum) behandles som «ratio»-variabler i analysen.

Tabell 6. Datasett for vurderingsrubrikk 1m\* fra høst 2023 for å eksemplifisere strukturen i datasettet.

| Vurderingsrubrikk 1m, høst 2023 | Sensor 1 | Sensor 2 | Sensor 3 | Sensor 4 | Sensor ... 32 |
|---------------------------------|----------|----------|----------|----------|---------------|
| Kandidat 1                      | 1        | 1        | NA       | NA       | NA            |
| Kandidat 2                      | 0        | NA       | 0        | NA       | NA            |
| Kandidat 3                      | 1        | NA       | NA       | 1        | NA            |
| Kandidat 4                      | NA       | 2        | 2        | NA       | NA            |
| Kandidat 5                      | NA       | 1        | NA       | 2        | NA            |
| Kandidat 6                      | NA       | NA       | 2        | 2        | NA            |
| Kandidat 7                      | NA       | NA       | NA       | NA       | NA            |
| Kandidat 8                      | NA       | NA       | NA       | NA       | NA            |
| Kandidat 9                      | NA       | NA       | NA       | NA       | NA            |
| Kandidat ... 300                | NA       | NA       | NA       | NA       | NA            |

\*Vurderingsrubrikk 1m: «Tolker ordlyden i bvl. § 6-2 første ledd bokstav c)». 0 poeng: «Blir ikke gjort». 1 poeng: «Tolker ordlyden». 2 poeng: «God tolkning».

## 5.2 Tematisk inndeling

Som tidligere nevnt er eksamensoppgaven en caseoppgave der studentene skal bruke juridisk metode for å belyse casets rettslige problemstillinger. Caseoppgaven skal teste kandidatens kunnskap utfra emnets læringsmål. Læringsmålene er knyttet til juridisk metode. Ved å kategorisere vurderingsrubrikkene etter de ulike stegene i juridisk metode kan vi analysere sensorenes samsvar tematisk slik at vi kan si enda mer om hva som skyldes variasjonen i samsvar i vurderingsrubrikkene.

Basert på læringsmålene og undervisningspraksis, samt ordlyd i vurderingsrubrikkene, har vi kategorisert rubrikkene inn i åtte temaer. I tabell 7 kan vi se hvordan rubrikkene fordeler seg. De fem første temaene reflekterer stegene i juridisk metode og er av mer praktisk karakter tilknyttet eksamensoppgavens case (f.eks. rubrikken fra høst 2022 «Bruker/tolker hva som ligger i «tiltak»). De fleste rubrikkene inngår i disse temaene. I tillegg skal sensorene vurdere besvarelsenes helhet, som de tre siste temaene på ulike måter reflekterer (f.eks. rubrikken «Språkføring generelt» fra vår 2023). Her ser vi færre rubrikker i hvert tema, og ved noen gjennomføringer ble temaet ikke identifisert. Vi velger likevel å skille disse temaene fra hverandre for å få frem nyanser i helhetsvurderingen.

Tabell 7. Tematisk inndeling av antall rubrikker per gjennomføring.

| Tematikker                                                                                   | Høst 2022 | Vår 2023 | Høst 2023 | Vår 2024 | Høst 2024 |
|----------------------------------------------------------------------------------------------|-----------|----------|-----------|----------|-----------|
| 1. Identifisere og formulere rettsspørsmål (og underspørsmål)                                | 7         | 3        | 9         | 11       | 12        |
| 2. Finne relevante rettskilder                                                               | 6         | 3        | 3         | 4        | 10        |
| 3. Klarlegge juss (tolke vilkår)                                                             | 13        | 19       | 11        | 13       | 12        |
| 4. Sammenlikne juss og faktum (subsumere og drøfte)                                          | 7         | 10       | 12        | 7        | 9         |
| 5. Konkludere (underproblemstilling og hovedproblemstillinger)                               | 7         | 11       | 8         | 11       | 10        |
| 6. Klarlegger et enkelt rettsspørsmål som skal fastslås                                      | 1         | 0        | 2         | 1        | 1         |
| 7. Overordnede rettsprinsipper (trekker inn barnets beste, medvirkning, minste inngrep osv.) | 0         | 2        | 5         | 6        | 2         |
| 8. Totalvurdering (utviser god juridisk metode og godt språk)                                | 7         | 5        | 5         | 5        | 2         |

### 5.3 Analysemetode

Vi brukte R-pakken «krippendorffsalpha» (Hughes 2022) i utregningen av Krippendorffs alpha med tilhørende konfidensintervaller. Analysen ble gjort per vurderingsrubrikk og total poengsum på alle gjennomføringer. Vi skilte mellom nominelle og ordinale rubrikker i analysen, og den totale poengsummen ble behandlet som «ratio». I bearbeidingen av datafilene brukte vi Microsoft Excel. Kategoriseringen per rubrikk ble også gjort i Excel, mens visualiseringen foregikk i R med R-pakkene «dplyr» og «ggplot2» (Wickham 2023; 2024).

For å vite hva som er et godt og mindre godt samsvar mellom sensorene, kan vi se til Krippendorffs (2019:356) beskrivelse av alphaverdiene:

- Alpha = 1: Indikerer perfekt samsvar mellom sensorer. Her har alle sensorer gitt akkurat samme poeng for hvert evaluerte kriterium.
- Alpha  $\geq 0,80$ : Anses som tilfredsstillende samsvar og indikerer en reliabel sensur. En alpha lik 0,80 eller over, er i mange forskningssammenhenger en akseptabel alpha å trekke konklusjoner ut fra basert på datagrunnlaget.
- Alpha [0,67–0,79]: Anses som det lavere sjiktet for å trekke tentative konklusjoner. Indikerer moderat samsvar, men resultatene må tolkes med forsiktighet, og roten til spredt sensur bør undersøkes.
- Alpha < 0,67: Indikerer et svakt samsvar mellom sensorer. En alpha under 0,67 anses ofte som ikke-reliabel.
- Alpha = 0: Indikerer null samsvar mellom sensorer annet enn det som kan forventes av tilfeldig samsvar.



- $\alpha < 0$ : En negativ alpha indikerer systematisk uenighet mellom sensorer. Kan forekomme der sensorene trekker systematisk i motsatt retning i sensuren.

Hvis vi får en alpha på 1, har vi med andre ord perfekt samsvar mellom sensorene, mens en alpha under 0,67 tilsier at vi har lav reliabilitet. En alpha på 1 tilsier at sensorene er helt enige på hvert punkt. En alpha på 0 tilsier at sensorenes sensurering er helt vilkårlig, og at de like gjerne kunne satt poeng tilfeldig uten en sensorveiledning eller en forståelse for innholdet i vurderingen (Marzi et al. 2023). Konfidensintervallene forteller oss sannsynligheten for at reliabiliteten ligger i et gitt område (her: med 95% sannsynlighet). Jo smalere konfidensintervall, desto sikrere kan vi være på at alphakoeffisienten stemmer. Til Krippendorffs alpha anbefales bootstrap-metoden for å regne ut konfidensintervaller (Zapf et al. 2016), og det er den vi bruker her.

I resultatene viser vi alphakoeffisienten til den totale poengsummen per gjennomføring samt variasjonen i alphakoeffisientene til vurderingsrubrikkene per gjennomføring. Til slutt viser vi gjennomsnittlig alphakoeffisient per tema. For faggruppen som utvikler den nasjonale eksamenen, vil variasjonen i vurderingsrubrikkene være nyttig for å se hvilke rubrikker sensorene har mindre samsvar på enn andre. For studenten derimot, vil den totale poengsummen være mest interessant i og med at denne setter karakteren. Et viktig poeng i denne sammenhengen er riktignok at det ikke er sensorenes individuelle sensur som setter karakteren, men deres felles sensurering (samsensur) til slutt. Det er likevel den totale poengsummen som er det nærmeste vi kommer karakterrelevant score i disse analysene av reliabiliteten i ordinær sensur.

### 5.3.1 Samsensur i klagesensur

Vi får anledning til å teste reliabiliteten i den felles sensureringen på en liten gruppe kandidater gjennom klagesensur. Kandidater som klager får besvarelsen sin vurdert av to uavhengige kommisjoner som begge foretar en samsensur først i ordinær sensur og deretter i samsensur. Disse kommisjonene kjenner ikke til hverandres vurderinger, det er altså blind sensur. Utvalget blir derimot noe skjevt i disse analysene siden det kun er den svakere delen av kandidatene som inkluderes. I tillegg er konteksten til de to parene noe ulik: Klagesensorene antar at kandidaten allerede er blitt vurdert til ikke bestått, siden karakterskalaen i denne eksamen er dikotom. En kandidat som har bestått, har ikke motiv for klage på resultatet. Slik sett er ikke klagesensur fullstendig blind ved en slik eksamen. Samtidig vet ikke sensorene hvilken poengsum kandidaten har fått tidligere og kan dermed ikke vite noe om hvor langt unna beståttgrensen kandidaten har vært.

Til disse analysene har vi valgt å kun bruke Krippendorffs alpha til analyser av total poengsum på tvers av gjennomføringer. For å gå dypere, men ikke gjøre oss avhengig av en modell som fordrer lik kontekst for dataene, har vi valgt å gjøre en enkel numerisk analyse av vurderingsrubrikkene på de to siste ordinære gjennomføringene høst 2023 og høst 2024.

Hvis vi ser på kandidatene samlet, får vi et mål på om klagesensur og ordinær sensur har slått ulikt ut. Summen av poeng som er gitt på hver rubrikk i klagesensuren, dvs. vi adderer alle poeng på en og samme rubrikk på tvers av kandidatene, holdes sammen med summen av poengene hos de samme kandidatene i ordinær sensur. Da får vi en slik tabell:

Tabell 8. Utdrag av resultater på de seks første rubrikkene i gjennomføring høst 2024.

| Rubrikk | Rubrikksum klage | Rubrikksum ordinær | Variasjon |
|---------|------------------|--------------------|-----------|
| 1       | 66               | 67                 | 0,99      |
| 2       | 66               | 63                 | 1,05      |
| 3       | 65               | 66                 | 0,98      |
| 4       | 33               | 37                 | 0,89      |
| 5       | 22               | 17                 | 1,29      |
| 6       | 11               | 11                 | 1,00      |

Perfekt samsvar er en variasjon på 1,00, et høyere tall er «mildere» sensur i klage og et lavere tall enn 1 indikerer at klagesensuren har vært strengere.

I det følgende presenteres først resultater fra reliabilitetsanalysene av ordinær sensur, og deretter resultatene av samsensur i ordinær- og klagesensur.

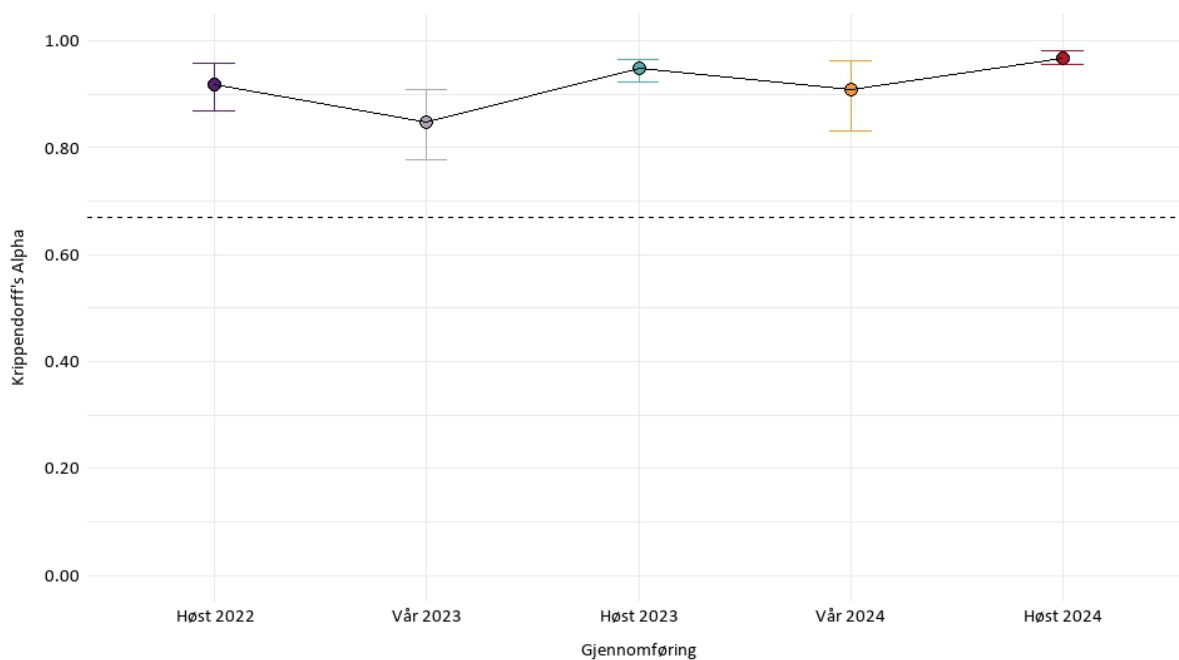
## 5.4 Resultater

Figur 5 viser alphakoeffisienten til total poengsum på tvers av gjennomføringer, og det er denne summen som er mest eksamensrelevant for karakter i individuell vurdering. Utfra figuren ser vi en svært god sensurreliabilitet på nasjonal deleksamen i rettsanvendelse. Vi ser likevel noe variasjon.

Høst 2024 og høst 2023 skiller seg positivt ut med nærmest perfekt samsvar mellom sensorene (hhv.  $\alpha = 0,97$  og  $\alpha = 0,95$ ). I tillegg ser vi at konfidensintervallene til disse to gjennomføringene er smalere enn resterende gjennomføringer, noe som indikerer stor sannsynlighet for at estimatet vårt treffer godt.

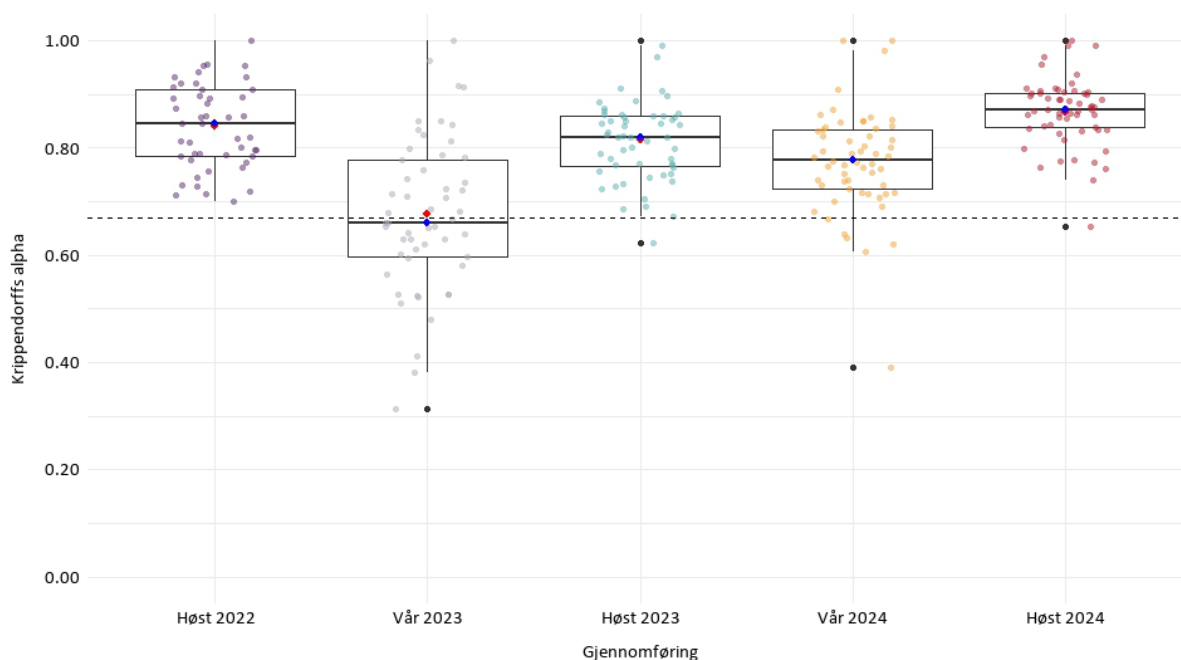
Vår 2023 og vår 2024 skiller seg noe negativt ut med lavere alphaverdi og bredere konfidensintervaller, noe som antyder at estimatene er noe mer usikre. Dette kan forklares av få kandidater, og dermed mer usikker variasjon i dataene, og at de fleste av disse kandidatene tok eksamen på nytt etter å ha fått ikke bestått høsten før. Sistnevnte poeng bør ikke nødvendigvis ha noe å si for at sensorene ikke er enige på poeng, men en mulig måte å tolke resultatet på er at besvarelsene til svake kandidater er mer tvetydige enn andres og dermed vanskeligere å sensurere.

Figur 5. Krippendorffs alpha med tilhørende konfidensintervaller av total poengsum mellom sensor 1 og sensor 2 i ordinær sensur på tvers av gjennomføringer. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensur.



Den totale poengsummen vil imidlertid, og naturlig nok, gi oss et utjevnet bilde av sensorsamsvaret blant sensorene. Hvis vi derimot ser på sensorsamsvaret per vurderingsrubrikk, får vi et mer nyansert bilde. I figur 6 visualiseres alphakoeffisientene til rubrikkene (uten konfidensintervaller), der enkelte rubrikker viser en del variasjon mellom sensorene. Dersom vi ser på konfidensintervallene til hver rubrikk (se figur i vedlegg 13.1), ser vi enda større variasjon.

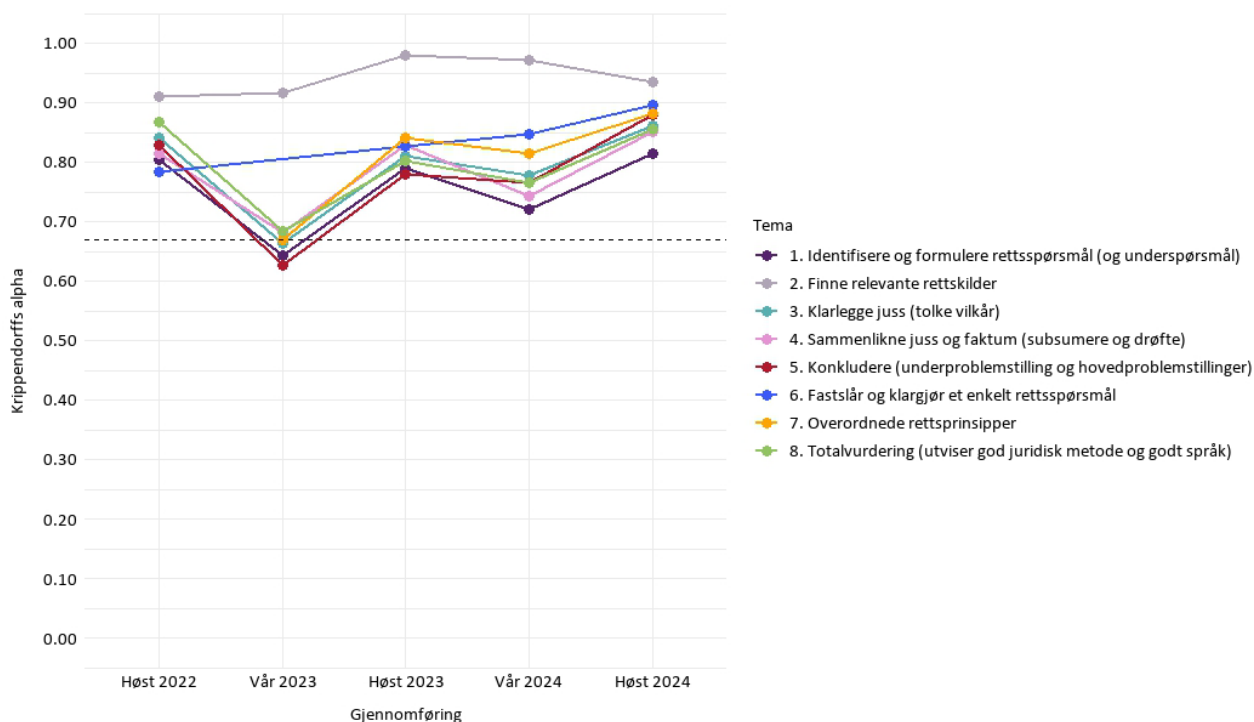
Figur 6. Boksplott av Krippendorffs alpha per vurderingsrubrikk på tvers av gjennomføringer. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensor. Blå prikk illustrerer median, mens rød prikk illustrerer gjennomsnitt.



På tvers av rubrikker ser vi at høstgjennomføringene har høy grad av sensorsamsvar. Flere rubrikker har lavere sensorreliabilitet i vårsemesteret. Vår 2023 har eksempelvis lavest samsvar med  $\alpha=0,31$  på vurderingsrubrikken 3c «Klarlegger utgangspunktet at barn under 15 år ikke har partsrettigheter» (med konfidensintervall  $-0,18-0,71$ ), og samtidig høyest mulig samsvar på vurderingsrubrikk 1b «Finner frem til relevant lovbestemmelse Bvl. 2021 § 5-7 første ledd» (med konfidensintervall  $1-1$ ). Vi ser ikke like store sprik i samsvar blant rubrikkene i høstsemesteret, men estimatene er usikre på enkelte rubrikker. For eksempel i høst 2024 hvor vi stort sett ser høyt samsvar og smale konfidensintervaller, har rubrikk 1 «Identifiserer og formulerer problemstillingen» en  $\alpha=0,80$  med et konfidensintervall fra  $0,39$  til  $1$ . Variasjonen antyder at vi ikke kan være helt sikre på reliabiliteten på hver rubrikk, særlig på vårgjennomføringene. Samtidig ser vi overordnet at godt over 80 prosent av rubrikkene har et samsvar å anse som moderat (over  $0,67$ ) og 60 prosent av konfidensintervallenes nedre grense er over  $0,67$ .

For å forstå enda mer av variasjonen i reliabiliteten på tvers av rubrikker ser vi på den tematiske inndelingen av rubrikkene. I figur 7 ser vi en oversikt over gjennomsnittlig alpha-verdier på de ulike temaene på tvers av år. Vi ser tilsvarende «dupper» ved våreksamen som vi har sett tidligere på alle temaene, hvor fire av temaene i vår 2023 får en alpha under  $0,67$  – grensen for et moderat samsvar mellom sensorene. Dette vitner om at det ikke er målt et reliabelt sensorsamsvar, enten at sensorene sensurerer vilkårlig på disse rubrikkene (vanskelig å sensurere) eller at det er lite data.

Figur 7. Gjennomsnittlig alpha per tema på tvers av gjennomføringer. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensor.

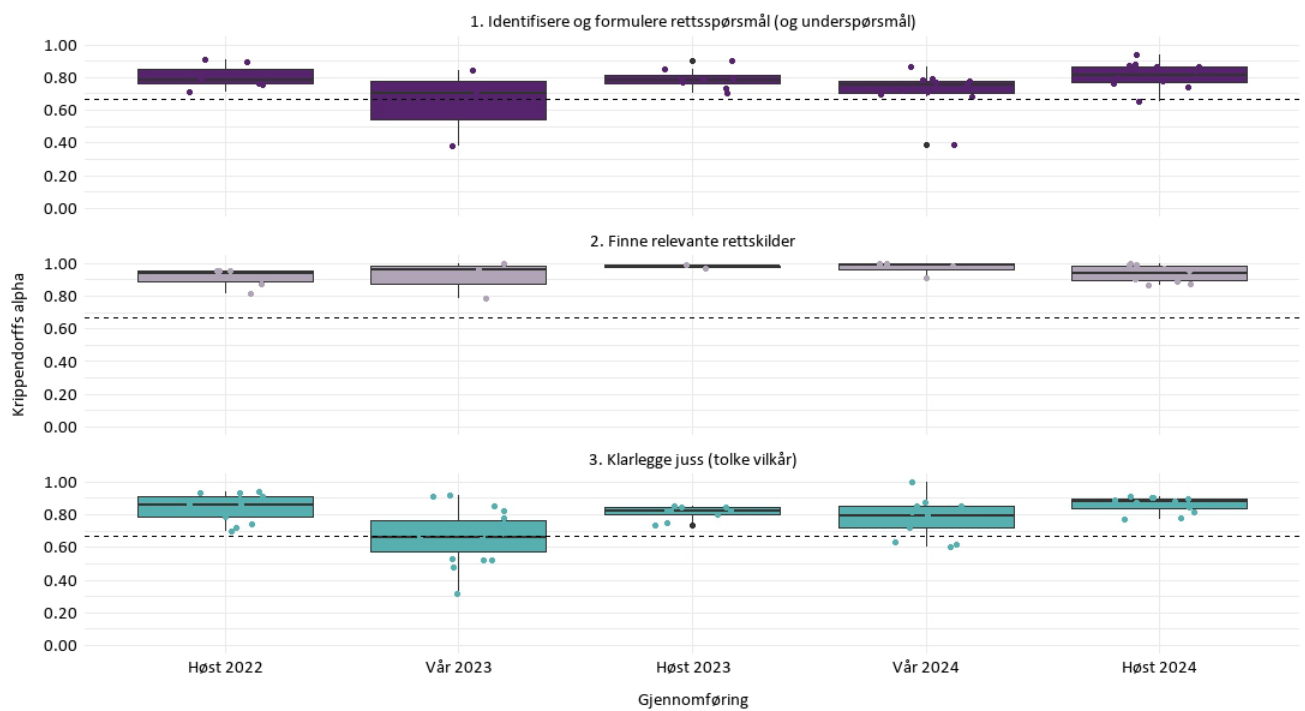


Overordnet ser vi at «2. Finne relevante rettskilder» har en svært høy reliabilitet på tvers av år, med høyest reliabilitet høst 2023. Her kan vi antyde at det er relativt rett frem å vurdere kandidatens evne til å finne relevante rettskilder sammenlignet med for eksempel deres evne til å identifisere rettsspørsmål og konkludere. Det kan tolkes som at hvorvidt en kandidat har funnet en relevant rettskilde, er en mer konkret vurdering (enten er rettskilden funnet eller ikke), mens identifisering og formulering av rettsspørsmål og konkludering kan være en mer skjønsmessig vurdering. Samtidig ser vi at selv tematikker man kunne tenkt var skjønsmessige vurderinger, og dermed mer utsatt for uenighet blant sensorer, har nokså god til moderat reliabilitet.

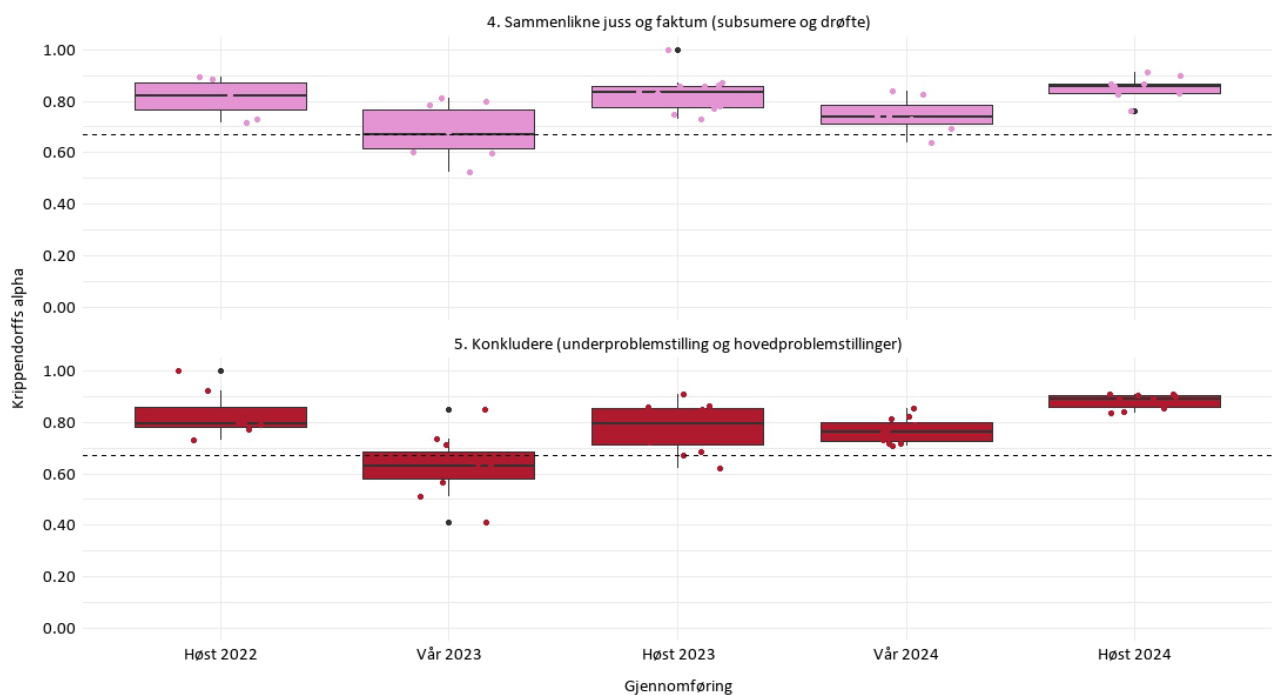
I figur 8–10 fremstilles variasjonen i reliabiliteten innad i hvert tema på tvers av år. Her ser vi ikke stor variasjon innad i temaene, igjen med unntak av vår 2023 som har relativt stor variasjon i alle temaene. Vi ser imidlertid også at konkluderings temaet (tema 5) har noe variert reliabilitetsnivå på samtlige gjennomføringer, særlig høst 2023. Dette antyder at konkludering er en vanskelig og uklar kategori å oppnå godt sensorsamsvar på.

Flesteparten av tematikkene befinner seg likevel rundt en alpha fra 0,80 og indikerer dermed en reliabel sensor. Avhengig av hvilket tema det er snakk om, er det med andre ord sannsynlig at sensorene kommer frem til samme konklusjon. Vi ser altså en tendens til at det er lite sensoravhengighet på tvers av temaer. Dersom en kandidat er sterkere på et tema enn et annet, er det ikke en stor fare for at de taper på dette.

Figur 8. Boksplott av Krippendorffs alpha av temaene 1–3. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensur.



Figur 9. Boksplott av Krippendorffs alpha for temaene 4 og 5. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensur.



Figur 10. Boksplott av Krippendorffs alpha for temaene 6–8. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensor.



#### 5.4.1 Reliabilitet i samsensur mellom klage og ordinær

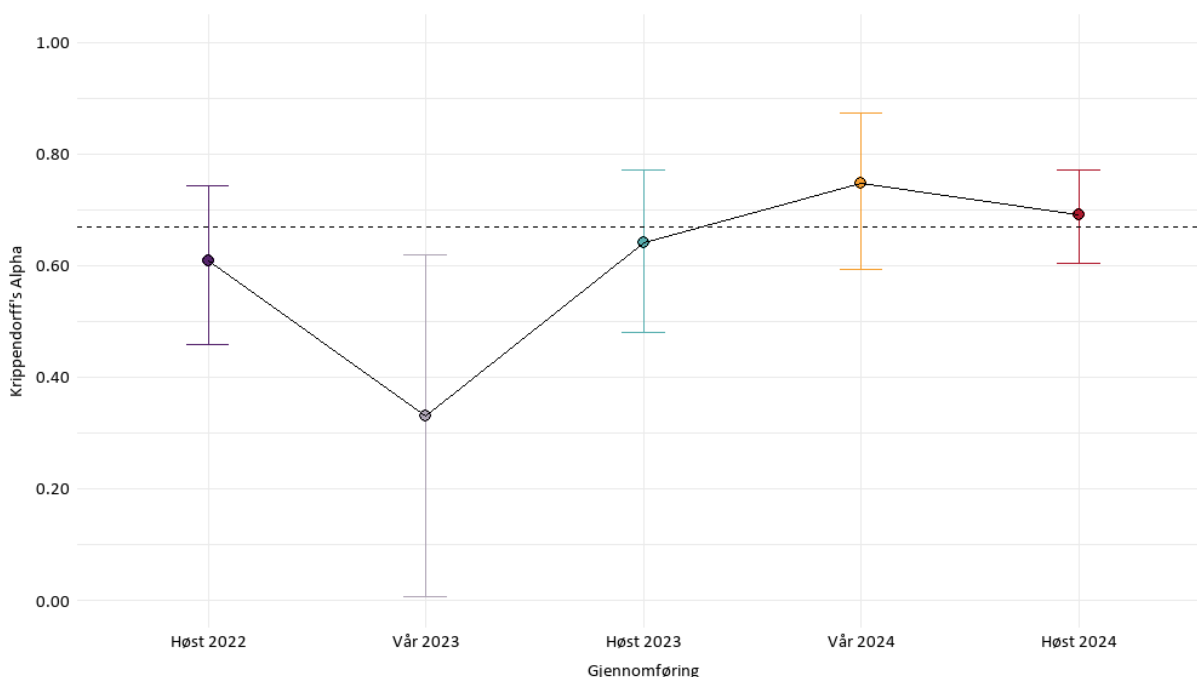
Som nevnt, har vi også undersøkt samsensur mellom sensorpar i ordinær- og klagesensur. I tabell 9 ser vi antall klager og karakterendringer i hver gjennomføring. Vi ser relativt få karakterendringer, noe som kan enten skyldes at de to sensorparene er samstemte om kandidatens endelige karakter eller at studentene som klager befinner seg mange poeng unna beståttgrensen.

Tabell 9. Oversikt over datamaterialet i klagesensur.

| Gjennomføring | Antall klager | Andel klager | Antall karakterendringer | Andel karakterendringer (blant klager) | Andel karakterendringer (totalt) |
|---------------|---------------|--------------|--------------------------|----------------------------------------|----------------------------------|
| Høst 2022     | 58            | 35 %         | 14                       | 24 %                                   | 8 %                              |
| Vår 2023      | 34            | 49 %         | 4                        | 12 %                                   | 6 %                              |
| Høst 2023     | 84            | 28 %         | 11                       | 13 %                                   | 4 %                              |
| Vår 2024      | 43            | 32 %         | 9                        | 21 %                                   | 7 %                              |
| Høst 2024     | 77            | 29 %         | 11                       | 14 %                                   | 4 %                              |

I figur 11 ser vi at sensorene i ordinær sensur og klagesensorene tenderer til å være mindre samstemte enn de to sensorene i ordinær sensur. Kun vår 2024 og høst 2024 indikerer reliabel sensur, men likevel svakere enn i ordinær sensur. I vår 2023 ser det ut til at sensorene hadde lavt samsvar i ordinær- og klagesensur. Her er det imidlertid stor usikkerhet knyttet til estimatet, noe det lave antallet klager kan forklare.

Figur 11. Krippendorffs alpha med tilhørende konfidensintervaller av total poengsum mellom samsensur i ordinær- og klagesensur på tvers av gjennomføringer. Stiplet linje illustrerer  $\alpha = 0,67$ , som regnes som grensen for reliabel sensur.



I de numeriske analysene har vi kun sett på høst 2023 og høst 2024. Her ser vi et snitt over rubrikkene på 0,97 i høst 2023 og 1,05 i høst 2024. På rubriknivå observerer vi at klagesensuren var noe strengere høst 2023 og noe mildere høst 2024, men begge tallene ligger svært nær perfekt samsvar, noe som gir tillit til at klagesensur og ordinær sensur faktisk opererer etter samme kriterier.

Hvis vi ser på enkeltrubrikkene, kan vi observere samme mønster. Jevnt over er det lite variasjon, men vi ser også noen få rubrikker hvor variasjonen er noe større. Gitt det begrensede antallet kandidater som ber om ny sensur ved hver eksamen, kan en rubrikk som slik viser seg utfordrende å sensurere, gi tydelig utslag i oversikten. Her er ikke plass til å vise full oversikt over rubrikkene (hhv. 55 og 58 i antall), men snittvariasjon kategorisert etter tematikk gir en slik oversikt:

Tabell 10. Samsensur i ordinær- og klagesensur på tvers av tematikker.

| Tematikk:                                                                                    | Høst 2023 | Høst 2024 |
|----------------------------------------------------------------------------------------------|-----------|-----------|
| 1. Identifisere og formulere rettsspørsmål (og underspørsmål)                                | 0,95      | 1,10      |
| 2. Finne relevante rettskilder                                                               | 0,99      | 1,07      |
| 3. Klarlegge juss (tolke vilkår)                                                             | 0,96      | 1,09      |
| 4. Sammenlikne juss og faktum (subsumere og drøfte)                                          | 0,90      | 1,10      |
| 5. Konkludere (underproblemstilling og hovedproblemstillinger)                               | 1,00      | 1,16      |
| 6. Fastslår og klargjør et enkelt rettsspørsmål                                              | 1,04      | 1,11      |
| 7. Overordnede rettsprinsipper (trekker inn barnets beste, medvirkning, minste inngrep osv.) | 1,04      | 1,01      |
| 8. Totalvurdering (utviser god juridisk metode og godt språk)                                | 0,75      | 1,21      |



Det moderate utslaget på kategori 5 (høst 2024) skyldes spesielt to enkeltribrikker (begge har ordlyden «konkluderer i samsvar med drøftelsen»), men samtidig finner vi ikke igjen denne utfordringen i reliabilitetsanalysen nevnt ovenfor. Det som synes vanskeligst å vurdere er det siste temaet *overordnede rettsprinsipper*. Man kan også tenke seg at utvalget av besvarelser i en analyse av klagesensur inneholder en større andel av litt vage tekster, som da åpner et større tolkningsrom for sensorene.

## 5.5 Oppsummering

Basert på reliabilitetsanalysene av samsvaret mellom sensorene på tvers av gjennomføringer kan vi si at den nasjonale eksamen i rettsanvendelse fremholder høy sensurreliabilitet i ordinær sensur. Vi ser at høstsemestrene, hvorav de fleste institusjonene har ordinær gjennomføring, har høyest reliabilitet. Samtidig ser vi noe variasjon i analysene på tvers av gjennomføringer og tematikker, hvorav særlig vårsemestrene og tematikkene som omhandler identifisering av problemstillinger og konkludering viser svakere reliabilitet. I tillegg ser vi at sensorparene i ordinær- og klagesensur er mindre samstemte enn et sensorpar i ordinær sensur på studentenes totale poengsum.

Vår 2023 skiller seg ut gjennomgående i analysene. Dette var andre gjennomføring av eksamen og inkluderte få studenter. De brede konfidensintervallene kan skyldes nettopp det lave antallet besvarelser i analysen, men estimatet er også under grensen for hva som kan regnes som et godt reliabilitetsestimat.

Det første temaet, identifisering av problemstillinger, er interessant å se i lys av sensorkorpsets diskusjoner i begynnelsen. I kalibreringsmøtene etter de første gjennomføringene ble det nemlig diskutert hvorvidt problemstillinger og underproblemstillinger måtte formuleres eksplisitt eller om det holdt å behandle problemstillingene. Dette er et eksempel på at formell og uformell kalibrering har gitt en mer enhetlig praksis etter hvert. Høst 2024 viser høyere reliabilitet på dette temaet sammenlignet med foregående gjennomføringer. Tilsvarende utvikling ser vi for øvrig i flere av figurene – sensurreliabiliteten er stort sett høyest høst 2024. Det gir gode forutsetninger for å jobbe enda mer med sensorsamsvar fremover.

## 6 Sensorenes vurdering av eksamenene, vurderingsrubrikker og vurderingsprosessen

I det forutgående kapittelet i denne rapporten er det blitt gitt et innblikk i sensurreliabilitet gjennom sensorsamsvar på poenggivning i sensurprosess. En annen viktig faktor ved påliteligheten i vurderingen (reliabilitet) er hvordan sensorene gjør forståelse av eksamen, sensorveiledningen og vurderingsrubrikkene.

Ifølge Bjølseth et al. (2011) bør vurderinger baseres på faglig skjønn, som utvikles gradvis gjennom deltakelse i faglige og profesjonelle fellesskap. Vurderingskriterier og rubrikker i seg selv fører ikke nødvendigvis til mer pålitelige vurderinger. En pålitelig vurdering forutsetter at sensorene både anerkjenner dem og faktisk bruker dem i praksis. Det er dermed viktig at sensorene har en felles forståelse av hva som vurderes og hvordan slik at vurderingene blir mer konsistente.

Som nevnt er sensorene organisert i tverrinstitusjonelle par, og hvert sensorpar får tildelt et sett med besvarelser som skal vurderes. Hver sensor får tilsendt sensorveiledningen, inkludert vurderingsrubrikker, samt et gitt antall besvarelser. Sensorene er også organisert i grupper som sensurerer sammen. Vurderingsprosessen består av flere trinn: først en individuell vurdering der hver sensor leser og vurderer besvarelsene uavhengig. I denne prosessen deltar sensorene på et innledende sensormøte med de andre sensorene hvor utfordringer i sensurering kan løftes og diskuteres. Deretter foretar sensorparene en samsensur hvor sensorparet diskuterer seg frem til en felles karakter og poengsum. Det er samsensuren som danner grunnlaget for studentens endelige karakter og begrunnelse. For å understøtte dette med faglige fellesskap og kalibrering deltar sensorene også kalibreringsmøte med resten av fagmiljøet hvor faglig diskusjoner kan løftes og poengfordeling i vurderingsrubrikkene og grense for bestått-karakter settes.

I dette kapittelet presenteres resultater fra sensorundersøkelser gjennomført i tilknytning til den nasjonale eksamenen for master i barnevern og barnevernsfaglig arbeid. Undersøkelsene gir innsikt i sensorenes vurderinger av eksamensoppgavene, vurderingsrubrikkene og selve vurderingsprosessen. Formålet er å belyse hvordan sensorer opplever kvaliteten i sensurprosessen, gjennom blant annet i hvilken grad vurderingskriteriene legger til rette for rettferdig og konsistent sensur. Kapittelet belyser også hvordan sensorer opplever nytteverdien av vurderingsrubrikkene – både som verktøy i sensurprosessen og som grunnlag for tilbakemelding til studentene – og hvordan deltakelse i sensurarbeidet påvirker deres egen undervisningspraksis.

### 6.1 Datagrunnlag

Etter hver gjennomføring av nasjonal deleksamen sender NOKUT ut en spørreundersøkelse til alle sensorer.<sup>19</sup> Formålet er å innhente erfaringer og vurderinger knyttet til eksamensoppgavene, sensorveiledningen og selve sensurprosessen. Undersøkelsene gir verdifull innsikt i hvordan sensorer opplever gjennomføringen av sensurprosessen og danner en del av kunnskapsgrunnlaget for videre utvikling og forbedring. Undersøkelsene

---

<sup>19</sup> Våren 2023 ble ikke dette gjennomført da dette i all hovedsak var en konteeksamen etter pilotgjennomføringen høst 2022, og et lite utvalg av sensorene deltok på grunn av få kandidater. I tillegg hadde NOKUT tett dialog med alle de som var sensorer slik at vurderingen og evalueringen ble ivarettatt i samtaler.

gjennomføres i spørreskjemaverktøyet SurveyXact.<sup>20</sup> I Tabell 11 ser vi undersøkelsenes start- og sluttidspunkt på tvers av gjennomføringene. Undersøkelsene sendes etter klagenes er ferdigbehandlet slik at sensorene også kan ta stilling til klagesensureringen når de svarer, og de er åpne i omtrent to uker.

Tabell 11. Utsendelsestidspunkt av sensorundersøkelse på tvers av gjennomføringer.

| Gjennomføring | Utsendelsestidspunkt | Svarfrist        |
|---------------|----------------------|------------------|
| Høst 2022     | 6. februar 2023      | 20. februar 2023 |
| Høst 2023     | 2. februar 2024      | 9. februar 2024  |
| Vår 2024      | 30. april 2024       | 5. mai 2024      |
| Høst 2024     | 4. februar 2025      | 18. februar 2025 |

Spørreundersøkelsen består av både lukkede spørsmål (avkrysningsalternativer) og åpne fritekstfelt. Spørsmålsformuleringer har blitt justert noe mellom gjennomføringene (se vedlegg 13.2). Der spørsmålene ikke er tilnærmet like, vil formuleringene enten beskrives i teksten eller gjengis i vedlegget. Spørsmålene i undersøkelsen er stilt med en femdelt Likert-skala, fra «I liten grad» til «I stor grad», som kategorisørsmål og som fritekstspørsmål.

Resultatene presenteres både som figurer og sitater. Figurene viser prosentfordeling og gjennomsnitt i parentes. Når det gjelder fritekstsvarene, er disse gjennomgått og analysert tematisk. Gjentakende temaer er identifiserte og vil bli omtalt i tilknytning til relevante funn. Temaene er 1) beståttgrenser, 2) sensurmøter, 3) vurderingsrubrikk, 4) samsensur og 5) effekter.

### 6.1.1 Utvalget

Sensorkorpset består av fagpersoner fra institusjonene som tilbyr master i barnevern og barnevernsfaglig arbeid samt nemndledere. Dette gir en solid forankring både i akademisk og i profesjonell praksis. Det er 13 institusjoner som tilbyr master i barnevern og master i barnevernsarbeid, og rundt 17 fagpersoner som underviser i juridisk metode og rettsanvendelse fordelt på disse. Nesten alle av disse underviserne er med i sensorkorpset. I tillegg er det noen sensorer som underviser i juss i andre studieprogram på institusjonene, slik som vernepleie. Noen av underviserne har også undervisning på flere institusjoner.

Sensorkorpset, dvs. alle som sensurerte innværende eksamen, er grunnlaget for utvalget i denne undersøkelsen. I Tabell 12 presenteres antall sensorer som mottok undersøkelsen etter hver gjennomføring, og svarprosent.

Tabell 12. Utvalg og svaroversikt per sensorundersøkelse.

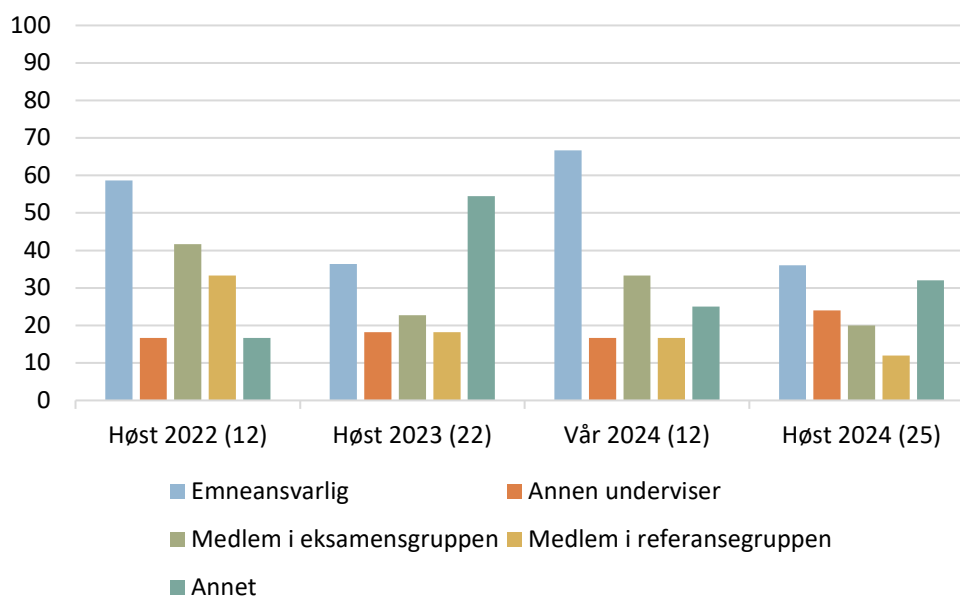
| Gjennomføring | Antall sensorer | Antall svarende | Svarprosent |
|---------------|-----------------|-----------------|-------------|
| Høst 2022     | 14              | 12              | 85,7 %      |
| Høst 2023     | 32              | 22              | 68,8 %      |
| Vår 2024      | 16              | 12              | 75,0 %      |
| Høst 2024     | 30              | 25              | 83,3 %      |

<sup>20</sup> Xact by Ramboll (u.å.).

I snitt er det 77 prosent av sensorene som har besvart undersøkelsen, hvorav 70 prosent gjennomførte hele undersøkelsen, og 8 prosent hadde noen svar. Høyest svarprosent ser vi høstsemesteret 2022 og 2024, og lavest svarprosent ser vi høst 2023 med 68,8 prosent.

I Figur 12 ser vi hvilken faglig rolle sensorene har oppgitt at de hadde. Hver sensor hadde mulighet til å huke av for flere svar om det var relevant for dem. Eksempelvis kan de både være medlem av eksamensgruppen og emneansvarlig.

Figur 12. Prosentfordeling av sensorenes faglige rolle i utvalget. Antall i parentes.



Av sensorene som har huket av for «Annet», har de opplyst om at de enten har hatt rollen som tidligere emneansvarlig, sensor eller ekstern sensor. En del av sensorkorpset er ansatt i praksisfeltet og jobber blant annet som nemndledere. Disse representeres i annet-kategorien.

Høsten 2022 var det 12 respondenter som besvarte undersøkelsen. Det var første gjennomføring av denne eksamen, og store deler av sensorkorpset var med i enten eksamens- eller referansegruppen (75 prosent). Videre oppga 59 prosent at de var emneansvarlig, 17 prosent var underviser i emnet og 17 prosent huket av for «Annet».

For høst 2023 var 41 prosent av 22 respondenter i eksamensgruppen eller referansegruppen. 36 prosent var emneansvarlig, 18 prosent underviste i emnet og 55 prosent oppga «Annet».

Vår 2024 var det færre sensorer igjen, ettersom dette for mange institusjoner er en kotteksamen. Av de 12 respondentene som besvarte undersøkelsen, var 50 prosent av dem med i eksamensgruppen eller referansegruppen. 67 prosent var emneansvarlig, 17 prosent underviste i emnet og 25 prosent huket av for «Annet». I vårsemesterne er det færre kandidater som tar eksamen, og dermed er det behov for færre sensorer. Da prioriteres først og fremst de som underviser i emnet.

Høst 2024 har det største utvalget med 25 respondenter. Av disse var 32 prosent i eksamensgruppen eller referansegruppen. 36 prosent oppga at de var emneansvarlig, og 12

prosent oppga at de underviste i emnet, mens 32 prosent definerte seg med annet-kategorien.

I kapittel 3 ble det vist til at eksamensgruppen består av 4 til 6 fagpersoner og referansegruppen 5 til 6 fagpersoner. Høst 2022 var 10 personer i disse gruppene undervisere fra ulike utdanningsinstitusjoner, og for høst 2024 var de 8 undervisere fra ulike institusjoner. Det at så mange av dem som har besvart undersøkelsen var med i eksamens- eller referansegruppen, har dermed en logisk forklaring med at de utgjør en relativ stor del av fagmiljøet.

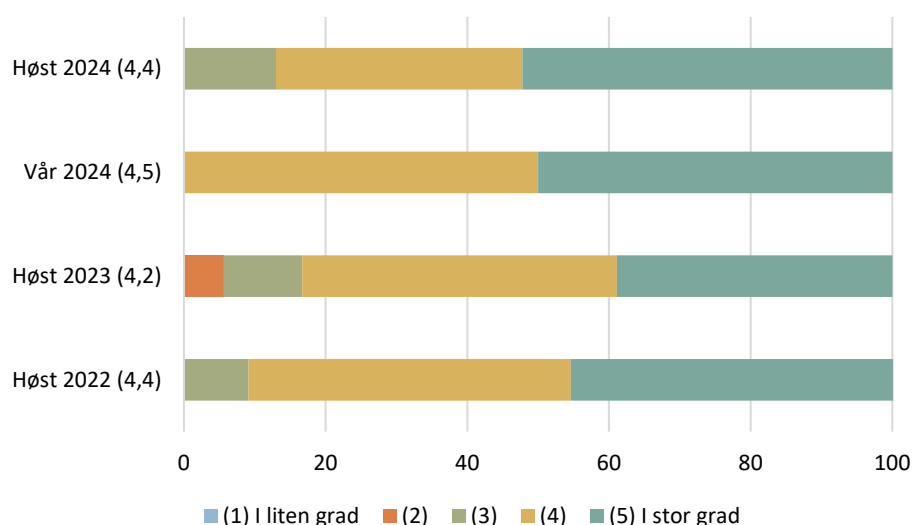
## 6.2 Eksamen

Det er satt av betydelig ressurser og tid for å utarbeide den nasjonale eksamenen. Dette gjøres av eksamensgruppen, som jobber sammen om dette over lang tid og diskuterer og kvalitetssikrer arbeidet. De utarbeider eksamensteksten som skal til studentene samt sensorveiledning og vurderingsrubrikker med poengfordelinger. I tillegg er det en referansegruppe som kvalitetssikrer dette arbeidet. Det legges dermed ned mye arbeidsinnsats i forkant av hver eksamen. I etterkant av eksamenen gjøres det psykometriske analyser samt analyser av eksamensresultatene. Videre er det tilbakemeldinger fra sensorene på eksamen, sensorveiledning og vurderingsrubrikkene som bidrar som et annet ledd i dette kvalitetsarbeidet. Eksamensgruppen, referansegruppen og mange av sensorene er de som underviser ved universitetene og høyskolene, slik sikres det at undervisning og vurdering ses i sammenheng.

I undersøkelsen blir sensorene bedt om å gi tilbakemelding på eksamenen de sensurerer. Flere kommer med utdypende skriftlige tilbakemeldinger om eksamenen. Disse viderefremmes anonymt til eksamensgruppen som et ledd i videreutviklingen av eksamen. I de følgende to figurene som presenteres her, vil vi trekke frem de mer overordnede tilbakemeldingene på eksamen.

I Figur 13 er sensorene blitt bedt om å vurdere i hvilken grad eksamenen var i samsvar med læringsmålene for den aktuelle eksamenen.

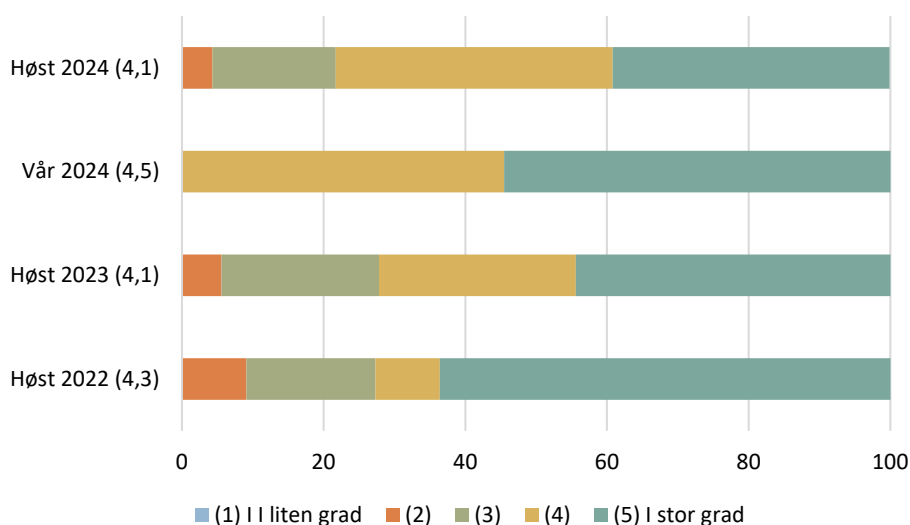
Figur 13. Svarfordeling på spørsmålet «Var eksamen i samsvar med læringsmålene?». Gjennomsnitt i parentes.



I gjennomsnitt, på tvers av alle gjennomføringer, ga 90 prosent av sensorene tilbakemelding om at eksamen i høy grad<sup>21</sup> var i samsvar med læringsmålene studentene skulle oppnå. Vår 2024 ser vi at alle respondentene har gitt tilbakemelding om at eksamenen samsvarer med læringsmålene, mens høst 2023 er det 83 prosent som mener det i høy grad er samsvar, og 6 prosent som oppgir at de i lavere grad opplevde at det var samsvar mellom eksamenssettet og læringsmålet.

I et beslektet spørsmål, i Figur 14, ble sensorene bedt om å vurdere i hvor høy grad eksamen testet juridisk metode.

Figur 14. Svarfordeling på spørsmålet «I hvilken grad testet eksamen juridisk metode?». Gjennomsnitt i parentes.



Sett under ett mente i snitt 81 prosent av sensorene at eksamenen testet juridisk metode i høy grad. Høyest gjennomsnitt var vår 2024 da alle ga tilbakemelding om at eksamen testet juridisk metode i høy grad. Høst 2022 og høst 2023 var noe lavere, med henholdsvis 73 og 72 prosent som mente at eksamen i høy grad testet juridisk metode. Høst 2024 var det 78 prosent som mente at eksamen i høy grad testet juridisk metode. Pedersen og Bråten (2025) viser at underviserne fra de første gjennomføringene i større grad vektla rettsanvendelse fremfor juridisk metode. Det er imidlertid uklart hvilken betydning dette har for resultatene.

### 6.3 Vurderingsrubrikker og sensorveiledning

I denne delen vil vi trekke frem refleksjoner og tilbakemeldinger knyttet til vurderingsrubrikker og sensorveiledning, slik som nytte av vurderingsrubrikkene, erfaring med poengsetting og om sensorene opplever at vurderingen gir et riktig bilde av studentenes prestasjoner.

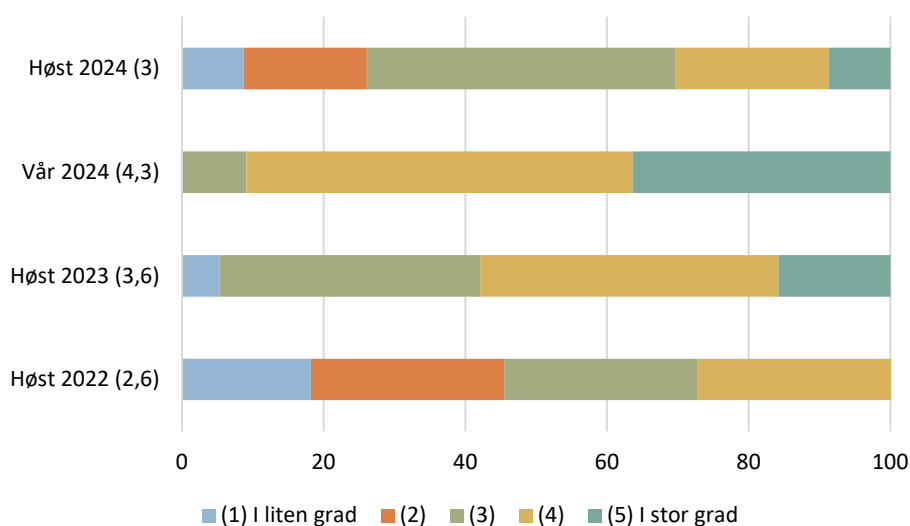
<sup>21</sup> I gradspørsmålene vil (1) og (2) slås sammen til høy grad, og (3) og (4) slås sammen til lav grad.

### 6.3.1 Nytte av vurderingsrubrikker

Som nevnt tidligere ble det ved oppstarten av denne nasjonale eksamenen samlet inn informasjon om ulike vurderingspraksiser. Overordnet fant vi at sensur i fagmiljøet i stor grad baserte seg på helhetsvurderinger og at tydelige vurderingsrubrikker ikke var en utbredt praksis i fagmiljøet. I oppstarten av denne eksamenen var bruk av vurderingsrubrikker dermed en ny prosess for mange sensorer. NOKUT var derfor interessert i å finne ut hvordan sensorene har opplevd vurderingsrubrikkene og skjønnsmessige vurderinger.

I Figur 15 har vi bedt sensorene vurdere i hvilken grad det er gode rammer for skjønnsmessig vurdering. I de ulike semestrene har sensorene fått litt ulike spørsmål. Disse utdypes nedenfor.

Figur 15. Svarfordeling på spørsmålet «I hvilken grad ga sensorveiledning/rubrikkene gode rammer for den skjønnsmessige vurderingen?». Gjennomsnitt i parentes.



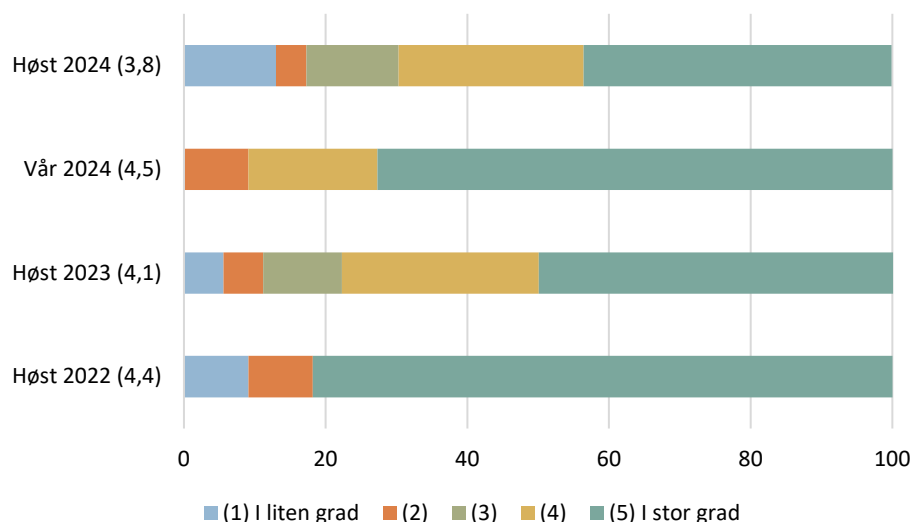
Fra høsten 2022 og frem til våren 2024 var spørsmålsformuleringen «I hvilken grad bidro sensorveiledningen til å gi gode rammer for skjønnsmessige vurdering». I denne perioden ble sensorveiledningen og vurderingsrubrikkene utlevert som to ulike dokumenter. For de tre semestrene kan kommentarer som ble gitt som fritekst, tyde på at det er litt uklart om sensorene har vurdert rammene til å være det som overordnet som står i sensorveiledningen, vurderingsrubrikkene eller begge dokumentene i dette spørsmålet.

Høsten 2024 ble det arbeidet med å gi ett dokument som ga rammer for sensuren. Noe av årsaken var at de hadde kommet frem at noen sensorer ikke forholdt seg så aktivt til sensorveiledningen, men fokuserte på vurderingsrubrikkene. Derfor ble vurderingsrubrikkene og sensorveiledningen sammenslått til ett dokument. Høsten 2024 ble spørsmålet endret til «I hvilken grad ga rubrikkene gode rammer for den skjønnsmessige vurderingen».

I gjennomsnitt for alle gjennomføringer har 52 prosent svart i høy grad og 29 prosent i noen grad. Sensorene har blitt mer positive til at det er gode skjønnsmessige rammer fra første gjennomføring, hvor kun 27 prosent svarte i høy grad. For høst 2023 svarte 58 prosent i høy grad, og for vår 2024 svarte 91 prosent i høy grad. Høst 2024 var det en nedgang igjen, hvor

30 prosent svarte i høy grad. Her endret spørsmålet seg fra «sensorveiledning» til «vurderingsrubrikk», noe som kan tyde på at sensorene opplever at det er litt mindre gode rammer for skjønnsmessige vurderinger når det gjelder vurderingsrubrikkene enn den mer overordnede sensorveiledningen.

Figur 16. Svarfordeling på spørsmålet «Var det nyttig å bruke vurderingsskjema og matrise som vurderingsverktøy?». Gjennomsnitt i parentes.



I Figur 16 ble sensorene stilt spørsmål om i hvilken grad det var nyttig å bruke vurderingsrubrikker som vurderingsverktøy.<sup>22</sup> I gjennomsnitt, på tvers av de fire gjennomføringene, mente 80 prosent av sensorene at det i høy grad var nyttig å bruke vurderingsrubrikkene i sensureringen. Vår 2024 mente 91 prosent av sensorene, som er den høyeste skåren, at vurderingsrubrikkene var nyttige, mens høst 2024 mente 70 prosent av sensorene, som er den laveste skåren, at vurderingsrubrikkene var nyttige. Da var det også 17 prosent som i lav grad synes vurderingsrubrikkene var et nyttig verktøy.

Det er flere sensorer som kommenterer hva og hvorfor de ser nytten av vurderingsrubrikker. Flere trekker frem at dette er en erfaring som de tar med seg til undervisning og vurdering i andre emner:

*Fordelen med sensuren ved deleksamenen er dog at oppstykkningen som vurderingsskjemaet bidrar til, gjør det enklere å se hva slags ferdigheter som er svake og sterke hos den jevne student. (Sensor, høst 2024)*

*Jeg anser dette for å være veldig nyttig, særlig å anvende matrisen, noe som jeg har brukt i egne emner. (Sensor, høst 2024)*

### 6.3.2 Poengsetting i vurderingsrubrikkene

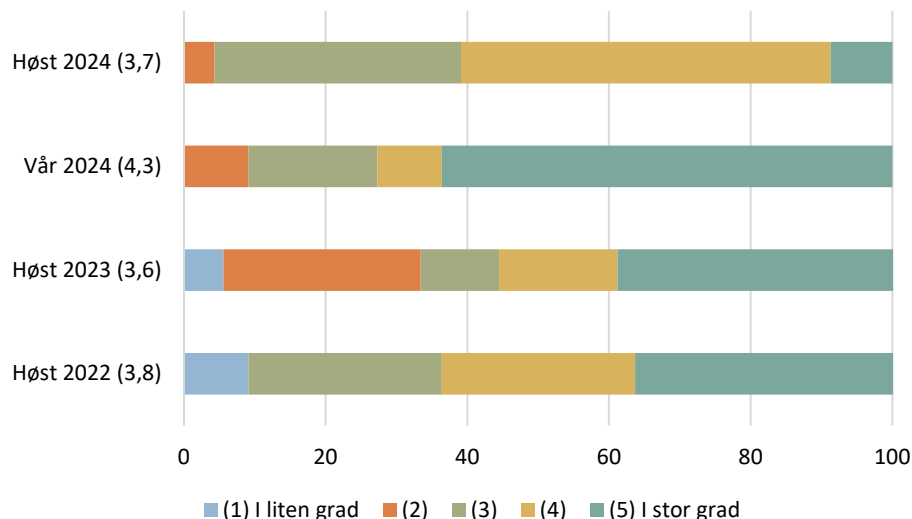
I denne delen vil vi se nærmere på hvordan, og om, sensorene er enige i poengfordelingen i vurderingsrubrikkene. Videre om de opplever at det gir et godt bilde av besvarelsen, samt om de bidrar til at studentene blir vurdert på likt grunnlag.

<sup>22</sup> I tidligere spørsmålsformuleringer har «vurderingsrubrikker» blitt betegnet som «vurderingsskjema» eller «vurderingsmatrise».



I Figur 17 ble sensorene spurt om i hvilken grad de var enige i poengsettingen i vurderingsrubrikkene.

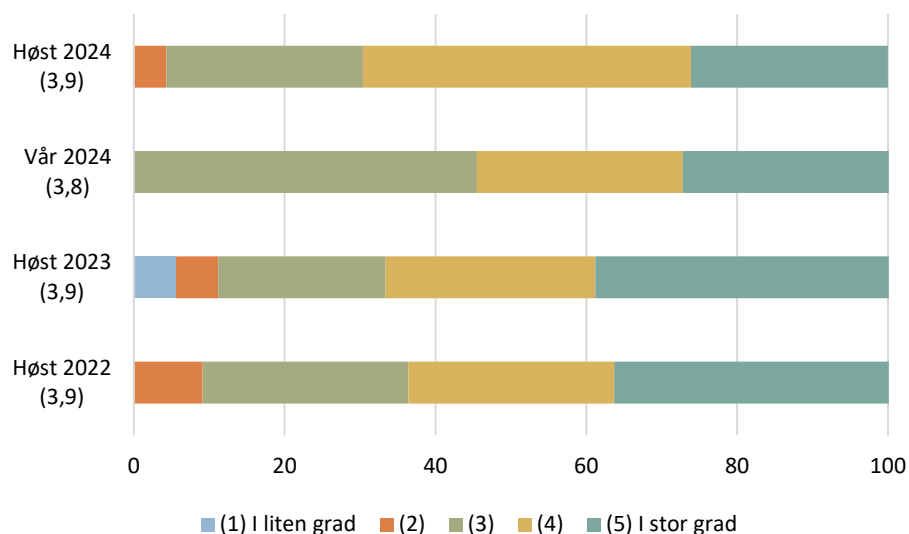
Figur 17. Svarfordeling på spørsmålet «Er du enig i poengsettingen i matrisen?». Gjennomsnitt i parentes.



I gjennomsnitt, på tvers av fire gjennomføringer, var 63 prosent i høy grad enig i poengsettingen i vurderingsrubrikkene, mens 14 prosent i lavere grad var enig i poengsettingen. Hvor enige sensorene er i poengsettingen har variert mellom semestrene. Det kan tyde på at det har blitt opplevd ulikt på de ulike eksamenene. Høst 2023 var det året som færrest, 56 prosent, i høy grad var enig i poengsettingen, mens vår 2024 var det året der flest, 73 prosent, i høy grad var enig i poengsettingen. Siden hver eksamen har ulike case, problemstillinger og vurderingsrubrikker, kan det variere i hvilken grad sensorene opplever at de er treffende.

I sensuren av denne eksamenen, vurderes studentene etter karakterskalaen bestått/ikke bestått. Dermed er det kun to utfall for sensurvedtak for studentene. Det gjøres derfor en vurdering i etterkant av hver eksamen på hvor denne grensen skal settes, basert på forslag fra eksamensgruppen og kalibrering. I Figur 18 ble sensorene spurt om poenggrensen for bestått og ikke bestått svarte til sensorenes vurdering av kandidaten.

Figur 18. Svarfordeling på spørsmålet «Svarte poenggrensen for bestått/ikke bestått til sensorens vurdering av kandidaten?».



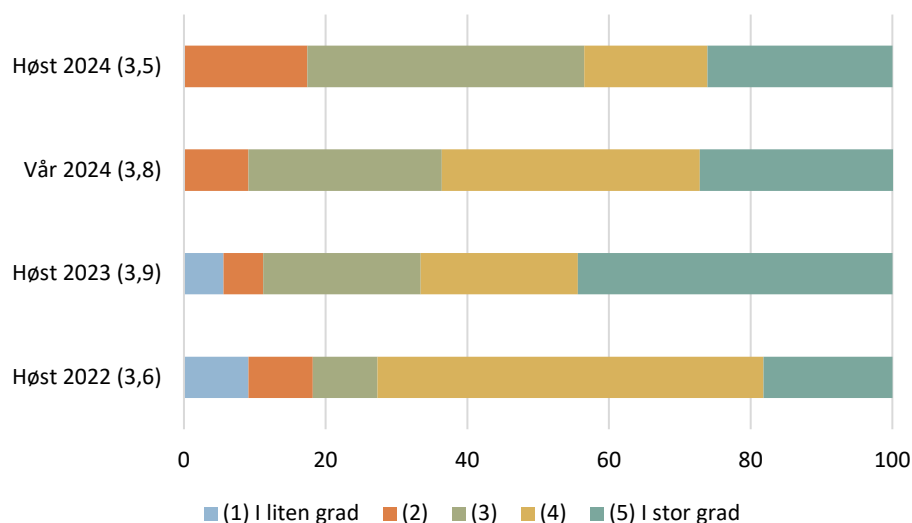
På tvers av de fire gjennomføringene mente 64 prosent av sensorene at poenggrensen for bestått/ikke bestått i høy grad svarte til vurderingen deres av kandidaten. 6 prosent svarte at det i lav grad svarte til vurderingen deres av kandidaten. Vår 2024 var det semesteret der færrest, 55 prosent, mente bestått/ikke bestått i høy grad svarte til vurderingen deres av kandidaten. Selv om det var det semesteret hvor færrest mente at det samsvarte i høy grad, var det ingen som mente at det i lav grad svarte til vurderingen deres av kandidaten. Høst 2024 var det flest, 70 prosent, som mente det i høy grad var samsvar.

I fritekstvarene blir sensorenes vurdering av karaktergrensene noe mer utdypet. For høst 2024 vurderer en sensor at poenggrensen er satt noe høyt, mens fire kommentarer beskriver hvordan poenggrensen er satt for lavt. En av de sensorene som kommenterte at det var for lavt, og skriver:

*De jeg sensurerte med hevdet det samme, da vi opplevde store svakheter i besvarelser vurdert til bestått. (Sensor, høst 2024)*

Dette synspunktet skal vi få videre bekreftet i analysene og fritekstvarene knyttet til et spørsmål om hvorvidt vurderingsrubrikker gir et riktig bilde på studentenes prestasjon. Figur 19 viser sensorene synspunkt på om vurderingsrubrikkene gir et riktig bilde av studentenes prestasjon på eksamen.

Figur 19. Svarfordeling på spørsmålet «I hvilken grad ga vurderingsskjemaet et riktig bilde av studentenes prestasjon på eksamen?». Gjennomsnitt i parentes.



På tvers av gjennomføringer mente i snitt 60 prosent av sensorene at vurderingsrubrikkene i høy grad ga et riktig bilde av studentenes prestasjon på eksamen. Andelen som svarte i høy grad har gradvis sunket for hvert semester, og andelen som har svart i noen grad har økt. Høst 2022 var det 73 prosent som mente at vurderingsrubrikkene i høy grad ga et riktig bilde, mens høsten 2024 var det 44 prosent som mente at vurderingsrubrikkene i høy grad ga et riktig bilde av studentenes prestasjon på eksamen.

I Figur 18 så vi at det høsten 2024 var 70 prosent som mente at poenggrensen for bestått/ikke bestått i høy grad samsvarte til deres vurdering av kandidaten. Samtidig ser vi i Figur 19 at det kun var 44 prosent som mente at vurderingsrubrikkene i høy grad ga et riktig bilde av studentenes prestasjon på eksamen. Dette kan tyde på at selv om flere mente at grensen for å bestå samsvarte med hva studenten presterte, så opplever sensorene at det ikke nødvendigvis gir et like godt og helhetlig bilde på hvordan studenten har prestert.

Hvis vi ser nærmere på hva sensorene skrev som utdypende svar i undersøkelsen, så er det særlig en ting som går igjen. Det er at de studentene som sensorene vurderer som faglige svakere studenter, belønnes i for stor grad når vurderingene knyttes til vurderingsrubrikkene og poenguttelling. Under presenteres et utdrag fra det sensorene beskriver:

*Det ble noe 'statisk', slik at kandidatene ikke fikk poeng for ellers gode drøftelser. Men i det store og hele fungerte det bra. (Sensor, høst 2022)*

*Matrisen inneholder mange relevante punkter relatert til juridisk metode. Men fordi sensurmetoden er slik den er, kan man komme unna med en dårlig strukturert besvarelse, så lenge man treffer de punktene matrisen spør etter. (Sensor høst 2023)*

*Denne gangen var det flere studenter som 'ble reddet' av matrisen, dvs. at de fikk flere poeng enn jeg tror deres metodeforståelse egt. skulle tilsi. (Sensor vår 2024)*

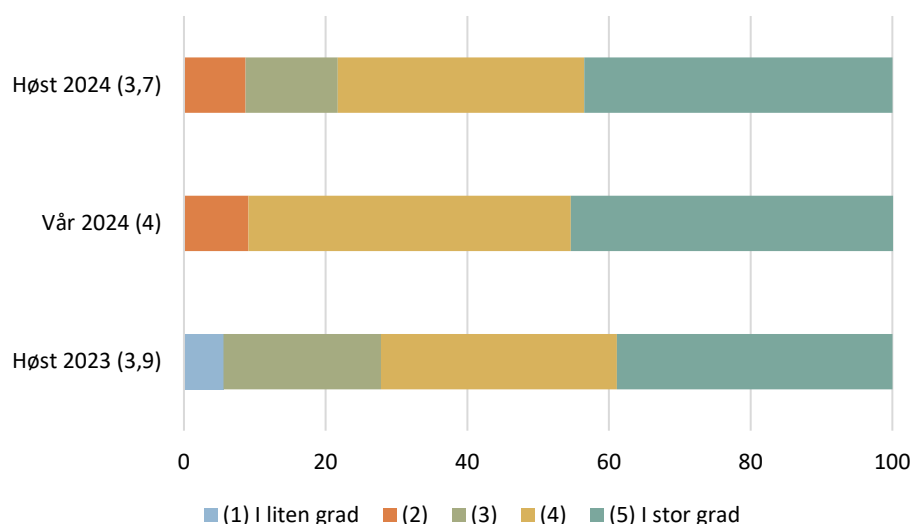
*Matrisene gjør det mulig å score bra med dårlig forståelse, samtidig som enkelte kommer dårlig ut fordi den forståelsen de har ikke premieres i matrisen. (Sensor, høst 2024)*

*Jeg tenker at matrisa for det meste gir god sensurering. Men i noen tilfeller gir den poeng til studenter som virker lite kunnskapsrike men likevel litt snublende og ute av sammenheng får med punktene som er nevnt i matrisa. Bruk av matrise har den ulempen at studenten kan samle poeng og samtidig gi sensor en mistanke om at studenten ikke har forstått så mye. (Sensor, høst 2024)*

Her er det flere som beskriver at studentene kan tildeles poeng til tross for at det overordnede inntrykket er at studentene ikke har gjort en så god prestasjon som det poengene tilsier. Flere sensorer mener altså at vurderingen av studentene ikke er for streng, men snarere tvert imot; vurderingsrubrikker kan redde inn faglig svakere studenter.

I Figur 20 ble sensorene bedt om å vurdere i hvilken grad vurderingsrubrikkene bidro til at studentene ble vurdert etter samme kriterier.<sup>23</sup>

Figur 20. Svarfordeling på spørsmålet «I hvilken grad bidro matrisen til at studentene ble vurdert etter samme kriterier?». Gjennomsnitt i parentes.



I gjennomsnitt på tvers av de tre gjennomføringene svarte 81 prosent av sensorene at vurderingsrubrikkene i høy grad bidro til at studentene ble vurdert etter samme kriterier. Høst 2023 mente 72 prosent av sensorene at de i høy grad bidro til likere sensur, mens 6 prosent av sensorene mente de i lav grad bidro til lik sensur. Semesteret etter, vår 2024, mente 91 prosent av sensorene at vurderingsrubrikkene i høy grad bidro til likere vurdering, mens 9 prosent mente at de i lavere grad bidro til det. Høst 2024 mente 78 prosent av sensorene at de i høy grad bidro til likere sensur, og 9 prosent av sensorene mente de i lavere grad bidro til det.

Mange påpeker også at vurderingsrubrikkene gjør at det ble enklere for sensorene å sensurere besvarelsene, og at det ble likere. Likevel påpeker noen sensorer at det var ulik tolkning av hva de ulike kravene innebar.

<sup>23</sup> Dette spørsmålet ble ikke stilt høst 2022.

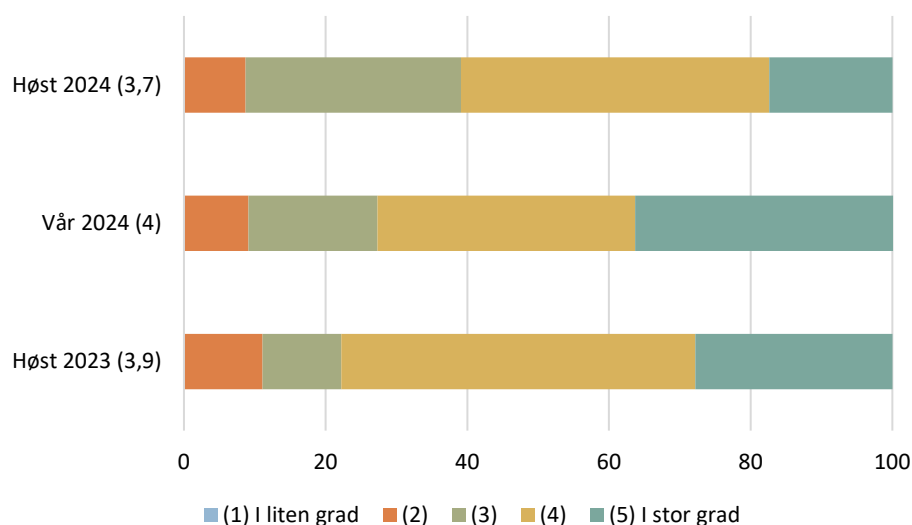
*Hvor langt de ulike sensorene 'strekker' matrisen har jeg inntrykk av at en ikke får gjort noe med, ettersom det alltid vil være tale om skjønn. (Sensor, høst 2024)*

Dette understreker viktigheten av kalibreringsdialoger og faglige diskusjoner om kriterier, faglige standarder og forståelse av ord (Bloxham et al., 2018), slik som gjennom innledende sensormøter og kalibreringsmøter. Dette er møter hvor fagmiljøet samles til dialog, og de aller fleste sensorene, på tvers av gjennomføringer, kommenterer at både sensormøtene og kalibreringsmøtene er nyttige og fungerer bra. De aller fleste sensorene kommenterer videre at også samsensur-ordningen fungerer fint. Og selv om noen påpeker at de har litt ulik tolkning av kriteriene, så er det svært høyt samsvar statistisk sett, som vi så i kapittel 4.

### 6.3.3 Vurderingsrubrikker som begrunnelse og tilbakemelding

Sensorveiledningen, vurderingsrubrikkene og poengfordelingen i vurderingsrubrikkene, utgjør studentenes begrunnelse for karaktervedtaket. Vi har derfor bedt sensorene om å vurdere i hvilken grad de anser at vurderingsrubrikkene gir studentene god tilbakemelding på sin eksamen og begrunnelse for karakter. Svarene presenteres i Figur 21.<sup>24</sup>

Figur 21. Svarfordeling på spørsmålet «I hvilken grad anser du matrisen og sensorveiledningen å kunne gi studentene god tilbakemelding på sin eksamen/begrunnelse for karakter?». Gjennomsnitt i parentes.



I gjennomsnitt på tvers av de tre gjennomføringene svarte 71 prosent av sensorene at de i høy grad ga en god tilbakemelding og begrunnelse for karakter. Høst 2023 var det 78 prosent som mente vurderingsrubrikkene og sensorveiledningen i høy grad gir studentene god tilbakemelding på sin eksamen og begrunnelse for karakter, mens høst 2024 var 61 prosent som i høy grad mente det samme. Samtidig har de 11 prosentene som mente at vurderingsrubrikkene i lavere grad gir studentene god tilbakemelding på sin eksamen og begrunnelse for karakter høst 2023, gått ned til 9 prosent høst 2024.

<sup>24</sup> Dette spørsmålet ble ikke stilt høst 2022.

*Matrisen bidrar til å gi kandidatene en solid begrunnelse for resultatet - dette er bra. (Sensor, vår 2024)*

*Jeg er usikker på om studentene som ikke består forstår rubrikkene. (Sensor, høst 2024)*

I rapporten «Studentstemmen: om undervisning, emneinnhold og vurdering i rettsanvendelse for barnevernsfaglige masterprogram» (Karlsen & Pedersen, 2025) fant vi at av de studentene som ikke besto ordinær eksamen på første forsøk, oppga over halvparten at de hadde nytte av begrunnelsen sin fra eksamenen.

## 6.4 Nytteeffekt for sensorene

I undersøkelsen finner vi også at flere av sensorene trekker frem nytteeffekter for dem selv ved å være sensor på denne eksamenen.

I delkapittelet om «nytte av vurderingsrubrikker» så vi at en sensor påpekte at vurderingsrubrikkene bidro til å gjøre det synlig hvilke ferdigheter som var svakere og sterkere hos den jevne studenten. I forlengelsen av dette er det flere sensorer som er undervisere, som opplever at sensurprosessen gir verdifulle tilbakemeldinger som hjelper dem å tilpasse fagopplegg og stille relevante krav til studentene. Flere påpeker at erfaring fra sensurprosessen bidrar til utvikling av undervisningen.

*Det er lærerikt å se hvordan det svares på spørsmål – det gir grunnlag for å tilpasse videre undervisning. Nyttig å diskutere med medsensor. (Sensor, høst 2024)*

*Oppdaga det grunnleggende problemet at studentene ikke klarer å lese lovtekst. (Sensor høst 2024)*

*Erfaringen fra sensur spiller direkte inn i undervisningsoppbyggingen, det kan være med tanke på temaer og på undervisningsopplegg. (Sensor høst 2024)*

*Jeg er lærer i faget og det å være med på sensur er nyttig og bidrar til bedre forståelse av det jeg må kreve av egne studenter. (Sensor høst 2024)*

I tillegg til at sensurering av studentbesvarelser bidrar til utvikling i undervisningen, er det også flere sensorer som trekker frem at kontakt med andre fagpersoner i fagmiljøet bidrar til utvikling. De fleste av sensorene trekker frem at sensormøter og kalibreringsmøter er nyttige for dem, utover det å sensurere den konkrete eksamenen. De trekker frem at møtene gir faglig utbytte, nettverksbygging og inspirasjon til utvikling av emner. Samarbeid med kollegaer på tvers av institusjoner og erfaringsutveksling fremheves som særlig nyttig.

*Fint å få kontakter. (Sensor, høst 2024)*

*Nyttig å samarbeide med andre jurister og knytte nettverk. (Sensor, høst 2024)*

*Nyttig for videre utvikling av eget emne, knytte kontakter på tvers av institusjoner. (Sensor, høst 2024)*

*Nyttig for arbeid også i andre jussemner på studiet. (Sensor, høst 2024)*

*Brukes i stor grad for å planlegge, utvikle og tilpasse undervisningen. (Sensor, høst 2024)*

## 6.5 Oppsummering

Sensorundersøkelsene bidrar til å nyansere forståelsen av vurderingsarbeidet i nasjonal eksamen i rettsanvendelse. Gjennomgående viser resultatene at en stor andel av sensorene opplever vurderingsrubrikkene som et nyttig sensurverktøy for å sikre konsistent og rettferdig vurdering, samtidig som det pekes på utfordringer knyttet til skjønnsmessig vurdering og helhetsinntrykk av studentenes prestasjoner.

Flertallet av sensorene mener at eksamenene i stor grad tester juridisk metode og samsvarer med læringsmålene, og at vurderingsgrunnlaget gir studentene en begrunnelse for karakteren. Samtidig fremheves det at enkelte rubrikker kan gi poeng til kandidater med svak forståelse for juridisk metode, og at det er behov for videre utvikling av vurderingsverktøyene for å fange opp helhet og dybde i besvarelsene.

Sensorene rapporterer også om faglig utbytte av å delta i sensurarbeidet i emnet knyttet til nasjonal eksamen og utover dette, både gjennom erfaringsutveksling i sensormøter og som inspirasjon til egen undervisning. Dette understreker at sensurprosessen ikke bare er en vurderingsaktivitet, men også en arena for faglig utvikling og kvalitetsarbeid.

## Del 3 Eksamenene og studentenes resultater

### 7 Kjennetegn ved de første eksamenene

Syv institusjoner startet opp høst 2022, og seks av disse gjennomførte nasjonal eksamen denne høsten. Dette var, som vi ser i tabell 13, HVL, NTNU, UiA, UiB, UiS og VID med totalt 165 kandidater. I lys av oppmeldingstall, som innhentes før endelig trekkfrist for studentene, var gjennomføringsgraden 94 prosent. Reell gjennomføringsgrad vil trolig være noe høyere. USN startet opp et deltidsstudium høst 2022 og avla dermed eksamen vår 2023 isteden. USN inngår derfor innunder institusjonene som startet opp høst 2022 selv om de har et annet eksamenstidspunkt.

Tabell 13. Oversikt over institusjonenes oppstart og eksamenstidspunkt.

| Institusjon                                            | Forkortelse | Oppstart  | Ordinær gjennomføring |
|--------------------------------------------------------|-------------|-----------|-----------------------|
| Høgskulen på Vestlandet                                | HVL         | Høst 2022 | Høst                  |
| Norges teknisk -naturvitenskapelige universitet - NTNU | NTNU        | Høst 2022 | Høst                  |
| Universitetet i Agder                                  | UiA         | Høst 2022 | Høst                  |
| Universitetet i Bergen                                 | UiB         | Høst 2022 | Høst                  |
| Universitetet i Stavanger                              | UiS         | Høst 2022 | Høst                  |
| VID vitenskapelige høyskole*                           | VID         | Høst 2022 | Høst                  |
| Universitetet i Sørøst-Norge                           | USN         | Høst 2022 | Vår                   |
| UiT Norges arktiske universitet                        | UiT         | Høst 2023 | Høst                  |
| Høgskolen i Østfold                                    | HiØ         | Høst 2023 | Høst                  |
| Høgskulen i Volda                                      | HVO         | Høst 2023 | Høst                  |
| Nord universitet                                       | Nord        | Høst 2023 | Høst                  |
| OsloMet – storbyuniversitet                            | OsloMet     | Høst 2023 | Høst                  |
| Universitetet i Innlandet                              | INN         | Høst 2023 | Høst                  |

\*VID endret ordinær gjennomføring fra høst til vår fra og med opptak høsten 2024.

Den første gjennomføringen ble kalt en pilot slik at det var rom for å gjøre justeringer i det videre arbeidet med eksamenen. Resultatene var imidlertid tellende for studentene, men de ble ikke offentliggjort slik NOKUT har gjort ved senere gjennomføringer. En



konteeksamen for pilotgjennomføringen ble avholdt 12. april 2023, mens for USN var dette som nevnt første gjennomføring.<sup>25</sup>

Høst 2023 ble første ordinære gjennomføring for alle 13 utdanningsinstitusjoner som tilbyr barnevernsfaglige masterprogrammer. Her var 363 kandidater registrert til eksamen og 300 gjennomførte eksamen, altså en gjennomføringsgrad på 83 prosent.

I videre gjennomføringer har høstsemesteret primært vært en ordinær eksamen, mens våreksamen avholdes som en konteeksamen. Tabell 14 vises en oversikt over gjennomførte eksamener, antall kandidater og andel bestått før og etter klagesensur. Totalt har rundt 900 barnevernsstudenter tatt nasjonal eksamen i rettsanvendelse og i snitt har 58 prosent bestått.

Tabell 14. Oversikt over gjennomføringer og eksamensresultater høsten 2022 til og med høsten 2024.

| Gjennomføring                           | Eksamenstidspunkt | Antall oppmeldt | Antall levert | Andel bestått før klage | Andel bestått etter klage |
|-----------------------------------------|-------------------|-----------------|---------------|-------------------------|---------------------------|
| Høst 2022                               | 07.12.2022        | 175             | 165           | 57 %                    | 65 %                      |
| Vår 2023                                | 12.04.2023        | 74              | 70            | 41 %                    | 47 %                      |
| Høst 2023                               | 06.12.2023        | 363             | 300           | 54 %                    | 58 %                      |
| Vår 2024                                | 10.04.2024        | 147             | 133           | 53 %                    | 59 %                      |
| Høst 2024                               | 04.12.2024        | 322             | 267           | 58 %                    | 63 %                      |
| <b>Totalt (snitt av prosentandeler)</b> |                   | 1081            | 935           | 53 %                    | 58 %                      |

Dersom vi ser nærmere på studentenes prestasjoner utover beståttprosent, kan vi i tabell 15 få inntrykk av stor variasjon i nivået til studentene. Til tross for ulike makspoeng og beståttgrenser, er det store sprik mellom svakeste og sterkeste student i samtlige gjennomføringer. Høst 2022 og høst 2024 er de eneste gjennomføringene hvor studentenes poengsum i snitt er høyere enn beståttgrensen.

Tabell 15. Oversikt over poeng på tvers av gjennomføringer.

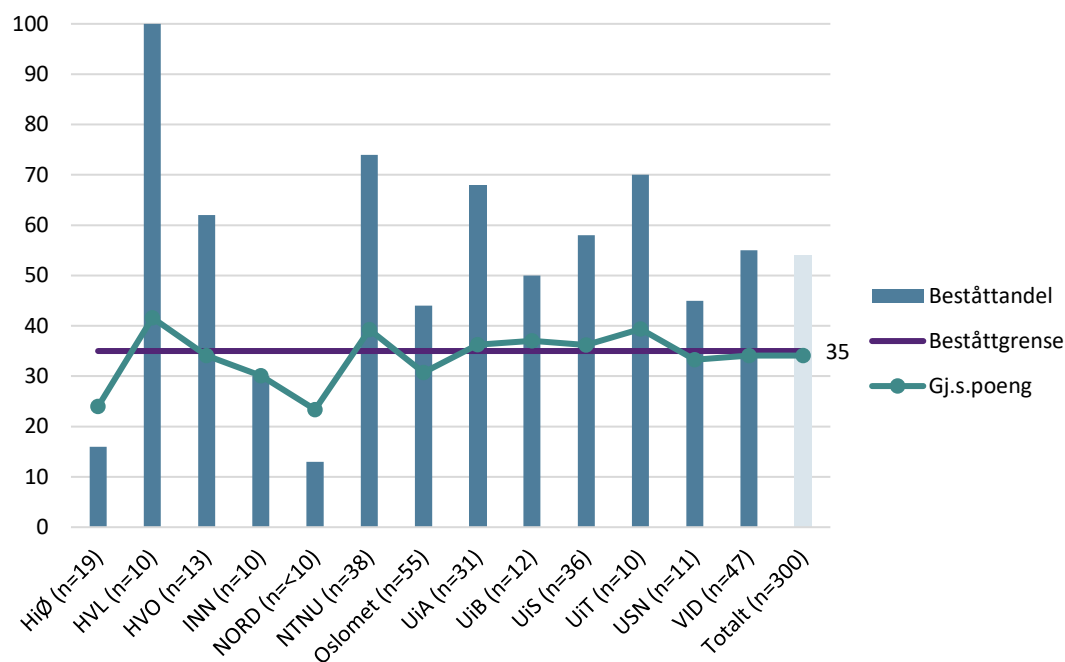
| Gjennomføring | Makspoeng | Beståttgrense | Gjennomsnittspoeng (nasjonalt) | Minste oppnådde poeng (nasjonalt) | Mest oppnådde poeng (nasjonalt) |
|---------------|-----------|---------------|--------------------------------|-----------------------------------|---------------------------------|
| Høst 2022     | 60        | 30            | 31,0                           | 12                                | 54                              |
| Vår 2023      | 65        | 35            | 31,4                           | 13                                | 47                              |
| Høst 2023     | 62        | 35            | 34,1                           | 3                                 | 60                              |
| Vår 2024      | 70        | 35            | 33,6                           | 8                                 | 63                              |
| Høst 2024     | 71        | 37            | 38,5                           | 5                                 | 66                              |

Ser vi på resultatene fra de to siste ordinære gjennomføringene fordelt på utdanningsinstitusjon, er det også stor variasjon. Høst 2023 (figur 22) viser blant annet at HVL, NTNU og UiT skilte seg ut med svært høye andeler studenter som bestod eksamen. Høst 2024 (figur 23) viser at UiB, HVL og UiA skiller seg ut med høye beståttandeler. Vi skal

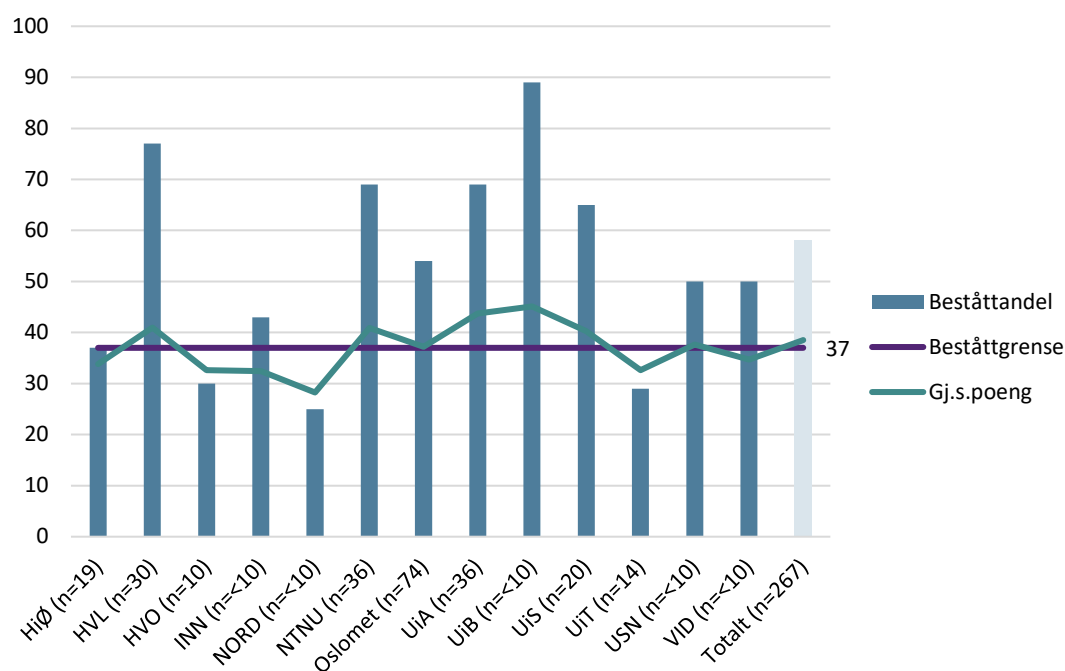
<sup>25</sup> Fra og med høsten 2024/våren 2025 gjaldt også dette VID. Både USN og VID har dermed konteeksamen i høstsemesteret. Nasjonale resultater i høstgjennomføringene må ses i lys av dette.

imidlertid legge merke til den varierende studentstørrelsen på tvers av institusjonene, hvorav flere har under og rundt ti studenter. Det gjør at resultatene er sårbare for tilfeldige variasjoner, eksempelvis opptakskravet til den enkelte institusjon det gjeldende semesteret.

Figur 22. Resultater på nasjonal eksamen i rettsanvendelse høst 2023 på tvers av institusjoner.



Figur 23. Resultater på nasjonal eksamen i rettsanvendelse høst 2024 på tvers av institusjoner.



Tidligere analyser av nasjonale eksamener (Haakens & Bråten, 2023; Haakens, Karlsen & Bråten, 2021;) har vist at det kan være faktorer både ved studentene selv og studieprogrammene som forklarer studentenes prestasjon på eksamen. Tilsvarende undersøkelser er gjennomført knyttet til denne eksamen, noe vi vil se nærmere på i del tre.

## 8 Studentens valg av master

### Overgang fra bachelor til master: Utdanningsvalg og mønstre

Siden master i barnevern og barnevernsarbeid har ulike opptakskrav, finner vi ulike bachelorbakgrunn hos studentene ved de to masterprogrammene. Mens master i barnevern krever en bachelor i barnevern, er master i barnevernsarbeid rettet mot studenter med bachelor i barnevern, sosialt arbeid, vernepleie og åpner samtidig for andre utdanningsgrupper.

Ved bruk av bakgrunnsdata om studentene som tar nasjonal eksamen, har vi sett på hvilke masterprogrammer studentene med ulike bachelorbakgrunnene velger.<sup>26</sup> Av studentene som hadde bachelorbakgrunn tilgjengelig, ble disse kategorisert inn i tre kategorier: Bachelor i barnevern, sosialt arbeid og vernepleie/andre profesjoner.<sup>27</sup>

Nesten 72 prosent tar en master i barnevernsarbeid, mens 28 prosent tar en master i barnevern. Tabell 16 viser videre at bachelorbakgrunn påvirker hvilken masterutdanning de velger.

Tabell 16. Valg av masterprogram fordelt på bachelorbakgrunn. Prosentandel av bachelorgruppe i parentes.

| Bachelorgruppe                      | Master i barnevern  | Master i barnevernsarbeid |
|-------------------------------------|---------------------|---------------------------|
| Barnevern (n=304)                   | 123 (40,5 %)        | 181 (59,5 %)              |
| Sosialt arbeid (n=99)               |                     | 99 (100 %)                |
| Vernepleie/Andre profesjoner (n=37) | 1 (2,7 %)           | 36 (97,3 %)               |
| <b>Totalt (n=440)</b>               | <b>124 (28,2 %)</b> | <b>316 (71,8 %)</b>       |

Som opptakskravene gir mulighet til, ser vi at studentene som følger masterutdanningen i barnevernsarbeid, har en variert faglig bakgrunn. Denne utdanningen fungerer som en tverrfaglig inngangsport til videre kompetanseutvikling innen barnevernsfeltet. Den største gruppen er studenter med bachelor i barnevern, med 181 personer, som utgjør rundt 60 prosent av både de med bachelor i barnevern og blant bachelorutdanningene i master i barnevernsarbeid. Rundt en tredjedel av bachelorutdanningene, 99 studenter, kommer fra sosialt arbeid, mens 36 studenter har bakgrunn fra vernepleie eller andre profesjoner.

Det er interessant å merke seg at studenter med bachelor i barnevern fordeler seg mellom begge masterutdanningene, der flertallet (59,5 prosent) velger master i barnevernsarbeid, mens resterende 40,5 prosent velger master i barnevern. Med unntak av én student som ble identifisert med «andre profesjoner», har samtlige av studentene i master i barnevern også en bachelor i barnevern. Dette skyldes at opptakskrav til denne masteren krever nettopp bachelor i barnevern eller barnevernspedagogutdanning.

Det at studentene med bachelor i barnevern velger master i barnevernsarbeid fremfor master i barnevern kan forklares med en oppfatning om at førstnevnte master er en bredere og mer fleksibel utdanning med flere yrkesmuligheter. Dette så vi i

<sup>26</sup> Her er det kun brukt bakgrunnsdata for studentene som tok eksamen høst 2023 og høst 2024. Av 567 studenter hadde 440 studenter tilgjengelig informasjon om bachelorbakgrunn. Mer om bruk av bakgrunnsdata og utvalg i kapittel «Analyse av studentenes resultater og bakgrunn».

<sup>27</sup> I datagrunnlaget var det ikke mulig å skille vernepleie fra andre profesjoner på grunn av svært lavt antall og uklar bachelorinformasjon. Andre profesjoner kan eksempelvis være utdanninger innen sykepleie, psykologi og pedagogikk.

studentundersøkelsen blant barnevernsstudentene som NOKUT gjennomførte høst 2024 (Karlsen & Pedersen, 2025), samt i intervjuer med studenter. Flere ønsket seg en bredere masterutdanning for å ha flere arbeidsmuligheter senere. En annen forklaring kan være at flere institusjoner tilbyr kun én av masterne, noe som kan påvirke valget deres på grunn av et ønske om å studere ved en spesifikk institusjon.

## 9 Analyse av studentenes resultater og bakgrunn

I dette kapittelet skal vi se nærmere på studentenes resultater på nasjonal eksamen i rettsanvendelse. Basert på bakgrunnsdata har vi undersøkt sammenhenger mellom studentenes eksamensresultater og deres utdanningsinstitusjonstilhørighet, inneværende karaktersnitt<sup>28</sup>, karakterer fra videregående skole, bachelorbakgrunn og masterprogram. Analysene er knyttet til de to siste ordinære gjennomføringene høst 2023 og høst 2024.

For å si noe om det generelle kompetansenivået til studentene fra videregående skole har vi valgt å se på karakterpoeng – et snitt av alle karakterer inkludert tilleggspoeng fra real- og språkfag. I tillegg ser vi på studentenes snittkarakter i norsk for studenter som tok eksamen høst 2024.<sup>29</sup> Norsk karakteren kan si noe om språkføring. Presist språk er viktig i juss og kan påvirke studentenes fremstilling av caseoppgavens rettslige problemstillinger. Utover norsk karakter tester nasjonal eksamen i rettsanvendelse nokså spesifikk kunnskap som ikke er direkte overførbart til fag fra videregående skole.<sup>30</sup>

### 9.1 Datagrunnlag

Analysen av studentenes resultater på nasjonal deleksamen i rettsanvendelse bygger på resultatdata og bakgrunnsdata hentet fra Felles studentsystem ved Sikt – Kunnskapssektorens tjenesteleverandør.

Bakgrunnsdata er koblet sammen med eksamensresultater med kandidatnummer som koblingsnøkkel. Inkludert i bakgrunnsdata er kjønn, alder, inneværende karaktersnitt, utdanningsinstitusjon, bachelorbakgrunn og karakterer fra videregående skole. Dette gir et bredt grunnlag for å analysere sammenhenger mellom studentenes karakterer og deres demografiske kjennetegn og utdanningsbakgrunn.<sup>31</sup>

I tabell 17 ser vi hvordan kandidatene fordeler seg på tvers av utdanningsinstitusjoner samt største aldersgruppe og kjønnskategori. Variasjonen i størrelse reflekterer forskjeller i eksamensavvikling og kandidatvolum mellom gjennomføringene. Analysene behandler hver eksamensavleggelse som en separat observasjon, noe som betyr at kandidater som har tatt eksamen i flere semestre registreres som separate enheter for hvert semester. Denne tilnærmingen gjør det mulig å analysere semestervariasjoner og institusjonsspesifikke forskjeller, men det må tas hensyn til at samme kandidat kan være representert i flere semestre.

---

<sup>28</sup> Inneværende karaktersnitt er et snitt av fag studentene har hatt i løpet av masterprogrammet, ekskludert resultatet på nasjonal deleksamen i rettsanvendelse.

<sup>29</sup> Dette var første gjennomføring hvor norsk karakter ble hentet ut av Sikt.

<sup>30</sup> Rettslære er kanskje det nærmeste man kommer, men dette er ikke informasjon vi har fra Sikt.

<sup>31</sup> Tilgjengelige bakgrunnsvariabler varierer noe mellom kandidatene, og det vil derfor ikke være alle kandidater med i alle analysene. Av 567 kandidater totalt var det 493 kandidater (87 prosent) med mer enn 50 prosent av bakgrunnsvariablene tilgjengelig. Vi anser dette som en tilstrekkelig andel for å foreta analyser av sammenhenger mellom bakgrunnsdata og resultatdata.

Tabell 17. Antall studenter per utdanningsinstitusjon og demografiske variabler i datagrunnlaget.

| Institusjon   | Høst 2023  | Høst 2024  | Totalt     |
|---------------|------------|------------|------------|
| UiB           | 12         | 9          | 21         |
| UiT           | 10         | 14         | 24         |
| NTNU          | 38         | 36         | 74         |
| UiA           | 31         | 36         | 67         |
| HVL           | 10         | 29         | 39         |
| NORD          | 8          | 4          | 12         |
| INN           | 10         | 7          | 17         |
| Oslomet       | 55         | 74         | 129        |
| UiS           | 36         | 20         | 56         |
| USN           | 11         | 6          | 17         |
| HVO           | 13         | 10         | 23         |
| HiØ           | 19         | 19         | 38         |
| VID           | 47         | 3          | 50         |
| <b>Totalt</b> | <b>300</b> | <b>267</b> | <b>567</b> |
| Kvinner       | 275        | 242        | 517        |
| 22–28 år      | 190        | 188        | 378        |

OsloMet er den største enkeltinstitusjonen i datamaterialet, etterfulgt av NTNU og UiA. Det er viktig å merke seg at enkelte undergrupper har svært lavt antall studenter, spesielt ved nedbrytning på utdanningsinstitusjon og bakgrunnsvariabler. Dette kan føre til ustabile estimer og større usikkerhet i resultatene. Derfor bør funn tolkes med varsomhet, og hovedvekten bør legges på mønstre og tendenser fremfor enkeltverdier.

Når det gjelder aldersfordeling, viser datasettet at aldersgruppen 22–28 år er klart størst på tvers av gjennomføringene. Det er også en tydelig overvekt av kvinner i studentmassen.

Bachelorbakgrunn har vi klassifisert ved hjelp av mønstergjenkjenning i tekstfeltet for bachelorgrad. Vi har kategorisert bachelorbakgrunn inn i tre hovedgrupper. Tabell 18 viser at det er klart flest som har en bachelor i barnevern, og denne kategorien brukes derfor som referansekategori i regresjonsanalysene.

Tabell 18. Oversikt over datamaterialet. Antall i hver bachelorgruppe.

| Gruppe                                                | Høst 2023  | Høst 2024  | Totalt     |
|-------------------------------------------------------|------------|------------|------------|
| Bachelor i barnevern                                  | 157        | 147        | <b>304</b> |
| Bachelor i sosialt arbeid                             | 52         | 48         | <b>100</b> |
| Bachelor i vernepleie/Andre profesjoner <sup>32</sup> | 91         | 72         | <b>163</b> |
| <b>Totalt</b>                                         | <b>300</b> | <b>267</b> | <b>567</b> |

Videre har vi plassert studentene i to grupper basert på hvilket masterprogram de følger: Master i barnevern (n = 232) og master i barnevernsarbeid (n = 335). I tabell 19 ser vi at

<sup>32</sup> Vi ønsket å bryte opp denne kategorien slik at vi fikk vernepleie for seg. Det var ikke mulig grunnet kategorisering i datagrunnlaget. Derimot ser vi at totalt antall vernepleie i datagrunnlaget vårt for alle institusjoner og alle semestre er omtrent ti studenter.

master i barnevernsarbeid utgjør en større gruppe enn master i barnevern totalt og i begge gjennomføringene.

Tabell 19. Antall studenter i hvert masterprogram.

| Masterprogram    | Høst 2023  | Høst 2024  | Totalt     |
|------------------|------------|------------|------------|
| Barnevern        | 125        | 107        | 232        |
| Barnevernsarbeid | 175        | 160        | 335        |
| <b>Totalt</b>    | <b>300</b> | <b>267</b> | <b>567</b> |

## 9.2 Metodiske betraktninger

I analysene benyttes ulike regresjonsmodeller. Formålet med å bruke regresjonsanalyse her er å gi en presis beskrivelse av de statistiske sammenhengene mellom utfallet – poengsum på nasjonal eksamen og de øvrige variablene i modellene. Ved å inkludere flere variabler i modellene kan vi se hvor robuste statistiske sammenhenger mellom to variabler er når vi tar høyde for rimelige drivere av statistiske sammenhenger. For eksempel kan vi ved å ta høyde karakterer fra videregående skole vurdere om det er sannsynlig at en forskjell mellom to institusjoner utelukkende drives av at den ene institusjonen tar inn de mest skoleflinke studentene.

Det er viktig å understreke at slutningene vi kan trekke fra modellen er deskriptive – modellene gir ikke grunnlag for å trekke kausale slutninger. En statistisk signifikant forskjell mellom to institusjoner kan selvsagt skyldes mange andre faktorer enn hvor skoleflinke studentene var på videregående og jobben studieprogrammet gjør med å forberede studentene til nasjonal eksamen. Derfor oppfordrer vi leseren til å ikke tolke resultatene lenger enn som et utgangspunkt for kausale hypoteser, som kan styrkes eller svekkes ved hjelp av andre data. For eksempel ville signifikante forskjeller mellom to institusjoner over en lengre tidsperiode enn vi dekker her kunne styrket en kausal hypotese om at det er vesentlige forskjeller i hvor godt studieprogram forbereder studentene til eksamen. Dette ville gjort det mulig å se bort fra at egenskaper ved enkeltkull, som at mange studenter ved en institusjon ble langvarig influensasyke, var den viktigste driveren av forskjeller i eksamensresultat.

## 9.3 Karaktersnitt fra videregående skole

For å analysere sammenhengen mellom studentenes karaktersnitt fra videregående skole (heretter kalt VGS-poeng) og prestasjonene deres på nasjonal eksamen har vi benyttet både lineær og logistisk regresjon. Den lineære regresjonsanalysen estimerer den statistiske sammenhengen mellom poengsummen på eksamen og VGS-poeng, mens den logistiske regresjonen estimerer sannsynligheten for å bestå eksamen. En odds ratio over 1 betyr at sannsynligheten for å bestå er høyere i den ene gruppen enn i referansegruppen. For å sikre mer intuitive tolkninger er VGS-poeng sentrert rundt gjennomsnittet i analysene.

Analysen er gjennomført i tre steg: Først analyseres hver eksamensgjennomføring separat med enkle modeller. Deretter kontrollerer vi for utdanningsinstitusjonelle forskjeller. Til slutt har vi laget en samlet modell som omfatter både gjennomføringseffekter og utdanningsinstitusjonelle variasjoner.

I tabell 20 ser vi resultatet fra lineær regresjon av enkel modell. Modellen viser at det er en svak positiv og statistisk signifikant sammenheng mellom VGS-poeng og poengsum.



Samtidig ser vi at styrken på og forklaringskraften av sammenhengen er sterkere i høst 2023 enn i høst 2024.

Tabell 20. Enkel lineær regresjonsmodell av poengsum på VGS-poeng.

| Gjennomføring                               | Høst 2023 | Høst 2024 |
|---------------------------------------------|-----------|-----------|
| VGS-poeng                                   | 0,56***   | 0,47***   |
| Gjennomsnittlig poengsum uten karaktersnitt | 34,78     | 38,66     |
| R <sup>2</sup>                              | 0,128     | 0,064     |
| Antall studenter                            | 256       | 229       |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Modellene predikerer at en student med karaktersnitt 0 fra videregående skole (som i praksis ikke eksisterer) ville oppnådd 34,78 poeng på eksamen i 2023, og for hver økning på 1 poeng i karaktersnitt øker den forventede poengsummen med 0,56 poeng. Tilsvarende i høst 2024 øker poengsummen med 0,47 for hver ekstra karakterenhet i VGS-poeng. Forklaringskraften i begge modellene er relativt svake, men svakest i høst 2024. Dette antyder at det er flere eller andre faktorer som spiller inn på poengsummen.

I neste analyse, presentert i tabell 21, har vi kontrollert for om studentenes utdanningsinstitusjon har betydning for sammenhengen mellom VGS-poeng og poengsum på eksamen. I høst 2023 ser vi at modellen nå forklarer 28,5 prosent av variasjonen i eksamensresultatene. Her øker poengsum med 0,64 poeng for hver ekstra karakterenhet i VGS-poeng. Sammenhengen er fremdeles statistisk signifikant. I høst 2024 er fortsatt den statistiske sammenhengen noe svakere. Her øker poengsummen med 0,45 poeng per karakterenhet, og modellen forklarer 14,1 prosent av variasjonen.

Vi ser også at flere utdanningsinstitusjoner har signifikant høyere poengsummer enn referansegruppen (OsloMet), særlig i høst 2023. Blant andre har studenter ved HVL, UiB, NTNU og VID over 7 poeng mer i snitt i poengsum enn OsloMet. Høst 2024 viser tre institusjoner med signifikant høyere poengsummer enn referansegruppen: Studenter ved USN, UiA og HVL med henholdsvis 12,81, 5,7 og 5,1 poeng mer enn OsloMet.

Tabell 21. Full lineær regresjonsmodell av poengsum på VGS-poeng med kontrollvariabler.

| Variabel                          | Høst 2023 | Høst 2024 |
|-----------------------------------|-----------|-----------|
| VGS-poeng                         | 0,635***  | 0,23      |
| <b>Kontrollvariabler</b>          |           |           |
| Alder                             | 0,057     | -0,282*   |
| Kvinne                            | 1,074     | 0,757     |
| <b>Institusjon (ref. OsloMet)</b> |           |           |
| HiØ                               | -4,848    | -0,782    |
| HVL                               | 9,79**    | 5,104*    |
| HVO                               | 6,129     | -1,088    |
| INN                               | 2,672     | 6,105     |
| Nord                              | -5,581    | 6,56      |
| NTNU                              | 7,672***  | 2,473     |
| UiA                               | 5,444*    | 5,775**   |
| UiB                               | 7,22*     | 6,393     |
| UiS                               | 6,686**   | 5,114     |
| UiT                               | 5,014     | -12,387   |
| USN                               | 5,237     | 12,81**   |
| VID                               | 7,856***  | 5,157     |
| Konstant                          | 29,558*** | 35,864*** |
| R <sup>2</sup>                    | 0,283     | 0,145     |
| Antall studenter                  | 239       | 220       |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Det er som nevnt ulikt antall makspoeng mellom gjennomføringene, og her har høst 2023 færre mulig oppnådde poeng enn høst 2024. Dette kan spille inn på forskjeller mellom styrken på sammenhengene. Vi ønsket derfor også å se på sannsynligheten for å få bestått eksamen for begge gjennomføringene.

Den samlede logistiske regresjonsmodellen viser at det er en positiv sammenheng mellom karaktersnitt fra videregående skole og sannsynligheten for å bestå eksamen. Oddsen for å bestå øker med 9,9 prosent for hver ekstra karakterenhet over gjennomsnittet, og denne sammenhengen er statistisk signifikant.

Modellen tar også hensyn til demografiske faktorer (alder og kjønn), tidspunkt for eksamen og hvilket studiested kandidatene tilhører. Resultatene viser at det er 32,2 prosent høyere odds for å bestå høst 2024 sammenlignet med høst 2023, men denne forskjellen er ikke statistisk signifikant.

Analysen avdekker betydelige forskjeller mellom studiestedene. Sammenlignet med referanseinstitusjonen (OsloMet) har studenter ved HVL over seks ganger høyere odds for å bestå (odds ratio = 6,17), mens studenter ved NTNU (2,5), UiA (2,1) og VID (3,3) ganger høyere odds for å bestå.

Tabell 22. Resultater av oddsanalyse av bestått og ikke-bestått på VGS-poeng med kontrollvariabler.

| Variabel                          | Odds ratio | Koeffisient | Standardfeil |
|-----------------------------------|------------|-------------|--------------|
| VGS-poeng                         | 1,099      | 0,094***    | 0,024        |
| Alder                             | 0,980      | -0,02       | 0,019        |
| Kvinne                            | 1,022      | 0,022       | 0,432        |
| H24 (ref. H23)                    | 1,322      | 0,279       | 0,223        |
| <b>Institusjon (ref. OsloMet)</b> |            |             |              |
| HiØ                               | 0,755      | -0,281      | 0,43         |
| HVL                               | 6,167      | 1,819***    | 0,535        |
| HVO                               | 2,014      | 0,7         | 0,509        |
| INN                               | 0,964      | -0,037      | 0,685        |
| Nord                              | 1,039      | 0,039       | 0,812        |
| NTNU                              | 2,505      | 0,918**     | 0,351        |
| UiA                               | 2,099      | 0,741*      | 0,342        |
| UiB                               | 1,368      | 0,313       | 0,568        |
| UiS                               | 2,099      | 0,741       | 0,423        |
| UiT                               | 1,012      | 0,012       | 0,924        |
| USN                               | 2,417      | 0,882       | 0,69         |
| VID                               | 3,341      | 1,206**     | 0,458        |
| <b>Konstant</b>                   | 0,747      | -0,292      | 0,478        |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Når vi ser på terskelverdiene, det vil si hvilket VGS-poeng som gir 50 prosent sannsynlighet for å bestå eksamen, finner vi forskjeller mellom utdanningsinstitusjonene. Ved HVL er terskelverdien lavest (23,6 poeng). Dette tilsier at en student med bare 23,6 i snitt fra videregående har like stor sjanse for å bestå som å stryke ved HVL. Når vi ser på andre siden av skalaen, finner vi HiØ og INN med terskelverdier på henholdsvis 45,9 og 43,3 poeng. Dette betyr at ved HiØ må en student ha nesten dobbelt så høyt karaktersnitt fra videregående som ved HVL for å ha samme sannsynlighet for å bestå eksamen.

Mellom disse ytterpunktene ser vi flere grupperinger: VID har en terskelverdi på 30,1, en gruppe bestående av NTNU, USN, UiA, UiS og HVO har terskelverdier mellom 33 og 36 poeng, UiB ligger på 39,6 poeng, mens Nord, UiT og OsloMet har terskelverdier rundt 42–43 poeng. Disse forskjellene antyder at institusjonelle faktorer har stor betydning for studentenes sannsynlighet for å lykkes, selv når vi kontrollerer for deres faglige utgangspunkt fra videregående skole.

Tabell 23. Resultat av predikerte terskelverdier for å bestå eksamen. N = 459.

| Institusjon | Terskelverdi (VGS-poeng) | Sannsynlighet |
|-------------|--------------------------|---------------|
| HVL         | 23,6                     | 50,0 %        |
| VID         | 30,1                     | 50,0 %        |
| NTNU        | 33,2                     | 50,1 %        |
| USN         | 33,6                     | 50,1 %        |
| UiA         | 35,1                     | 50,1 %        |
| UiS         | 35,1                     | 50,1 %        |
| HVO         | 35,5                     | 50,0 %        |
| UiB         | 39,6                     | 50,0 %        |
| Nord        | 42,5                     | 49,9 %        |
| UiT         | 42,8                     | 50,0 %        |
| OsloMet     | 42,9                     | 49,9 %        |
| INN         | 43,3                     | 49,9 %        |
| HiØ         | 45,9                     | 49,9 %        |

Notat: AIC=588,859, Log Likelihood: -277,429

### 9.3.1 Norskarakter

Vi brukte også her både lineær og logistisk regresjon for å undersøke sammenhengen mellom kandidatenes norskarakterer fra videregående skole og eksamensprestasjonene deres. I motsetning til analysen av VGS-poeng inkluderer denne analysen kun data fra eksamensgjennomføringen høst 2024. Her er 222 av 267 studenter med tilgjengelig bakgrunnsinformasjon om norskarakter. Under ser vi en oversikt over denne gruppens snittkarakter i norsk samt poengsum og beståttandel på nasjonal eksamen.

Tabell 24. Eksamensresultat høst 2024 og norskarakter fra videregående skole.

| Resultat      | Antall (andel)     | Gjennomsnittlig norskarakter | Gjennomsnittlig poengsum |
|---------------|--------------------|------------------------------|--------------------------|
| Bestått       | 139 (62,6 %)       | 4,2                          | 46,4                     |
| Ikke bestått  | 83 (37,4 %)        | 3,8                          | 27,5                     |
| <b>Totalt</b> | <b>222 (100 %)</b> | <b>4,1</b>                   | <b>39,3</b>              |

Den lineære regresjonsanalysen estimerer den statistiske sammenhengen mellom poengsum på eksamen og norskarakter, mens den logistiske regresjonen estimerer sannsynligheten for å bestå eksamen. For å sikre en mer konkret tolkning er norskarakteren sentrert rundt gjennomsnittet i analysene.

Analysen er gjennomført i tre steg: Først analyseres sammenhengen mellom norskarakter og resultat uten kontrollvariabler. Deretter inkluderes demografiske kontrollvariabler som alder og kjønn, før det til slutt kontrolleres for institusjonelle forskjeller.

Den enkle logistiske regresjonsmodellen undersøker hvordan norskarakter fra videregående skole henger sammen med sannsynligheten for å bestå nasjonal eksamen høst 2024. Modellen inkluderer kun norskarakter og tar ikke hensyn til andre faktorer som alder, kjønn eller utdanningsinstitusjon.

Resultatene viser at norskkarakter er en tydelig og statistisk signifikant variabel for om en kandidat består eksamen. For hver enhets økning i norskkarakter (på skala fra 1 til 6) øker oddsen for å bestå med omtrent 142 prosent. Dette tilsvarer en odds ratio på 2,42, som betyr at en kandidat med karakter 5 har over dobbelt så høy sannsynlighet for å bestå som en kandidat med karakter 4.

Modellen viser også at en kandidat med gjennomsnittlig norskkarakter (4,07) har en beregnet sannsynlighet på omtrent 63,6 prosent for å bestå eksamen. Dette stemmer godt overens med den faktiske andelen som bestod i utvalget (62,6 prosent).

Tabell 25. Resultater fra enkel logistisk regresjonsmodell av bestått og ikke bestått på norskkarakter.

| Variabel      | Koeffisient | Standardfeil | Odds ratio | 95% KI for odds ratio |
|---------------|-------------|--------------|------------|-----------------------|
| Konstant      | 0,559***    | 0,147        | 1,75       | -                     |
| Norskkarakter | 0,884***    | 0,208        | 2,42       | [1,61 - 3,65]         |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

I modell 2, som kontrollerer for alder og kjønn, viser resultatene at norskkarakter fortsatt er en sterk og signifikant variabel for å bestå eksamen. For hver enhets økning i norskkarakter øker oddsen for å bestå med omtrent 127 prosent (odds ratio = 2,27). Verken alder eller kjønn har signifikant effekt, noe som tyder på at disse faktorene har mindre betydning for eksamensutfallet enn norskkarakteren.

I modell 3, hvor det også kontrolleres for utdanningsinstitusjon, forblir effekten av norskkarakter stabil og signifikant (odds ratio = 2,21). Dette viser at sammenhengen mellom norskkarakter og bestått eksamen er solid selv når vi tar hensyn til både demografiske og utdanningsinstitusjonelle forskjeller. Blant utdanningsinstitusjonene skiller HVL seg ut med signifikant høyere sannsynlighet for å bestå sammenlignet med referanseinstitusjonen (HiØ).

Modell 3 gir det mest informative bildet av hva som påvirker sannsynligheten for å bestå nasjonal eksamen. Når både demografiske faktorer og utdanningsinstitusjon tas med i analysen, ser vi at variasjonen i beståttgrad forklares bedre enn i de enklere modellene. Samtidig er det verdt å merke seg at effekten av norskkarakter er bemerkelsesverdig stabil på tvers av alle modellene. Odds ratio varierer kun fra 2,42 til 2,21.

Tabell 26. Utvidet modell av sammenheng mellom norskkarakter og eksamensresultat. VID er tatt ut på grunn av kun 1 person med tilgjengelig norskkarakter.

| Variabel                          | Modell 2: + Demografi | Modell 3: + Institusjon |
|-----------------------------------|-----------------------|-------------------------|
|                                   | Odds ratio (SE)       | Odds ratio (SE)         |
| Norskkarakter                     | 2,27*** (0,214)       | 2,21*** (0,232)         |
| Alder                             | 0,97 (0,030)          | 0,98 (0,035)            |
| Kvinne                            | 1,45 (0,538)          | 1,46 (0,577)            |
| <b>Institusjon (ref. OsloMet)</b> |                       |                         |
| HiØ                               |                       | 0,57 (0,572)            |
| HVL                               |                       | 16,59* (1,085)          |
| HVO                               |                       | 0,43 (0,932)            |
| INN                               |                       | 0,82 (0,884)            |
| Nord                              |                       | 0,36 (1,272)            |
| NTNU                              |                       | 1,78 (0,478)            |
| UiA                               |                       | 1,75 (0,481)            |
| UiB                               |                       | 4,99 (1,129)            |
| UiS                               |                       | 2,72 (0,613)            |
| UiT                               |                       | 0,58 (0,869)            |
| USN                               |                       | 0,72 (1,322)            |
| <b>Konstant</b>                   | 1,24 (0,518)          | 0,89 (0,611)            |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ . VID er tatt ut på grunn av kun 1 person med tilgjengelig norskkarakter. Standardfeil i parentes med tre desimaler.

Samlet viser analysene at norskkarakter fra videregående skole har en tydelig sammenheng med sannsynligheten for å bestå nasjonal eksamen høst 2024. Denne sammenhengen holder seg stabil på tvers av alle modellene, også når det kontrolleres for alder, kjønn og utdanningsinstitusjon. Resultatene indikerer at både individuelle og institusjonelle forhold kan bidra til variasjon i eksamensutfall, men at norskkarakteren fremstår som en gjennomgående viktig forklaringsfaktor.

## 9.4 Bachelorbakgrunn

Regresjonsanalysen ble gjennomført med poengsum som avhengig variabel og bachelorgruppe og semester som uavhengige variabler. Det ble kjørt separate regresjonsmodeller for hvert semester samt en samlet modell med interaksjonsterm for å undersøke om effekten av bachelorbakgrunn varierer mellom semestrene.

For å finne ut om det var statistisk signifikante forskjeller i gjennomsnittlig poengsum mellom bachelorgruppene, ble det gjennomført en variansanalyse, også kalt ANOVA-test (*Analysis of Variance*). Denne testen sammenligner variasjonen mellom gruppene med variasjonen innad i hver gruppe. Hvis forskjellene mellom gruppene er store nok, kan testen vise at minst én gruppe skiller seg tydelig ut når det gjelder prestasjonsnivå.

I tillegg ble det brukt chi-kvadrat-tester for å undersøke om det var forskjeller i andelen som bestod mellom bachelorgruppene. Denne testen ser på om det er en statistisk sammenheng mellom hvilken bachelorbakgrunn studentene har og om de består eller ikke. Den gjør dette ved å sammenligne de observerte tallene med det man ville forventet dersom det ikke var noen forskjell mellom bachelorgruppene.

Den samlede regresjonsanalysen undersøkte sammenhengen mellom studentenes poengsum og bachelorbakgrunnen deres, samt hvilket semester eksamen ble gjennomført. Det ble benyttet to modeller: én enkel modell med kun bachelorgruppe som forklaringsvariabel og én utvidet modell som også inkluderte semester.

Modellen viser at hvilken bachelorutdanning kandidatene har, i liten grad forklarer forskjeller i poengsum mellom studentene. Høst 2023 forklarer utdanningsbakgrunn kun 0,03 prosent av variasjonen i resultatene, og det ble ikke funnet noen statistisk signifikante forskjeller mellom de ulike bachelorgruppene. I høst 2024 er forklaringskraften noe høyere på 1,8 prosent, og her viser analysen at kandidater med bachelor i sosialt arbeid har noe høyere poengsum enn kandidater med bachelor i barnevern. Forskjellen er statistisk signifikant. Det ble ikke funnet signifikante forskjeller mellom gruppen med bachelor i vernepleie eller andre profesjonsutdanninger sammenlignet med de øvrige.

Når vi inkluderer alder, kjønn og forskjeller mellom utdanningsinstitusjoner i analysen, øker modellens forklaringskraft betydelig. I høstsemesteret 2023 forklares 18,9 prosent av variasjonen i poengsum, og i høstsemesteret 2024 forklares 14,2 prosent. I 2023 er det først og fremst forskjeller mellom institusjonene som påvirker resultatene, og flere av disse forskjellene er statistisk signifikante. Det er fortsatt ingen tydelige forskjeller mellom bachelorgruppene. I 2024 ser vi at kandidater med bachelor i sosialt arbeid presterer noe bedre enn andre, og denne forskjellen er nær signifikant. I tillegg viser analysen at eldre kandidater i gjennomsnitt har noe lavere poengsum, og denne sammenhengen er statistisk signifikant.

Tabell 27. Resultater fra regresjonsanalyse av poengsum på bachelorgruppe med kontrollvariabler.

| Variabel                                 | Høst 2023 | Høst 2024 |
|------------------------------------------|-----------|-----------|
| Bachelorbackgrunn (referanse: Barnevern) |           |           |
| Sosialt arbeid                           | -0,45     | 3,53      |
| Vernepleie/Andre                         | -2,21     | 0,12      |
| <b>Kontrollvariabler</b>                 |           |           |
| Alder                                    | -0,01     | -0,21*    |
| Kvinne                                   | 3,19      | 1,51      |
| <b>Institusjon (ref. OsloMet)</b>        |           |           |
| UiB                                      | 6,86*     | 6,33      |
| UiT                                      | 8,78*     | -2,94     |
| NTNU                                     | 8,20***   | 3,08      |
| UiA                                      | 7,18**    | 6,4*      |
| HVL                                      | 10,73**   | 2,39      |
| NORD                                     | -7,97*    | -7,08     |
| INN                                      | -1,00     | -5,34     |
| UiS                                      | 5,41*     | 3,22      |
| USN                                      | 3,05      | 2,27      |
| HVO                                      | 3,89      | -3,48     |
| HiØ                                      | -6,59*    | -3,04     |
| VID                                      | 3,47      | -0,26     |
| Konstant                                 | 28,38***  | 35,04***  |
| R <sup>2</sup>                           | 0,189     | 0,142     |
| Antall studenter                         | 300       | 267       |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Den samlede modellen, som inkluderer data fra både høst 2023 og høst 2024, viser at sannsynligheten for å bestå eksamen har lite sammenheng med bachelorbakgrunn. Modellen viser ingen statistisk signifikante forskjeller mellom bachelorgruppene. Det er en svak tendens til at kandidater med bachelor i sosialt arbeid har høyere sannsynlighet for å bestå i høstsemesteret 2024 (interaksjonseffekt med odds ratio = 1,49), men denne forskjellen er ikke statistisk signifikant.

Det er imidlertid tydelige forskjeller mellom institusjonene. Studenter ved NTNU har 2,4 ganger høyere odds for å bestå, ved UiA er oddsen 2,6 ganger høyere og ved HVL er oddsen hele 4,6 ganger høyere sammenlignet med referanseinstitusjonen OsloMet. Studenter ved HiØ har derimot rundt 60 prosent lavere odds for å bestå (odds ratio = 0,41), og Nord viser en tilsvarende negativ tendens (odds ratio = 0,25).

Vi ser også at alder har en negativ sammenheng med sannsynligheten for å bestå. For hvert ekstra år reduseres oddsen med omtrent 2,7 prosent (odds ratio = 0,97).

Tabell 28 viser resultatene fra den logistiske regresjonsanalysen med interaksjonseffekter for sannsynligheten for å bestå eksamen.

Tabell 28. Resultater av oddsanalyse av bestått og ikke-bestått på bachelorbakgrunn med kontrollvariabler og interaksjonsledd.

| Variabel                                | Odds ratio | Koeffisient | Standardfeil |
|-----------------------------------------|------------|-------------|--------------|
| <b>Bachelorgrupper (ref. Barnevern)</b> |            |             |              |
| Sosialt arbeid                          | 0,89       | -0,12       | 0,35         |
| Vernepleie/Andre profesjoner            | 0,81       | -0,21       | 0,31         |
| H24 (ref. H23)                          | 1,04       | 0,04        | 0,25         |
| <b>Kontrollvariabler</b>                |            |             |              |
| Alder (sentrert)                        | 0,97       | -0,03*      | 0,01         |
| Kvinne                                  | 1,27       | 0,24        | 0,31         |
| <b>Institusjon (ref. OsloMet)</b>       |            |             |              |
| UiB                                     | 1,92       | 0,65        | 0,51         |
| UiT                                     | 1,12       | 0,11        | 0,47         |
| NTNU                                    | 2,41       | 0,88**      | 0,32         |
| UiA                                     | 2,63       | 0,97*       | 0,39         |
| HVL                                     | 4,56       | 1,52**      | 0,50         |
| NORD                                    | 0,25       | -1,37       | 0,81         |
| INN                                     | 0,60       | -0,52       | 0,55         |
| UiS                                     | 1,57       | 0,45        | 0,33         |
| USN                                     | 1,33       | 0,28        | 0,55         |
| HVO                                     | 1,33       | 0,29        | 0,49         |
| HiØ                                     | 0,41       | -0,90*      | 0,41         |
| VID                                     | 1,66       | 0,50        | 0,36         |
| <b>Interaksjonseffekter</b>             |            |             |              |
| Sosialt arbeid × H24                    | 1,49       | 0,40        | 0,53         |
| Vernepleie/Andre × H24                  | 0,99       | -0,01       | 0,42         |
| Konstant                                | 0,75       | -0,29       | 0,37         |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

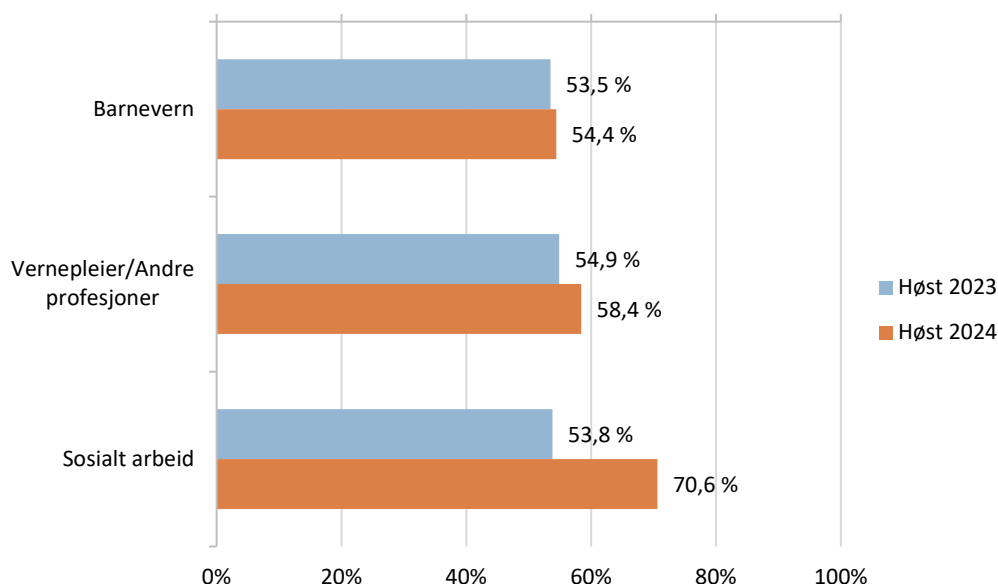


Regresjonsanalysen viser at bachelorbakgrunn alene har liten betydning for variasjonen i studentenes prestasjoner. De mest tydelige funnene i analysen er knyttet til forskjeller mellom utdanningsinstitusjonene. Selv om deskriptiv statistikk viser at studentene i høstsemesteret 2024 hadde noe høyere beståttprosent enn de i høstsemesteret 2023, er denne forskjellen ikke statistisk signifikant i den logistiske regresjonsmodellen (odds ratio = 1.04,  $p = 0.87$ ). Kandidater med bachelor i sosialt arbeid viser en tendens til å prestere bedre i 2024, mens gruppen med vernepleie eller andre profesjonsutdanninger har noe lavere poengsum i den samlede analysen. I tillegg viser analysen at eldre kandidater i gjennomsnitt presterer svakere, særlig i høstsemesteret 2024.

Det er verdt å merke seg at modellenes forklaringskraft øker betydelig når institusjonelle forskjeller inkluderes, fra 7,7 til 16,8 prosent, noe som understreker betydningen av institusjonsspesifikke faktorer for studentenes prestasjoner.

Figuren under viser beståttprosent for de tre bachelorgruppene høst 2023 og høst 2024.

Figur 24. Beståttandel høst 2023 og høst 2024 på tvers av bachelorbakgrunn.



Høst 2023 ligger beståttprosenten relativt jevnt mellom gruppene, med 53,5 prosent for barnevernspedagoger, 53,8 prosent for kandidater med bachelor i sosialt arbeid og 54,9 prosent for vernepleiere og andre profesjoner. Høst 2024 ser vi større forskjeller mellom gruppene, der særlig kandidater med bachelor i sosialt arbeid utmerker seg med en beståttprosent på 70,6 prosent, mens barnevernspedagoger ligger på 54,4 prosent og vernepleiere og andre profesjoner på 58,4 prosent.

## 9.5 Masterprogram

Regresjonsanalysen ble gjennomført med poengsum som avhengig variabel og semester og mastergradstype som uavhengige variabler. Det ble benyttet lineær regresjon for hvert enkelt semester samt en samlet modell med interaksjonsterm for å undersøke om effekten av utdanningsbakgrunn varierer mellom semestrene. I tillegg ble det gjennomført ANOVA for å teste hoved- og interaksjonseffekter for å undersøke forskjeller i beståttandel mellom gruppene.

### **Enkel modell (kun mastergradstype):**

Analysen viser at hvilken type mastergrad kandidatene har, bare i liten grad forklarer forskjeller i poengsum mellom studentene. I høstsemesteret 2023 forklarte utdanningsbakgrunnen kun 1,2 prosent av variasjonen i resultatene ( $R^2 = 0,012$ ). Forskjellene mellom de ulike masterkategoriene er her nesten statistisk signifikante ( $\beta = 2,37$ ,  $p = 0,056$ ).

I høstsemesteret 2024 var sammenhengen noe sterkere, med en forklaringskraft på 1,7 prosent ( $R^2 = 0,017$ ). Her viser analysen at studentene fra master i barnevern oppnådde signifikant høyere poengsum enn studentene fra master i barnevernsarbeid ( $\beta = 3,04$ ,  $p < 0,05$ ).

### **Full modell (med kontrollvariabler):**

Når vi tar hensyn til alder, kjønn og forskjeller mellom utdanningsinstitusjoner, øker modellens evne til å forklare variasjonen i studentenes poengsummer betydelig. I høstsemesteret 2023 forklares 19,7 prosent av variasjonen ( $R^2 = 0,197$ ), mens det i høstsemesteret 2024 forklarer 13,2 prosent ( $R^2 = 0,132$ ).

Høst 2023 viser analysen at både typen mastergrad og hvilken institusjon studenten kommer fra har betydning for resultatene. Studenter fra master i barnevern presterer signifikant bedre enn de andre ( $\beta = 3,47$ ,  $p < 0,05$ ), og flere institusjoner skiller seg ut med statistisk signifikante forskjeller.

Høst 2024 er det ikke en tydelig sammenheng mellom mastergradstype og poengsum ( $\beta = -0,10$ ,  $p = 0,960$ ). Derimot ser vi at alder har en svak negativ sammenheng med prestasjonene ( $\beta = -0,21$ ,  $p = 0,051$ ). Institusjonelle forskjeller fortsetter å være viktige, særlig for institusjon 201, hvor kandidatene oppnår betydelig høyere poengsummer enn andre ( $\beta = 6,31$ ,  $p < 0,01$ ).

Tabell 29. Resultater fra regresjonsanalyse av poengsum på masterprogram med kontrollvariabler.

| Variabel                          | Høst 2023 | Høst 2024 |
|-----------------------------------|-----------|-----------|
| Master i barnevern                | 3,47*     | -0,1      |
| <b>Kontrollvariabler</b>          |           |           |
| Alder                             | 0,00      | -0,21     |
| Kvinne                            | 2,78      | 1,68      |
| <b>Institusjon (ref. OsloMet)</b> |           |           |
| UiB                               | 6,46*     | 7,25      |
| UiT                               | 9,97**    | -2,74     |
| NTNU                              | 8,76***   | 3,32      |
| UiA                               | 3,96      | 6,31**    |
| HVL                               | 12,78***  | 4,5       |
| NORD                              | -5,71     | -7,38     |
| INN                               | -2,45     | -3,62     |
| UiS                               | 7,37***   | 3,61      |
| USN                               | 4,70      | 2,57      |
| HVO                               | 5,43      | -2,73     |
| HiØ                               | -4,79     | -3,16     |
| VID                               | 3,02      | -0,49     |
| R <sup>2</sup>                    | 0,197     | 0,132     |
| Antall studenter                  | 300       | 267       |
| Konstant                          | 26,31***  | 35,22***  |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

Samlet regresjonsanalyse med interaksjonseffekt ble gjennomført med bestått/ikke-bestått som avhengig variabel for å undersøke hvordan mastergradstype påvirker sannsynligheten for å bestå eksamen, og om denne effekten varierer mellom gjennomføringene.

Analysen viser at studenter fra master i barnevern har 32 prosent høyere odds for å bestå sammenlignet med studenter fra master i barnevernsarbeid. Dette tilsvarer en odds ratio på 1,32. I dette tilfellet er forskjellen imidlertid ikke statistisk signifikant, og det er derfor usikkert om den skyldes reelle forskjeller eller tilfeldigheter.

Studentene i høstsemesteret 2024 hadde noe høyere odds for å bestå enn studentene i høstsemesteret 2023 (odds ratio = 1,16), men også denne forskjellen er ikke signifikant.

Den faktoren som har størst betydning i modellen er alder. Her ser vi at høyere alder henger sammen med lavere sannsynlighet for å bestå. For hvert år økning i alder reduseres oddsen for å bestå med omtrent 3 prosent (odds ratio = 0,97).

Interaksjonseffekten mellom mastergradstype og semester er tilnærmet null (odds ratio = 0,99) og klart ikke signifikant. Dette tyder på at effekten av mastergradstype på beståttandel er svært lik i begge semestrene.

Tabell 30. Resultater av oddsanalyse av bestått og ikke-bestått på masterprogram med kontrollvariabler og interaksjonsledd.

| Variabel                  | Odds ratio | Koeffisient | Standardfeil |
|---------------------------|------------|-------------|--------------|
| Semester (H24)            | 1,16       | 0,14        | 0,22         |
| Master i barnevern        | 1,32       | 0,28        | 0,24         |
| <b>Kontrollvariabler</b>  |            |             |              |
| Alder                     | 0,97       | -0,03**     | 0,01         |
| Kvinne                    | 1,27       | 0,24        | 0,30         |
| <b>Interaksjonseffekt</b> |            |             |              |
| Master i barnevern × H24  | 0,99       | -0,01       | 0,35         |
| AIC                       | 772,91     |             |              |
| <b>Antall studenter</b>   | 567        |             |              |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

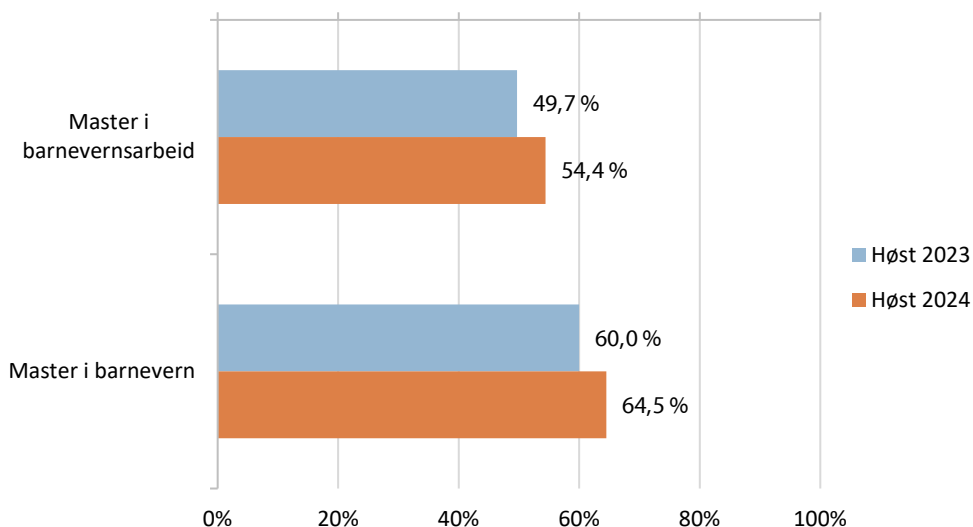
Resultatene viser at det ikke er noen signifikant samspill mellom mastergradstype og semester, og modellen med interaksjon gir ikke bedre forklaring enn modellen uten. Dette støtter konklusjonen om at effekten av mastergradstype på beståttandel er stabil mellom semestrene. Derimot gir modellen som inkluderer forskjeller mellom utdanningsinstitusjoner, også her en bedre forklaring på studentenes resultater.

Separate analyser for hvert semester viser ulike mønstre. Høst 2023 har kandidater med «Master i barnevern» høyere odds for å bestå enn de med «Master i barnevernsarbeid» (odds ratio = 1,80, men denne forskjellen er ikke signifikant. Høst 2024 er forskjellen mellom masterkategoriene minimal (odds ratio = 1,09) og tydelig ikke signifikant.

Når det gjelder institusjonelle forskjeller, viser analysen at i høst 2023 hadde kandidater fra NTNU signifikant høyere beståttandel. I høst 2024 var det HVL som utmerket seg med signifikant høyere beståttandel. Disse funnene understreker at utdanningsinstitusjonen spiller en viktig rolle for studentenes sannsynlighet for å bestå.

Figuren under viser beståttprosent for to mastergrupper fordelt på høst 23 og høst 24.

Figur 25. Beståttandel fordelt på mastergrupper høst 2023 og høst 2024.



I høst 2023 ser vi en tydelig forskjell i beståttprosent mellom mastergruppene, der kandidater med master i barnevern har en betydelig høyere beståttprosent på 60 prosent sammenlignet med kandidater med master i barnevernsarbeid som ligger på 49,7 prosent. I høst 2024 opprettholdes denne forskjellen, og begge gruppene viser en høyere andel bestått. Kandidater med master i barnevern har her en 64,5 beståttprosent, mens kandidater med master i barnevernsarbeid har 54,4 beståttprosent. Dette indikerer at forskjellen mellom de to mastergruppene har holdt seg relativt stabil på rundt 10 prosentpoeng fra høst 2023 til høst 2024, samtidig som begge gruppene har opplevd en positiv utvikling i beståttprosent.

## 9.6 Inneværende karaktersnitt på masterprogrammet

Noen undervisere mente å se ved første gjennomføring av eksamen at de studentene som ellers gjorde det bra i andre emner, ikke presterte så godt på nasjonal eksamen. For å undersøke om dette medførte riktighet og hvordan inneværende karaktersnitt i avlagte emner på masterstudiet påvirker eksamensresultatene, er det gjennomført regresjonsanalyser basert på data fra høst 2023 og høst 2024. Den samlede modellen ser på to ulike utfall: poengsum på eksamen og sannsynligheten for å bestå.

Poengsummen analyseres med lineær regresjon, som viser hvordan eksamensresultatet endrer seg når karaktersnittet varierer. Sannsynligheten for å bestå analyseres med logistisk regresjon, som gir innsikt i hvor stor sjanse studentene har for å bestå avhengig av karaktersnittet.

Analysene ble først gjennomført separat for hvert semester, både med og uten kontroll for forskjeller mellom utdanningsinstitusjoner. Deretter ble det laget en samlet modell som inkluderer data fra både høst 2023 og høst 2024. I denne modellen er både semestereffekter og institusjonelle forskjeller tatt med som forklaringsvariabler.

I den enkle modellen for høst 2023 viser analysen at karaktersnitt har en tydelig positiv sammenheng med eksamensresultatet. For hver karakterenhet øker poengsummen med 3,94 poeng, og modellen forklarer 17,5 prosent av variasjonen i resultatene. I høst 2024 er

sammenhengen noe svakere. Her øker poengsummen med 2,67 poeng per karakterenhet, og modellen forklarer 6 prosent av variasjonen.

Tabell 31. Resultat av enkel regresjonsanalyse av poengsum på inneværende karaktersnitt.

| Gjennomføring | Antall studenter | Koeffisient | R <sup>2</sup> |
|---------------|------------------|-------------|----------------|
| Høst 2023     | 283              | 3,941***    | 0,175          |
| Høst 2024     | 248              | 2,667***    | 0,060          |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$ . Det er kjørt separate analyser for hver gjennomføring.

I høst 2023 viser den fulle modellen at poengsummen på eksamen øker med 4,3 poeng for hver karakterenhet i karaktersnitt. Dette er en statistisk sikker sammenheng, og modellen forklarer 31,4 prosent av variasjonen i resultatene. Flere utdanningsinstitusjoner skiller seg ut med høyere poengsum i snitt sammenlignet med referanseinstitusjonen, blant andre HVO (12, 2 poeng), NTNU (10,6 poeng) og UiB (10,5 poeng). Forskjellene er statistisk signifikante.

Høst 2024 viser den fulle modellen at poengsummen på eksamen øker med 3,2 poeng for hver karakterenhet i karaktersnitt. Dette er en statistisk sikker sammenheng, og modellen forklarer 16,1 prosent av variasjonen i resultatene. Det er også forskjeller mellom institusjonene: Studenter ved NTNU har i snitt 4,8 poeng mer enn referanseinstitusjonen, og studenter ved UiA har 7,4 poeng mer. Studenter ved UiT skiller seg ut med lavere resultater og har i snitt 13,1 poeng mindre enn referansegruppen.

Tabell 32. Resultater av regresjonsanalyse av poengsum på inneværende karaktersnitt med kontrollvariabler.

| Variabel                          | Høst 2023 | Høst 2024 |
|-----------------------------------|-----------|-----------|
| Karaktersnitt                     | 4,300***  | 3,103***  |
| <b>Kontrollvariabler</b>          |           |           |
| Alder                             | -0,002    | -0,141    |
| Kvinne                            | 3,639     | 1,001     |
| <b>Institusjon (ref. OsloMet)</b> |           |           |
| HiØ                               | 2,701     | 1,591     |
| HVL                               | 9,761**   | 3,914     |
| HVO                               | 12,778*** | 0,539     |
| INN                               | 8,137*    | 9,416     |
| Nord                              | 6,251     | 10,538    |
| NTNU                              | 10,612*** | 4,312*    |
| UiA                               | 8,068***  | 7,246***  |
| UiB                               | 10,191*** | 0,48      |
| UiS                               | 9,833***  | 8,363     |
| UiT                               | 8,303**   | -11,728   |
| USN                               | 6,091*    | 7,283*    |
| VID                               | 6,108***  | 0,431     |
| R <sup>2</sup>                    | 0,323     | 0,168     |
| Antall studenter                  | 283       | 248       |

Notat: Tegnsetting for signifikansnivå: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$

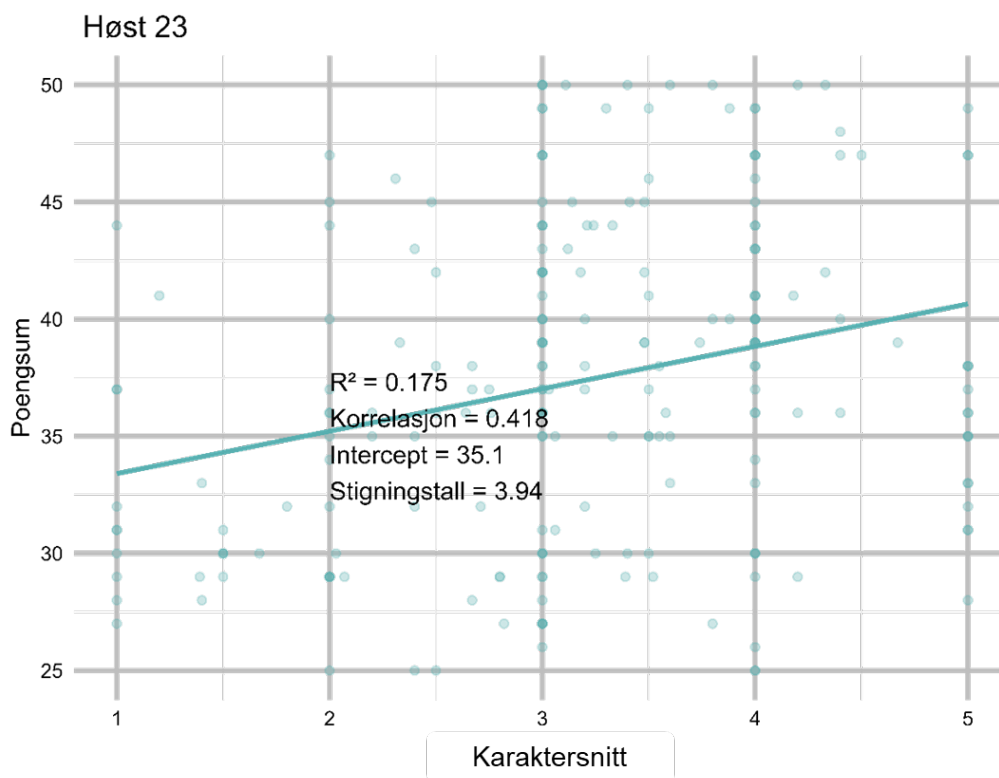
## Korrelasjonsanalyse

Korrelasjonsanalysen viser at det er en positiv sammenheng mellom karaktersnitt og eksamensresultat i begge semestre. I høstsemesteret 2023 er korrelasjonskoeffisienten 0,418, og i høstsemesteret 2024 er den 0,245. Begge sammenhengene er statistisk signifikante, noe som betyr at det er svært lite sannsynlig at resultatene har oppstått tilfeldig. Dette gir støtte til at det finnes en reell og målbar sammenheng mellom studentenes karaktersnitt og prestasjonene deres på eksamen.

Den lineære regresjonsmodellen viser at karaktersnitt har en positiv sammenheng med eksamensresultatet. I høst 2023 forklarer modellen 17,5 prosent av variasjonen i poengsum, mens den i høst 2024 forklarer 6 prosent. Dette betyr at karaktersnitt alene gir en viss forklaring på forskjellene i eksamensresultater mellom kandidatene, men at størstedelen av variasjonen skyldes andre forhold, enten observerbare eller uobserverbare, som ikke er inkludert i modellen.

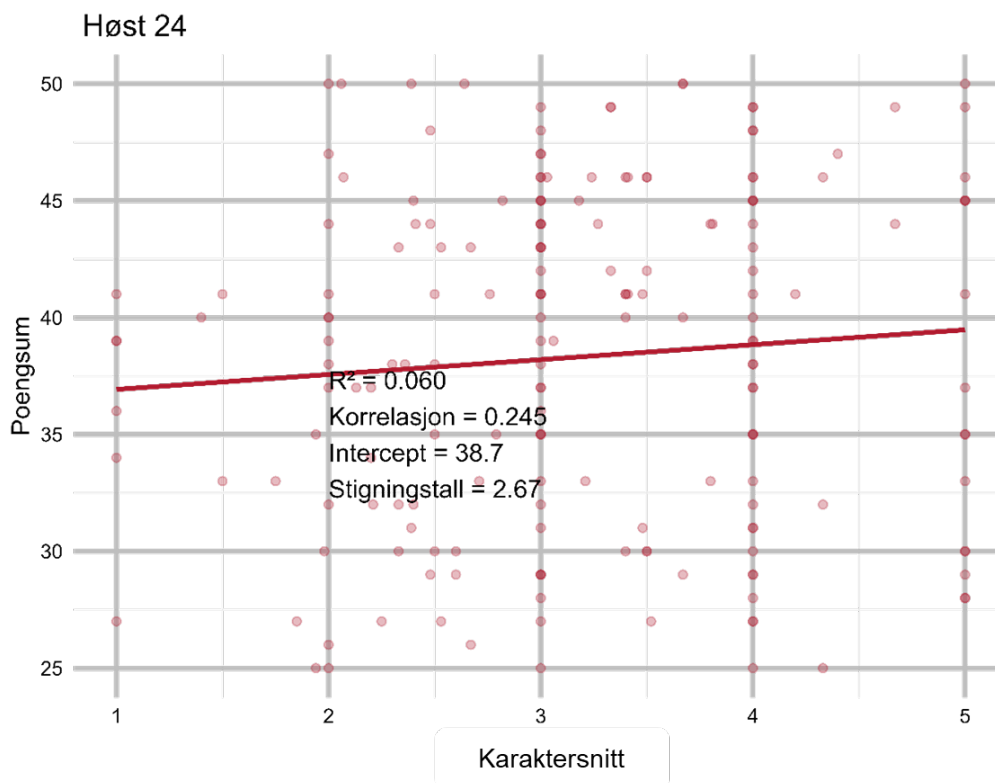
Regresjonsplottet for høstsemesteret 2023 viser spredningen av karakterer og eksamensresultater, med en stigende regresjonslinje som indikerer en positiv sammenheng. Datapunktene viser en tydelig, men moderat spredning rundt regresjonslinjen, noe som gjenspeiles i modellens forklaringskraft.

Figur 26. Visualisering av korrelasjonsanalyse mellom poengsum og inneværende karaktersnitt høst 2023.



I høst 2024 viser regresjonsplottet også en positiv sammenheng, men med en svakere stigning i regresjonslinjen sammenlignet med høst 2023. Datapunktene har større spredning rundt regresjonslinjen, noe som reflekteres i den lavere forklaringskraften til modellen.

Figur 27. Visualisering av korrelasjonsanalyse mellom poengsum og inneværende karaktersnitt høst 2024.



Samlet sett viser analysene at det er en positiv sammenheng mellom karaktersnitt og poengsum på eksamen, men styrken på denne sammenhengen varierer mellom de to gjennomføringene.

## 9.7 Oppsummering

Oppsummert kan vi se flere tendenser i hva som påvirker studentenes eksamensprestasjoner. Vi ser tydelig at studentenes resultater på nasjonal eksamen i rettsanvendelse har sterk sammenheng med inneværende karaktersnitt.

Når det gjelder tidligere karakterer fra videregående skole, var det en antakelse, basert på tidligere studier, at dette ville ha stor betydning for resultatet på nasjonal eksamen. Vi fant at det var en positiv sammenheng mellom VGS-poeng og eksamensresultat for barnevernsstudentene, og at denne sammenhengen var statistisk signifikant høst 2023, men ikke høst 2024. Det antyder at det er litt uklart hvorvidt vi kan slå fast at VGS-poeng har betydning for eksamensresultatet til studentene på generell basis. Vi kan forsøke å forklare dette med at kompetanse fra praksisfeltet og erfaring med juridiske problemstillinger fra tidligere studier eller arbeid er viktigere faktorer for om du klarer å forstå og besvare caseoppgaven på eksamen.

Utdanningsinstitusjon viser seg som en viktig forklaringsfaktor i alle analyser. Dette er riktignok kun basert på kullene i høst 2023 og høst 2024, og det kan være andre faktorer som spiller inn på resultatene deres utover utdanningsinstitusjonen de tilhører, som for eksempel opptakskrav. Vi ser ulike tendenser i høst 2023 og høst 2024 innenfor samme institusjon, noe som antyder at studentkullene kan variere.



På en annen side ser vi at norskkarakter fra VGS har nokså stor betydning for eksamenspoeng høst 2024. Dette tilsier at det er forkunnskaper som studentene har som kan påvirke resultatet deres utover institusjonens undervisning og egenlæring. Skrivetrening kan være en mulig løsning for studenter med svakere norskkarakterer fra VGS.

Samtidig er det viktig å bemerke begrensninger ved analysene. Vi ser kun på to gjennomføringer, hvorav grupperinger på institusjonsnivå og bachelornivå gir varierende størrelser noe som kan gi tilfeldige variasjoner. Dette viser seg også i forklaringskraften til flere av modellene. Videregående karakter, selv når vi kontrollerer for alder, kjønn og utdanningsinstitusjon, forklarer henholdsvis 28 og 24,5 prosent av variasjonen i dataene. Det antyder at det er flere aspekter som kan påvirke resultatet til studentene.

Flere analyser over lengre tidsperioder trengs for å styrke sannsynligheten for at det er en sammenheng mellom variablene vi undersøker her.

## 10 Eksamenstema og studentenes prestasjoner

I dette kapittelet presenteres en kort tematisk analyse av studentens prestasjoner på eksamen høsten 2023 og høsten 2024. Nasjonal eksamen i rettsanvendelse for barnevernsfaglige mastere er bygget opp rundt en casebasert oppgave. Studentene vurderes etter et sett med vurderingsrubrikker som speiler de sentrale elementene i juridisk metode. Rubrikkene er utviklet for å sikre enhetlig vurdering på tvers av institusjoner og for å gi detaljert innsikt i hvilke deler av den juridiske metoden studentene behersker godt og hvor det er utfordringer. Vi trekker veksler på studentenes vurderinger av læringsutbytte (Karlsen & Pedersen, 2025) og sensorenes vurderinger av studentenes besvarelser.

Basert på vurderingsrubrikkene og sammenstillinger av studentenes uttelling kan vi si noe om hvilke deler av den juridiske metoden og rettsanvendelsen studentene behersker på eksamen og hvilke elementer de ikke behersker.

Figur 28. Eksempel på vurderingsrubrikker i sensurskjemaet høst 2024 (1–11 av totalt 58 rubrikker).

|    | A                                                                                                                                                                                     | B                                                                                                                               | C                                                                   | D                                                 | E                                          | G        | H         | I                                                |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|--------------------------------------------|----------|-----------|--------------------------------------------------|
| 1  | <b>Nasjonal deleksamen i rettsanvendelse i barnevernsarbeid, 4. desember 2024</b>                                                                                                     |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 2  | Skriv kun i fargede ruter. Sensorpar (som f.eks 2bc) i C7. Sluttvurdering registreres i fellesskap på linje 79 etter diskusjonen på kal 16. desember.                                 |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 3  | Eventuelle ytterligere kommentarer til studenten kan gis i kandidatens fjerde kolonne. Samsensur og eventuelle kommentarer vil fungelse, så dere ikke trenger å skrive slike i ette   |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 4  | Protokoll genereres automatisk, signeres av begge sensorer.                                                                                                                           |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 5  | Send både sensurskjema og protokoll til deleksamen-barnevern@nokut.no senest fredag 3. januar. NOKUT bruker sensorenes individuellelekun til analyseformål, og de vil ikke viderefors |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 6  |                                                                                                                                                                                       |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 7  | <b>Sensorpar:</b>                                                                                                                                                                     |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 8  |                                                                                                                                                                                       |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 9  |                                                                                                                                                                                       |                                                                                                                                 | Kandidaten svarer ikke på spørsmålet, eller svarer feil/mangelfullt | Kandidaten svarer tilfredsstillende på spørsmålet | Kandidaten svarer utfyllende på spørsmålet | Sensor B | Samsensur | Eventuell kommentar (kan formidles til kandidat) |
|    |                                                                                                                                                                                       |                                                                                                                                 | 0 poeng                                                             | 1 poeng                                           | 2 poeng                                    |          |           |                                                  |
| 14 | <b>1 Er Mari part? (Inntil 13 poeng)</b>                                                                                                                                              |                                                                                                                                 |                                                                     |                                                   |                                            |          |           |                                                  |
| 15 | 1                                                                                                                                                                                     | Identifiserer og formulerer problemstillingen                                                                                   | Formulerer ikke problemstillingen                                   | Formulerer problemstillingen                      | –                                          |          |           |                                                  |
| 16 | 2                                                                                                                                                                                     | Viser til bvl. § 12-2                                                                                                           | Viser ikke til bestemmelsen                                         | Viser til bestemmelsen                            | –                                          |          |           |                                                  |
| 17 | 3                                                                                                                                                                                     | Viser til fvl. § 2 første ledd bokstav e                                                                                        | Viser ikke til bestemmelsen                                         | Viser til bestemmelsen                            | –                                          |          |           |                                                  |
| 18 | 4                                                                                                                                                                                     | Klargjør at det er to alternativer i fvl. § 2 første ledd bokstav e                                                             | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
| 19 | 5                                                                                                                                                                                     | Stiller spørsmål ved om Mari er «en person som en avgjørelse retter seg mot», fvl. § 2 første ledd bokstav e, første alternativ | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
| 20 | 6                                                                                                                                                                                     | Tolker første alternativ; part i saken er «en person som en avgjørelse retter seg mot» fvl. § 2 første ledd bokstav e           | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
| 21 | 7                                                                                                                                                                                     | Bruker relevant faktum i drøftingen av om Mari er «en person som en avgjørelse retter seg mot»                                  | Bruker ikke relevant faktum                                         | Bruker relevant faktum                            | –                                          |          |           |                                                  |
| 22 | 8                                                                                                                                                                                     | Konkluderer med hensyn til om avgjørelsen retter seg mot Mari                                                                   | Ingen konklusjon/feil konklusjon                                    | Riktig konklusjon                                 | –                                          |          |           |                                                  |
| 23 | 9                                                                                                                                                                                     | Stiller spørsmål ved om «saken ellers direkte gjelder» Mari, jf. fvl. § 2 bokstav e andre alternativ                            | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
| 24 | 10                                                                                                                                                                                    | Tolker ordlyden «som saken ellers direkte gjelder», jf. fvl. § 2 første ledd bokstav e andre alternativ                         | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
| 25 | 11                                                                                                                                                                                    | Bruker relevant faktum i drøftingen av om «saken ellers direkte gjelder» Mari                                                   | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
|    |                                                                                                                                                                                       | Konkluderer om saken ellers direkte gjelder Mari                                                                                | Blir ikke gjort                                                     | Blir gjort                                        | –                                          |          |           |                                                  |
|    | <b>Vurderingsskjema</b>                                                                                                                                                               |                                                                                                                                 | Protokoll                                                           | +                                                 |                                            |          |           |                                                  |

Hver rubrikk kan gi 0,1 eller 2 poeng, og rubrikkscore beregnes ved andel av mulige poeng per kategori. Dette gir grunnlag for å analysere studentenes juridiske metodeforståelse og rettsanvendelse.

Figur 29. Eksempel på sammenstilling av vurderingsrubrikker nasjonalt.

|                |               |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
|----------------|---------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| <b>Høst 23</b> | Rubrikk       | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   |
|                | Max poeng     | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 2    | 2    | 1    | 1    | 1    | 2    | 2    | 1    | 1    | 1    |
|                | Kategori      | 1    | 2    | 3    | 3    | 1    | 3    | 4    | 1    | 3    | 4    | 5    | 1    | 3    | 4    | 5    | 5    | 1    | 3    | 4    | 5    | 7    | 7    |
|                | Antall 0      | 10   | 6    | 62   | 76   | 77   | 75   | 110  | 116  | 114  | 80   | 102  | 113  | 95   | 77   | 100  | 139  | 92   | 126  | 99   | 117  | 173  | 186  |
|                | Antall 1      | 290  | 294  | 238  | 224  | 223  | 225  | 190  | 184  | 186  | 219  | 198  | 187  | 148  | 161  | 200  | 161  | 208  | 149  | 167  | 183  | 126  | 114  |
|                | Antall 2      | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 56   | 62   | 0    | 0    | 0    | 24   | 34   | 0    | 0    | 0    |
|                | Oppg. Snitt   | 0,97 | 0,98 | 0,79 | 0,75 | 0,74 | 0,75 | 0,63 | 0,61 | 0,62 | 0,73 | 0,66 | 0,62 | 0,87 | 0,95 | 0,67 | 0,54 | 0,69 | 0,66 | 0,78 | 0,61 | 0,42 | 0,38 |
|                | Rubrikk score | 0,97 | 0,98 | 0,79 | 0,75 | 0,74 | 0,75 | 0,63 | 0,61 | 0,62 | 0,73 | 0,66 | 0,62 | 0,43 | 0,48 | 0,67 | 0,54 | 0,69 | 0,33 | 0,39 | 0,61 | 0,42 | 0,38 |
| <b>Høst 24</b> | Rubrikk       | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   |
|                | Maks poeng    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 2    | 2    | 1    | 2    | 2    |
|                | Kategori      | 1    | 2    | 2    | 3    | 1    | 3    | 4    | 5    | 1    | 3    | 4    | 5    | 5    | 1    | 2    | 1    | 2    | 3    | 4    | 2    | 3    | 4    |
|                | Antall 0      | 2    | 17   | 4    | 70   | 121  | 163  | 108  | 116  | 127  | 153  | 104  | 122  | 31   | 16   | 26   | 111  | 100  | 158  | 166  | 106  | 176  | 119  |
|                | Antall 1      | 265  | 250  | 263  | 196  | 146  | 104  | 159  | 151  | 140  | 114  | 163  | 145  | 236  | 251  | 241  | 156  | 167  | 89   | 92   | 161  | 79   | 138  |
|                | Antall 2      | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 20   | 9    | 0    | 12   | 10   |
|                | Oppg. snitt   | 0,99 | 0,94 | 0,99 | 0,73 | 0,55 | 0,39 | 0,60 | 0,57 | 0,52 | 0,43 | 0,61 | 0,54 | 0,88 | 0,94 | 0,90 | 0,58 | 0,63 | 0,48 | 0,41 | 0,60 | 0,39 | 0,58 |
|                | Rubrikk score | 0,99 | 0,94 | 0,99 | 0,73 | 0,55 | 0,39 | 0,60 | 0,57 | 0,52 | 0,43 | 0,61 | 0,54 | 0,88 | 0,94 | 0,90 | 0,58 | 0,63 | 0,24 | 0,21 | 0,60 | 0,19 | 0,30 |

Vi har kategorisert vurderingsrubrikkene i åtte hovedkategorier. Dette er samme inndeling som er gjort i kapittel 5.

1. Identifisere og formulere rettsspørsmål (og underspørsmål)
2. Finne relevante rettskilder
3. Klarlegge juss (tolke vilkår)
4. Sammenlikne juss og faktum (subsumere og drøfte)
5. Konkludere (underproblemstilling og hovedproblemstillinger)
6. Fastslår og klargjør et enkelt rettsspørsmål
7. Overordnede rettsprinsipper (trekker inn barnets beste, medvirkning, minste inngrep osv.)
8. Totalvurdering (utviser god juridisk metode og godt språk)

Eksempler på kategorisering av fra rubrikker høsteksamen 2023:

10. Bruker relevant faktum for å vurdere om vilkåret i bvl. § 6-2 første ledd bokstav b) er oppfylt (1 poeng, **temakategori 4**)
11. Konkluderer mht. om vilkåret i bvl. § 6-2 første ledd bokstav b) er oppfylt (1 poeng, **temakategori 5**)
12. Formulerer underproblemstilling om vilkåret i bvl. § 6-2 første ledd bokstav c) annen form for utpreget normløs atferd er oppfylt (1 poeng, **temakategori 1**)
13. Tolker ordlyden i bvl. § 6-2 første ledd bokstav c) (2 poeng, **temakategori 3**)

## 10.1 Studentenes prestasjoner

I rapporten «Studentstemmen: om undervisning, emneinnhold og vurdering i rettsanvendelse for barnevernsfaglige masterprogram» fremgår det at studentene samlet sett opplever et høyt læringsutbytte (Karlsen & Pedersen, 2025). Analysene av studentenes eksamensresultater viser at vurderingsrubrikkene effektivt skiller mellom kandidater som består og de som ikke består. Samtidig identifiseres det enkelte temaer som studentene behersker godt, og andre områder hvor det fortsatt er utfordringer.

Ved sammenligning av oppnådde poeng på rubrikkene fra høsten 2023 og høsten 2024, ser vi noe variasjon. Dette kan ses i sammenheng med at det er ulike eksamener, case og studentkull. Til tross for disse forskjellene, viser tabellene fra eksamensgjennomføringene noen gjennomgående resultater på tvers av kull og eksamener.

Figur 30 under viser kategoriseringer og poengsetting per tema. De kategoriene som er vanskeligst for studentene er markert med farge.

Figur 30. Eksempel på sammenstilling av vurderingsrubrikker per tema nasjonalt.

| Kategori                                                                                     | Snittscore h23 | Snitt rubrikk h23 | Antall 2p | Snittscore h24 | Snitt rubrikk h24 | Antall 2p |
|----------------------------------------------------------------------------------------------|----------------|-------------------|-----------|----------------|-------------------|-----------|
| 1. Identifisere og formulere rettsspørsmål (og underspørsmål)                                | 0,72           | 0,72              | 0         | 0,75           | 0,75              | 0         |
| 2. Finne relevante rettskilder                                                               | 0,70           | 0,70              | 0         | 0,74           | 0,74              | 0         |
| 3. Klarlegge juss (tolke vilkår)                                                             | 0,66           | 0,57              | 3         | 0,53           | 0,37              | 6         |
| 4. Sammenlikne juss og faktum (subsumere og drøfte)                                          | 0,59           | 0,48              | 3         | 0,63           | 0,47              | 4         |
| 5. Konkludere (underproblemstilling og hovedproblemstillinger)                               | 0,65           | 0,65              | 0         | 0,67           | 0,67              | 0         |
| 6. Fastslå og klargjør et enkelt rettsspørsmål                                               | 0,53           | 0,53              | 0         | 0,73           | 0,73              | 0         |
| 7. Overordnede rettsprinsipper (trekker inn barnets beste, medvirkning, minste inngrep osv.) | 0,44           | 0,44              | 0         | 0,55           | 0,39              | 1         |
| 8. Totalvurdering (utviser god juridisk metode og godt språk)                                | 0,54           | 0,46              | 1         | 0,81           | 0,41              | 2         |

### 10.1.1 Studentene identifiserer rettsspørsmål og relevante rettskilder

Studentene har gjennomgående gode ferdigheter i å identifisere juridiske problemstillinger og finne frem til relevante rettskilder. Innledningsvis ble det uttrykt bekymring for om studentene ville oppleve det som for vanskelig å selvstendig identifisere juridiske problemstillinger. Erfaringene viser imidlertid at dette er noe studentene mestrer godt. De klarer også å finne frem til relevante rettskilder og lovparagrafer.

Det er likevel ikke alle studenter som tydeliggjør underspørsmål og -problemstillinger, og noen behandler dette kun implisitt. Dette var et tema som ble diskutert i sensormøter, hvordan slike besvarelser skulle vurderes. Vurderingspraksis ble etablert: Dersom slike spørsmål ble behandlet implisitt, skulle dette likevel gi uttelling.

Studentene oppnår også relativt høye resultater når det gjelder å trekke konklusjoner, både på hoved- og underproblemstillinger (kategori 5). Ifølge Karlsen og Pedersen (2025) oppgir 64 prosent av studentene selv at de i stor grad kunne konkludere, mens kun 4 prosent opplever at de i liten grad kunne gjøre dette. Enkelte sensorer påpeker at noen studenter konkluderer riktig, men ikke alltid på grunnlag av juridisk metode.

### 10.1.2 Studentenes utfordringer: tolke vilkår og barnets beste

På tvers av eksamensgjennomføringene er det særlig to aspekter som er utfordrende for studentene, hvor de gjennomgående opplever eksamen som vanskelig og oppnår lav uttelling i vurderingsrubrikkene. Dette gjelder tolkning av vilkår – altså evnen til å forstå og tolke lovteksten – samt det vi har definert som *overordnede rettsprinsipper* (kategori 7),

som omfatter vurderinger av barnets beste, barnets rett til medvirkning og «minste inngreps prinsipp».<sup>33</sup>

Den mest krevende delen av juridisk metode for studentene er å tolke vilkår og klarlegge jussen (kategori 3). Mange studenter har utfordringer med å lese lovteksten, forstå begreper og skille mellom alternative og kumulative vilkår. Når det gjelder dette temaet, ser vi også en nedgang i studentenes prestasjoner fra høsten 2023 til høsten 2024. Det er tydelig at selv om studentene identifiserer juridiske problemstillinger og relevante rettskilder, er det utfordrende for dem å tolke loven og rettsanvendelse.

Subsumsjon og drøfting (kategori 4) er heller ikke noe studentene mestrer godt, men dette kan henge sammen med svakheter i tolkningen av vilkår. Det er imidlertid interessant at studentene selv har noe større tro på egen kompetanse på dette område (Karlsen & Pedersen, 2025) enn det resultatene faktisk viser.

Et noe overaskende funn er at mange studenter har utfordringer med, eller utelater, vurdering knyttet til barnets beste og det vi har kalt overordnede rettsprinsipper i kategori 7. Det er svært mange studenter som ikke inkluderer slike vurderinger i besvarelsene sine. Mange nøyer seg med å nevne at barnets beste skal vurderes, uten å konkretisere hva dette innebærer i praksis for de ulike casene og juridiske problemstillingene. Å trekke inn prinsipper som «barnets beste» og «minste inngreps prinsipp» viser seg å være krevende for mange kandidater, til tross for at dette er sentralt i barnevernsretten og arbeidet i barnevernet.

Totalvurdering er en egen rubrikk i skjemaet (kategori 8 her). Her vurderer sensorene holistisk studentenes juridiske metodeforståelse, samt i hvilken grad fremstillingen er formulert i et presist språk med god struktur. Presis språkføring er et nødvendig element i juridisk arbeid. Selv om snittscoren øker i 2024, er rubrikkscoren fortsatt noe lav. Dette kan tyde på at studentene fortsatt har noen utfordringer knyttet til helhetlig metodeforståelse og språkføring. Her ser vi samsvar med studentenes egenvurdering (Karlsen & Pedersen, 2025). «Å formulere seg presist» er det læringsutbyttet flest av studentene selv vurderer at de har begrenset kunnskap om. Det er likevel kun 11 prosent som sier de i lav grad har kunnskap om å formulere seg presist, mens halvparten (54 prosent) sier de har høy kunnskap om dette.

## 10.2 Avslutning: Rubrikker som grunnlag for kvalitetsutvikling

Analysen av studentenes prestasjoner etter rubrikk-kategorier viser at de fleste kandidater identifiserer rettslige problemstillinger og finner relevante rettskilder, men at det er større utfordringer knyttet til tolkning av vilkår og anvendelse av overordnede rettsprinsipper som barnets beste. Dette kan indikere at det er områder som er vanskelige for studentene, og det kan peke på behovet for ytterligere fokus på disse delene av juridisk metode i undervisningen.

Analyser av rubrikkene kan gi et godt grunnlag for ulike institusjoner til å sammenlikne seg mot det nasjonale snittet og identifisere områder hvor undervisninger treffer godt og mindre godt. Slik kan vurderingsrubrikkene og sammenstillinger gi et godt grunnlag for å følge utviklingen over tid og for å målrette kvalitetsutviklingen i både undervisning og vurdering.

---

<sup>33</sup> I henhold til Barnevernsloven §1-3, §1-4 og §1-5.

## 11 Avslutning og veien videre

Denne rapporten har dokumentert implementeringen av nasjonal eksamen i juridisk metode og rettsanvendelse for masterutdanningene i barnevern og barnevernsfaglig arbeid i perioden 2022–2024 samt analysert vurderingskvalitet og studentenes resultater.

Innføringen av eksamenen var et direkte svar på et vedvarende kompetansebehov i barnevernet, særlig juridisk kompetanse, med mål om å styrke barn og familiers rettssikkerhet. NOKUTs rolle har vært å administrere og utvikle eksamenen slik at utdanningsinstitusjonene har oppmerksomhet på juridisk kompetanse, at studentene sikres et tilstrekkelig juridisk kunnskapsnivå og at sektoren etablerer felles faglige standarder.

### 11.1 Arena for samarbeid og kvalitetsutvikling

Gjennom tett samarbeid med fagmiljøene og etablering av en eksamensgruppe, referansegruppe og et nasjonalt sensorkorps har den nasjonale eksamenen skapt en arena for samarbeid, kvalitetsutvikling og erfaringsdeling. Arbeidet har ført til økt vurderingsfaglig kunnskap i fagmiljøene og skapt et nasjonalt nettverk for erfaringsdeling og faglig utvikling, som er særlig viktig for jurister i barnevern som arbeider i små miljøer. Sett i lys av målsetningen om sammenlikning, kvalitetsutvikling og felles standarder er dette positive funn. Vi har sett her hvordan sensorene, i likhet med undervisere, trekker frem dette (Pedersen & Bråten 2025).

Nasjonale vurderingsdialoger er en viktig arena for faglig kalibrering, erfaringsdeling og utvikling av felles standarder. Dette styrker både kvaliteten på vurderingene og fagmiljøets kompetanse. Sensorene rapporterer at nasjonal eksamen gir faglig samarbeid og utvikling med faglige nettverk og inspirasjon til egen undervisning.

### 11.2 Eksamen med høy kvalitet og reliabilitet

Testteoretiske analyser (IRT) og reliabilitetsanalyser (Krippendorffs alpha) viser at eksamenen og vurderingsrubrikkene fungerer godt og har høy reliabilitet, høyt sensorsamsvar. Rubrikkene fungerer godt for å skille mellom kandidater med sterk og svak fagkunnskap og skiller tydelig mellom ulike kompetansenivåer og sikrer konsistent vurdering, spesielt rundt grensen for bestått/ikke-bestått. Høstgjennomføringene har vist særlig sterkt sensorsamsvar.

Analysene viser at den nasjonale eksamenen holder høy kvalitet og har høy reliabilitet. Vurderingsrubrikkene er nyttige vurderingsverktøy for å sikre en konsistent og rettferdig vurdering, og sensorene vurderer at det fungerer godt. Videre er erfaringer fra sensorer at eksamenene samsvarer med læringsmålene studentene skal oppnå. De fleste studentene oppgir selv at de raskt forsto hva problemene i eksamensteksten var, men det er også noen som argumenterer for at eksamen er krevende (Karlsen & Pedersen, 2025). I lys av IRT-analyser fremstår ikke eksamenene totalt sett som vanskelige. Rubrikkene anses som nyttige, samtidig ser vi at sensorene påpeker noen utfordringer med at vurderingsrubrikkene i noen tilfeller kan gi poeng til studenter med svakere faglig forståelse, slik at de i noen tilfeller «reddes inn» av rubrikkene.

## 11.3 Institusjonell variasjon

Utdanningsinstitusjon er en betydelig forklaringsfaktor for studentenes prestasjoner, selv når det kontrolleres for bakgrunnsvariabler. Dette indikerer at det er betydelige institusjonsspesifikke faktorer som påvirker studentenes resultater. HVL og NTNU skiller seg positivt ut.

Det er en sterk positiv og statistisk signifikant sammenheng mellom studentenes inneværende karaktersnitt på masterprogrammet og poengsummen på nasjonal eksamen. Dette avkrefter antakelser fra underviserne om at studenter som presterer godt i andre emner ikke nødvendigvis lykkes på denne eksamenen.

Det kan virke som norsk språkføring og presisjon kan være en forklaringsfaktor for studentens prestasjoner. Norsk karakteren fra videregående skole har en tydelig, positiv og stabil sammenheng med sannsynligheten for å bestå eksamen. I «Studentstemmen: om undervisning, emneinnhold og vurdering i rettsanvendelse for barnevernsfaglige masterprogram» (Karlsen & Pedersen, 2025) så vi at 65 prosent av studentene hadde fått erfaring med å arbeide med case-oppgaver, mens det er 36 prosent som i høy grad opplever at de har fått erfaring med organisert skrivetrening i juridisk metode. Når vi ser at norsk karakteren har en så tydelig sammenheng med sannsynligheten for å bestå eksamen, kan det tale for viktigheten av at studentene får skrivetrening.

Alder har en liten, negativ effekt på sannsynligheten for å bestå. Dette funnet kan bidra til å avkrefte frykten enkelte undervisere hadde om at studenter som kom rett fra bachelor presterte dårligere enn eldre studenter med mer praksiserfaring (Pedersen & Bråten, 2025). Videre analyser av den negative effekten til alder hadde vært interessant å gjennomføre. Samtidig vet vi at mange som starter på master etter at de har vært i arbeidslivet, som regel er i full stilling ved siden av studiet (Karlsen & Pedersen, 2025), og de kan ha andre forkunnskaper enn studenter som kommer rett fra bachelor.

Selv om det er ulike opptakskrav knyttet til masterprogrammene, ser vi at masterprogram samt bachelorbakgrunn i liten grad forklarer variasjonen i studentenes eksamensresultater. Den større gjennomgangen av implementeringen av eksamen viser at eksamen har gitt mer fokus på juridisk kompetanse både på bachelor- og masternivå (Pedersen & Bråten, 2025). Videre vil det være interessant å se om sammenhengen mellom bachelorbakgrunn og eksamensresultat vil forsterkes som følge av dette økte fokuset på juridisk kompetanse på bachelornivå.

## 11.4 Utfordringer og veien videre

Til tross for eksamenens høye faglige kvalitet og reliabilitet, har gjennomsnittlig beståttandel ligget på 58 prosent i perioden 2022–2024. Dette har gitt gjenstand for bekymring. Dette indikerer at det fremdeles er et betydelig behov for kompetanseheving i juridisk metode og rettsanvendelse i barnevernsfaglig arbeid.

Denne utfordringen settes i en ny kontekst av den nylige politiske beslutningen om å fjerne kravet om mastergrad i barnevernet for ansettelse i barnevernet som skulle tre i kraft 2031, med unntak av ledere og stedfortredere. Beslutningen har skapt usikkerhet rundt fremtidig kompetanseutvikling i sektoren.

Gitt at den nasjonale eksamenen ble etablert nettopp for å sikre en felles, høy juridisk kompetanse, understreker den fortsatt relativt høye andelen ikke-bestått-studenter at behovet for å sikre tilstrekkelig juridisk metodeforståelse vedvarer.



Funn i denne rapporten kan peke på at NOKUT og fagmiljøene kan arbeide videre med:

- 1. Kontinuerlig kvalitetsutvikling av vurderingsprosessene:** Det er nødvendig å videreutvikle vurderingsrubrikkene for bedre å fange opp helhet og dybde i studentenes juridiske metodeforståelse og rettsanvendelse, særlig i lys av sensorenes bekymring for at dagens rubrikker kan overvurdere svake besvarelser. Dette arbeidet vil fortsatt skje gjennom test-teoretiske analyser, sensorundersøkelser og faglige dialoger.
- 2. Styrking av det faglige nettverket:** Det etablerte fellesskapet og samarbeidet har vist sin nytte ved å bidra til både kvalitetsarbeid og egen faglig utvikling blant underviserne. Dette nettverket må vedlikeholdes og videreføres for å sikre felles faglige standarder og praksiser. En oppgave for NOKUT og nettverket kan være å synliggjøre og dokumentere effekt av en slik eksamen og juss-kompetanse for ledere og ansatte i barnevernet med mastergrad. Funn fra studentundersøkelsen (Karlsen & Pedersen 2025) indikerer allerede at ansatte med juridisk kompetanse endrer praksis og sprer læring til kollegiet.
- 3. Undervisning:** Funnene knyttet til sterk korrelasjon med norskkarakter fra videregående skole og studentenes prestasjon på eksamen kan indikere at skriveferdigheter og presis språkføring er viktig for juridisk metodeforståelse. Utdanningsinstitusjonene kan vurdere tiltak som skrivetrening for studenter med svakere norskbakgrunn for å bedre forberede dem til eksamen. Flere institusjoner arbeider allerede med dette (Pedersen & Bråten 2025), og erfaringsdeling og spredning av kunnskap kan gi innsikt til å justere undervisningsopplegg og læringsaktiviteter.

Denne rapporten, sammen med de øvrige rapportene i NOKUTs gjennomgang av de nasjonale eksamenene, danner et viktig kunnskapsgrunnlag for den videre oppfølgingen og evalueringen av ordningen. Dette er imidlertid kun en status etter to ordinære gjennomføringer med alle institusjoner. Det vil kreves flere analyser over lengre tidsperioder for fullt ut å forstå og se årsakssammenhengene mellom variablene og effekten av nasjonal eksamen som tiltak i barnevernsutdanningene.



## 12 Litteraturliste

Andrich, D., & Sheridan, B. (2019). *RUMM2030 Plus [Computer software]*. Rumm Laboratory Pty Ltd.

Barne- og likestillingsdepartementet (BLD) (2017). Mer kunnskap – bedre barnevern. Kompetansestrategi for det kommunale barnevernet 2018-2024. <https://www.regjeringen.no/contentassets/6e31905299774f5681d57311e284d519/bedre-barnevern.pdf>

Barne- og familiedepartementet (BFD). (2021, 9. april). *Styrker rettssikkerheten for barn og foreldre*. Regjeringen Solberg. <https://www.regjeringen.no/no/dokumentarkiv/regjeringen-solberg/aktuelt-regjeringen-solberg/bld/nyheter/2021/ny-barnevernslov/id2843407/>

Barnevernsloven. (2021). Lov om barnevern (barnevernsloven) (LOV-2021-06-18-97). Lovdata. Hentet fra <https://lovdata.no/dokument/NL/lov/2021-06-18-97>

Barne-, ungdoms- og familiedirektoratet (Bufdir) (2019). *Utredning av kompetansehevingstiltak i barnevernet*. Rapport 04/2019. [https://cms.bufdir.no/siteassets/rapporter/barnevern/Utredning\\_av\\_kompetansehevingstiltak\\_i\\_barnevernet.pdf](https://cms.bufdir.no/siteassets/rapporter/barnevern/Utredning_av_kompetansehevingstiltak_i_barnevernet.pdf)

Bjølseth, G., Havnes, A., & Lauvås, P. (2011). Lavt sensorsamsvar – kan det bedres? *Uniped*, 34(4), 4–16. <https://doi.org/10.18261/ISSN1893-8981-2011-04-01>

Bloxham, S., Reimann, N. & Rust, C. (2018). Calibration of Standards: What, Why and How? Advanced HE, December 2018. [https://www.advance-he.ac.uk/sites/default/files/2022-01/Calibration%20synthesis%20report\\_0.pdf](https://www.advance-he.ac.uk/sites/default/files/2022-01/Calibration%20synthesis%20report_0.pdf)

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10.18637/jss.v048.i06>

Dahlen, W. N. & Langsrud, E. (2021). Ikke-juristers bruk av rettsregler En analyse av profesjonsutøveres beskrivelse av rettsanvendelsen i helse- og velferdstjenestene. *Kritisk juss*, 47 (3), 185-206. <https://doi.org/10.18261/issn.2387-4546-2021-03-04>

Forskrift om masterutdanning i barnevern (2025). *Forskrift om nasjonal retningslinje for masterutdanning i barnevern*. (FOR-2025-07-04-1456) Lovdata. <https://lovdata.no/dokument/SF/forskrift/2025-07-04-1456>

Forskrift om masterutdanning i barnevernsarbeid (2025). *Forskrift om nasjonal retningslinje for masterutdanning i barnevernsarbeid*. (FOR-2025-07-04-1455). Lovdata. <https://lovdata.no/dokument/SF/forskrift/2025-07-04-1455?q=barnevernsarbeid>

Gynnild, V. (2013). «Kriteriebasert vurdering» – hva innebærer det i praksis? *Uniped* 36 (1): 26-41. <https://doi.org/10.3402/uniped.v36i1.20959>

Gynnild, V. (2019). «Blind» klagesensur ved Universitetet i Oslo: Pålitelig og rettferdig eller et steg i feil retning? *Uniped*, 42(4), 340–354. <https://doi.org/10.18261/issn.1893-8981-2019-04-02>

Haakens, M. & Bråten, H. (2023). *Grunnskolelæreres resultater på nasjonal deleksamen i matematikk: betydningen av forkunnskaper*. MNT-konferansen 2023, 16. – 17. mars 2023, Stavanger. [grunnskolelæreres-resultater-pa-nasjonal-deleksamen-i-matematikk\\_betydningen-av-forkunnskaper\\_2023.pdf](https://www.mnt.no/medias/136121/grunnskolelæreres-resultater-pa-nasjonal-deleksamen-i-matematikk-betydningen-av-forkunnskaper-2023.pdf),

<https://www.uis.no/sites/default/files/2023-03/Trykt%20utgave%20konferansebidrag%20MNT-konferansen.pdf>

Haakens, M., Karlsen, H. & Bråten, H. (2021) *Resultater på nasjonal deleksamen i anatomi, fysiologi og biokjemi: Gode studenter eller gode studieprogrammer?* NOKUT rapport 4-2021. NOKUT. [resultater-pa-nd-i-anatomi-fysiologi-og-biokjemi\\_gode-studenter-eller-gode-studieprogrammer\\_4-2021.pdf](#)

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications, Inc.

Higher Education Quality Council (1997). *Assessment in higher education and the role of 'graduateness'* London: HEQC.

HEA (2013). Quality Assurance Agency, Higher Education Academy (2013) *External examiners understanding and use of standards*.  
<https://www.heacademy.ac.uk/system/files/downloads/external-examiners-report.pdf>

Hjemås, G., Beyer, S. & Dyrhaug, T. (2023). *Kompetanse i barnevernet*. Kartlegging 2020. Rapport 2023/8. SSB. [https://www.ssb.no/sosiale-forhold-og-kriminalitet/barne-og-familievern/artikler/kompetanse-i-barnevernet.kartlegging-2020/\\_attachment/inline/2f1fcaec-cd1d-4227-9e6e-0075eb35fa07:e15df9519042d641bc66eff715a751b2fb7e2682/RAPP2023-08.pdf](https://www.ssb.no/sosiale-forhold-og-kriminalitet/barne-og-familievern/artikler/kompetanse-i-barnevernet.kartlegging-2020/_attachment/inline/2f1fcaec-cd1d-4227-9e6e-0075eb35fa07:e15df9519042d641bc66eff715a751b2fb7e2682/RAPP2023-08.pdf)

Hughes, J. (2022). *\_krippendorffsalpha: Measuring Agreement Using Krippendorff's Alpha Coefficient\_*. R package version 2.0. <https://CRAN.R-project.org/package=krippendorffsalpha>

Jørgensen, C. & Bråten, H. (2019). *Can 'passed with distinction' as a new grading scale favour the transition towards formative assessment?* *Nordic Journal of STEM Education*. 3: 7-11. [doi:10.5324/njsteme.v3i1.2992](https://doi.org/10.5324/njsteme.v3i1.2992) ]  
<https://www.ntnu.no/ojs/index.php/njse/article/view/2992/2918>

Karlsen, M. L. & Pedersen, M. (2025). *Studentstemmen: om undervisning, emneinnhold og vurdering i rettsanvendelse for barnevernsfaglige masterprogram*. NOKUT rapport 7-2025. NOKUT.

Krippendorff, K. (2013). Commentary: A Dissenting View on So-Called Paradoxes of Reliability Coefficients. *Annals of the International Communication Association*, 36:1, 481-499. DOI: 10.1080/23808985.2013.11679143

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA.

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology* (4th Ed.). SAGE Publications. <https://doi.org/10.4135/9781071878781>

Kunnskapsdepartementet (2021, 12. mai). Oppstart og utvikling av nasjonal deleksamen for barnevernsfaglige masterutdanninger. [Brev]. Kunnskapsdepartementet.

Marzi, G., Balzano, M. & Marchiori, D. (2024). Krippendorff's Alpha Calculator: A user-friendly tool for computing Krippendorff's Alpha inter-rater reliability coefficient. *MethodsX* 12. <https://doi.org/10.1016/j.mex.2023.102545>

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47, 149–174. <http://dx.doi.org/10.1007/BF02296272>

NOKUT (u.å.). Nasjonal eksamen for master i barnevern og master i barnevernslaglig arbeid. <https://www.nokut.no/utdanningskvalitet/nasjonal-deleksamen/nasjonal-deleksamen-for-barnevern/>

NOU 2017: 12. (2017). «Svikt og svik — Gjennomgang av saker hvor barn har vært utsatt for vold, seksuelle overgrep og omsorgssvikt». Barne- og likestillingsdepartementet. <https://www.regjeringen.no/no/dokumenter/nou-2017-12/id2558211/>

NOU 2023: 7. (2023). Trygghet, sikker fremtid. Gjennomgang av rettsikkerheten for barn og foreldre i barnevernet. (Rettsikkerhetsutvalget). Barne- og familiedepartementet. <https://www.regjeringen.no/contentassets/a1eb3f73d3d94fa6b146dbd08f8a4253/no/pdfs/nou202320230007000dddpdfs.pdf>

Oslo Economics (2019). Samfunnsøkonomisk analyse av kompetansehevingstiltak i barnevernet. Rapportnummer 2019-1. <https://osloeconomics.no/wp-content/uploads/2019/10/Samfunnsokonomisk-analyse-av-kompetansehevingstiltak-i-barnevernet-juni-2019-1.pdf>

Pedersen, M. & Bråten, H. (2025). Undervisning i juridisk metode i barnevernslaglige mastere. Innsikt og erfaringer fra de første nasjonale eksamenene. NOKUT rapport 8. NOKUT.

Price, M. (2005). Assessment standards: the role of communities of practice and the scholarship of assessment. *Assessment & Evaluation in Higher Education*, 30:3, 215-230 <https://doi.org/10.1080/02602930500063793>

Prop. 133 L (2020 –2021). Lov om barnevern (barnevernsloven) og lov om endringer i barnevernloven. Barne- og familiedepartementet. <https://www.regjeringen.no/contentassets/f325e4de00fb472f85a7a2b94124f531/no/pdfs/prp202020210133000dddpdfs.pdf>

Rasch, G. (1960). *Probabilistic model for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.

Reise, S. P., & Revicki, D. A. (2015). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. New York: Routledge.

Shabankhani, B., Yazdani Charati, J., Shabankhani, K., Kaviani Cherati, S. (2020). Survey of agreement between raters for nominal data using krippendorff's Alpha. *Archives of Pharmacy Practice*, 11 (S1):160-4.

Universitets- og høyskolerådet (UHR). (2015a). Retningslinjer for karakterskalaen bestått/ikke bestått. UHR. [https://www.uhr.no/f/p1/iec171c06-38f2-4436-bce4-6621a6d32657/rapport om bestatt ikkje bestatt 10 112015.pdf](https://www.uhr.no/f/p1/iec171c06-38f2-4436-bce4-6621a6d32657/rapport%20om%20bestatt%20ikkje%20bestatt%2010%20112015.pdf)

Universitets- og høyskolerådet (UHR). (2015b). Karakterbruk i UH-sektoren 2015 (Rapport fra en arbeidsgruppe oppnevnt av Universitets- og høyskolerådet). UHR. [https://www.uhr.no/f/p1/i709c7820-e1cf-494d-8d3d-847ca5694fe2/karakterrappport 2015 endelig version-1.pdf](https://www.uhr.no/f/p1/i709c7820-e1cf-494d-8d3d-847ca5694fe2/karakterrappport%202015%20endelig%20version-1.pdf)

Universitets- og høyskolerådet (UHR). (2016). Karakterbruk i UH-sektoren 2016 (Rapport fra en arbeidsgruppe oppnevnt av Universitets- og høyskolerådet). UHR. [https://www.uhr.no/f/p1/i8b32118b-4211-4f1e-94e4-600420720be3/karakterrappport 2016 sluttversion.pdf](https://www.uhr.no/f/p1/i8b32118b-4211-4f1e-94e4-600420720be3/karakterrappport%202016%20sluttversion.pdf)

Universitets- og høyskoleloven. (2024). *Lov om universiteter og høyskoler* (LOV-2024-03-08-9). Lovdata. <https://lovdata.no/dokument/NL/lov/2024-03-08-9>

Wenger, E. (1998) *Communities of practice: Learning, meaning, and identity*, Cambridge: Cambridge University Press.

Wickham, H. (2023). *\_dplyr: A Grammar of Data Manipulation\_*. R package version 1.1.4. <https://dplyr.tidyverse.org>

Wickham, H. (2024). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. Version: 3.5.1. <https://ggplot2.tidyverse.org>

Xact by Ramboll (u.å.). <https://rambollxact.no/surveyxact>

Zapf, A., Castell, S., Morawietz, L. & Karch, A. (2016). Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate? *BMC Medical Research Methodology* 16:93. DOI 10.1186/s12874-016-0200-9

## Tabeller

|                                                                                                                                                                                                                               |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabell 1. Oversikt over eksamenstema og eksamensdokumenter. ....                                                                                                                                                              | 18 |
| Tabell 2. Fordeling av besvarelser per gjennomføring. ....                                                                                                                                                                    | 21 |
| Tabell 3. Eksempel på rubrikker med sterk svaravhengighet på eksamen høsten 2024. ....                                                                                                                                        | 27 |
| Tabell 4. Eksempel på rubrikker fra høst 2024 med vanskegrad høyere enn dyktigheten til kandidater som ikke bestod eksamen (37 poeng, dyktighet = 0,46). For rubrikk 19 (polytom) har vi gjengitt kriteriet for poeng 1. .... | 28 |
| Tabell 5. Oversikt over datamaterialet. ....                                                                                                                                                                                  | 31 |
| Tabell 6. Datasett for vurderingsrubrikk 1m* fra høst 2023 for å eksemplifisere strukturen i datasettet. ....                                                                                                                 | 31 |
| Tabell 7. Tematisk inndeling av antall rubrikker per gjennomføring. ....                                                                                                                                                      | 32 |
| Tabell 8. Utdrag av resultater på de seks første rubrikkene i gjennomføring høst 2024. ....                                                                                                                                   | 34 |
| Tabell 9. Oversikt over datamaterialet i klagesensur. ....                                                                                                                                                                    | 39 |
| Tabell 10. Samsensur i ordinær- og klagesensur på tvers av tematikker. ....                                                                                                                                                   | 40 |
| Tabell 11. Utsendelsestidspunkt av sensorundersøkelse på tvers av gjennomføringer. ....                                                                                                                                       | 43 |
| Tabell 12. Utvalg og svaroversikt per sensorundersøkelse. ....                                                                                                                                                                | 43 |
| Tabell 13. Oversikt over institusjonenes oppstart og eksamenstidspunkt. ....                                                                                                                                                  | 56 |
| Tabell 14. Oversikt over gjennomføringer og eksamensresultater høsten 2022 til og med høsten 2024. ....                                                                                                                       | 57 |
| Tabell 15. Oversikt over poeng på tvers av gjennomføringer. ....                                                                                                                                                              | 57 |
| Tabell 16. Valg av masterprogram fordelt på bachelorbakgrunn. Prosentandel av bachelorgruppe i parentes. ....                                                                                                                 | 60 |
| Tabell 17. Antall studenter per utdanningsinstitusjon og demografiske variabler i datagrunnlaget. ....                                                                                                                        | 63 |
| Tabell 18. Oversikt over datamaterialet. Antall i hver bachelorgruppe. ....                                                                                                                                                   | 63 |
| Tabell 19. Antall studenter i hvert masterprogram. ....                                                                                                                                                                       | 64 |
| Tabell 20. Enkel lineær regresjonsmodell av poengsum på VGS-poeng. ....                                                                                                                                                       | 65 |
| Tabell 21. Full lineær regresjonsmodell av poengsum på VGS-poeng med kontrollvariabler. ....                                                                                                                                  | 66 |
| Tabell 22. Resultater av oddsanalyse av bestått og ikke-bestått på VGS-poeng med kontrollvariabler. ....                                                                                                                      | 67 |

|                                                                                                                                                                |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabell 23. Resultat av predikerte terskelverdier for å bestå eksamen. N = 459. ....                                                                            | 68 |
| Tabell 24. Eksamensresultat høst 2024 og norskkarakter fra videregående skole. ....                                                                            | 68 |
| Tabell 25. Resultater fra enkel logistisk regresjonsmodell av bestått og ikke bestått på norskkarakter. ....                                                   | 69 |
| Tabell 26. Utvidet modell av sammenheng mellom norskkarakter og eksamensresultat. VID er tatt ut på grunn av kun 1 person med tilgjengelig norskkarakter. .... | 70 |
| Tabell 27. Resultater fra regresjonsanalyse av poengsum på bachelorgruppe med kontrollvariabler. ....                                                          | 71 |
| Tabell 28. Resultater av oddsanalyse av bestått og ikke-bestått på bachelorbakgrunn med kontrollvariabler og interaksjonsledd. ....                            | 72 |
| Tabell 29. Resultater fra regresjonsanalyse av poengsum på masterprogram med kontrollvariabler. ....                                                           | 75 |
| Tabell 30. Resultater av oddsanalyse av bestått og ikke-bestått på masterprogram med kontrollvariabler og interaksjonsledd. ....                               | 76 |
| Tabell 31. Resultat av enkel regresjonsanalyse av poengsum på inneværende karaktersnitt. ....                                                                  | 78 |
| Tabell 32. Resultater av regresjonsanalyse av poengsum på inneværende karaktersnitt med kontrollvariabler. ....                                                | 78 |

## Figurer

|                                                                                                                                                                                                                                                                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figur 1. Illustrasjon av at ett latent trekk eller én faktor F (endimensjonal modell) styrer de observerte verdiene på variablene a–d. ....                                                                                                                                                                                                                                                                     | 24 |
| Figur 2. Sannsynligheten for riktig svar (y-aksen) på rubrikk 30 høst 2024 som funksjon av dyktighet (x-aksen). Målepunktene viser faktisk andel kandidater som fikk poeng i hvert av fire «dyktighetsintervallene» fra lav til høy dyktighet. De observerte verdiene (punktene) passer godt til de teoretisk forventede verdiene (kurven). ....                                                                | 25 |
| Figur 3. Sannsynligheten for riktig svar (y-aksen) på rubrikk 27 høst 2024 som funksjon av dyktighet (x-aksen). Målepunktene viser faktisk andel kandidater som fikk poeng i hvert av fire «dyktighetsintervallene» fra lav til høy dyktighet. De observerte verdiene (punktene) passer ikke godt til de teoretisk forventede verdiene (kurven). ....                                                           | 26 |
| Figur 4. Kurvene beskriver sannsynligheten for 0, 1 og 2 poeng (y-aksen) på rubrikk 22 høst 2024 som funksjon av dyktighet (x-aksen). Poengene fungerte godt, fordi noen dyktigheter har 0, 1 eller 2 poeng som sin mest sannsynlige poengverdi. De vertikale stiplede strekene angir vanskegraden til poeng 1 (0,4) og poeng 2 (3,9). På rubrikk 22 fikk 52 % av kandidatene 1 poeng og 5 % fikk 2 poeng. .... | 28 |
| Figur 5. Krippendorffs alpha med tilhørende konfidensintervaller av total poengsum mellom sensor 1 og sensor 2 i ordinær sensur på tvers av gjennomføringer. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. ....                                                                                                                                                       | 35 |
| Figur 6. Boksplott av Krippendorffs alpha per vurderingsrubrikk på tvers av gjennomføringer. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. Blå prikk illustrerer median, mens rød prikk illustrerer gjennomsnitt. ....                                                                                                                                                | 36 |
| Figur 7. Gjennomsnittlig alpha per tema på tvers av gjennomføringer. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. ....                                                                                                                                                                                                                                               | 37 |
| Figur 8. Boksplott av Krippendorffs alpha av temaene 1–3. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. ....                                                                                                                                                                                                                                                          | 38 |
| Figur 9. Boksplott av Krippendorffs alpha for temaene 4 og 5. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. ....                                                                                                                                                                                                                                                      | 38 |

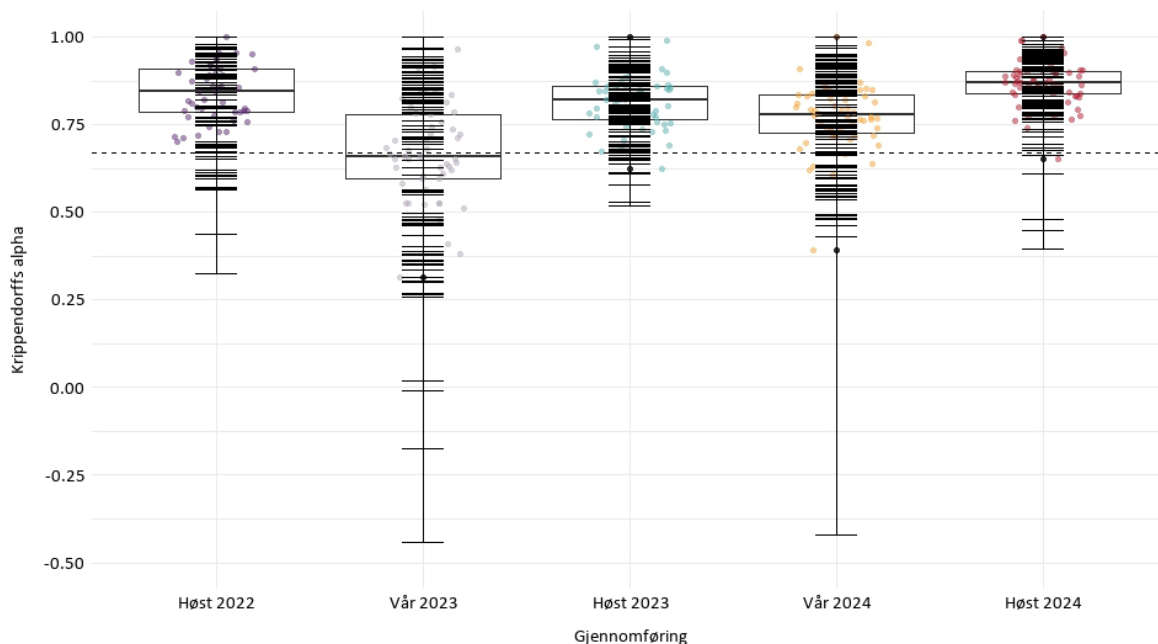
|                                                                                                                                                                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figur 10. Boksplott av Krippendorffs alpha for temaene 6–8. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur. ....                                                                                                | 39 |
| Figur 11. Krippendorffs alpha med tilhørende konfidensintervaller av total poengsum mellom samsensur i ordinær- og klagesensur på tvers av gjennomføringer. Stiplet linje illustrerer $\alpha = 0,67$ , som regnes som grensen for reliabel sensur..... | 40 |
| Figur 12. Prosentfordeling av sensorenes faglige rolle i utvalget. Antall i parentes.....                                                                                                                                                               | 44 |
| Figur 13. Svarfordeling på spørsmålet «Var eksamen i samsvar med læringsmålene?». Gjennomsnitt i parentes. ....                                                                                                                                         | 45 |
| Figur 14. Svarfordeling på spørsmålet «I hvilken grad testet eksamen juridisk metode?». Gjennomsnitt i parentes. ....                                                                                                                                   | 46 |
| Figur 15. Svarfordeling på spørsmålet «I hvilken grad ga sensorveiledning/rubrikkene gode rammer for den skjønnsmessige vurderingen?». Gjennomsnitt i parentes. ....                                                                                    | 47 |
| Figur 16. Svarfordeling på spørsmålet «Var det nyttig å bruke vurderingsskjema og matrise som vurderingsverktøy?». Gjennomsnitt i parentes.....                                                                                                         | 48 |
| Figur 17. Svarfordeling på spørsmålet «Er du enig i poengsettingen i matrisen?». Gjennomsnitt i parentes. ....                                                                                                                                          | 49 |
| Figur 18. Svarfordeling på spørsmålet «Svarte poenggrensen for bestått/ikke bestått til sensorens vurdering av kandidaten?». ....                                                                                                                       | 50 |
| Figur 19. Svarfordeling på spørsmålet «I hvilken grad ga vurderingsskjemaet et riktig bilde av studentenes prestasjon på eksamen?». Gjennomsnitt i parentes.....                                                                                        | 51 |
| Figur 20. Svarfordeling på spørsmålet «I hvilken grad bidro matrisen til at studentene ble vurdert etter samme kriterier?». Gjennomsnitt i parentes.....                                                                                                | 52 |
| Figur 21. Svarfordeling på spørsmålet «I hvilken grad anser du matrisen og sensorveiledningen å kunne gi studentene god tilbakemelding på sin eksamen/begrunnelse for karakter?». Gjennomsnitt i parentes.....                                          | 53 |
| Figur 22. Resultater på nasjonal eksamen i rettsanvendelse høst 2023 på tvers av institusjoner. ....                                                                                                                                                    | 58 |
| Figur 23. Resultater på nasjonal eksamen i rettsanvendelse høst 2024 på tvers av institusjoner. ....                                                                                                                                                    | 58 |
| Figur 24. Beståttandel høst 2023 og høst 2024 på tvers av bachelorbakgrunn. ....                                                                                                                                                                        | 73 |
| Figur 25. Beståttandel fordelt på mastergrupper høst 2023 og høst 2024. ....                                                                                                                                                                            | 77 |
| Figur 26. Visualisering av korrelasjonsanalyse mellom poengsum og inneværende karaktersnitt høst 2023. ....                                                                                                                                             | 79 |
| Figur 27. Visualisering av korrelasjonsanalyse mellom poengsum og inneværende karaktersnitt høst 2024. ....                                                                                                                                             | 80 |
| Figur 28. Eksempel på vurderingsrubrikker i sensurskjemaet høst 2024 (1–11 av totalt 58 rubrikker).....                                                                                                                                                 | 82 |
| Figur 29. Eksempel på sammenstilling av vurderingsrubrikker nasjonalt.....                                                                                                                                                                              | 83 |
| Figur 30. Eksempel på sammenstilling av vurderingsrubrikker per tema nasjonalt. ....                                                                                                                                                                    | 84 |
| Figur 31. Krippendorffs alpha for hver rubrikk med tilhørende konfidensintervaller. ....                                                                                                                                                                | 95 |



## 13 Vedlegg

### 13.1 Krippendorffs alpha for hver rubrikk med tilhørende konfidensintervaller

Figur 31. Krippendorffs alpha for hver rubrikk med tilhørende konfidensintervaller.



### 13.2 Sensorundersøkelse – Endringer av spørsmålsformuleringer på tvers av undersøkelser

I det følgende vises oversikt over spørsmålsformuleringer knyttet til figurene i kapittel 6. Sensorenes vurdering av eksamenene, vurderingsrubrikker og vurderingsprosessen.

Figur 2. Eksamen er i samsvar med læringsmålene

- H22: Overordnet om eksamen - Var eksamen i samsvar med læringsmålene?
- H23, V24 og H24: Overordnet om eksamen - Var eksamen i samsvar med læringsutbyttene denne eksamen skal vurdere?

Figur 3. Testet eksamen juridisk metode

- H22: Overordnet om eksamen - I hvilken grad testet eksamen juridisk metode?
- H23, V24 og H24: Overordnet om eksamen - Hvor godt testet eksamen juridisk metode?

Figur 4. Gode rammer for skjønnsmessig vurdering

- H22 til V24: I hvilken grad bidro sensorveiledningen til - å gi gode rammer for skjønnsmessige vurdering?

- H24: I hvilken grad: - Ga rubrikkene gode rammer for den skjønnsmessige vurderingen

*Figur 5. Nyttien av å bruke vurderingsrubrikker*

- H22: Var det nyttig å bruke vurderingsskjema og matrise som vurderingsverktøy?
- H23 og V24: Var det nyttig å bruke vurderingsskjema/matrise som vurderingsverktøy?
- H24: I hvilken grad var det nyttig å bruke vurderingsrubrikker som vurderingsverktøy

*Figur 6. Enig i poengsettingen i vurderingsrubrikkene*

- H22 – V24: Om matrisen/vurderingsskjemaet - Er du enig i poengsettingen i matrisen?
- H24: I hvilken grad: - Er du enig i poengsettingen i rubrikkene?

*Figur 7. Svarte poenggrensen for bestått/ikke bestått til sensorens vurdering av kandidaten*

- Alle: I hvilken grad svarte poenggrensen for bestått/ikke bestått til din vurdering av kandidaten?

*Figur 8. Gir vurderingsskjemaet et riktig bilde av studentenes prestasjon på eksamen*

- H22 – V24: Om matrisen/vurderingsskjemaet - I hvilken grad ga vurderingsskjemaet et riktig bilde av studentenes prestasjon på eksamen?
- H24: I hvilken grad: - Ga vurderingsrubrikkene et riktig bilde av studentenes prestasjon på eksamen

*Figur 9. Bidro matrisen til at studentene ble vurdert etter samme kriterier*

- H23 – V24: Om matrisen/vurderingsskjemaet - I hvilken grad bidro matrisen til at studentene ble vurdert etter samme kriterier?
- H24: I hvilken grad: - Bidro vurderingsrubrikkene til at studentene ble vurdert etter samme kriterier

*Figur 10. Gir vurderingsrubrikkene og sensorveiledningen studentene god tilbakemelding på sin eksamen/begrunnelse for karakter*

- H23-V24: I hvilken grad anser du matrisen og sensorveiledningen å kunne gi studentene god tilbakemelding på sin eksamen/begrunnelse for karakter?
- H24: I hvilken grad anser du at rubrikkene gir studentene god tilbakemelding på sin eksamen/begrunnelse for karakter?

## 13.3 Sensorundersøkelse høst 2024

### Sensorundersøkelse barnevern (eksamen høst 2024)

Eksamensgruppen og NOKUT ønsker å gjennomføre en sensorundersøkelse rundt avviklingen av nasjonal deleksamen i rettsanvendelse for barnevernfaglige mastere 4. desember 2024.



Vi ville derfor sette stor pris på om du som har sensurert svarer på noen spørsmål som kan hjelpe oss med videre utvikling av eksamen. Alle svar behandles anonymt. Svarene vil benyttes for å videreutvikle eksamen og i analyser av sensuren.

Tusen takk for flott innsats med sensuren!

Med vennlig hilsen

NOKUT

Dine roller er (du kan krysse av flere)

- (1) ☐ Emneansvarlig
- (2) ☐ Annen underviser
- (3) ☐ Medlem i eksamensgruppen
- (4) ☐ Medlem i referansegruppen
- (5) ☐ Annet (spesifiser) \_\_\_\_\_

| Overordnet om eksamen                                                  | (1) I liten grad | (2)   | (3)   | (4)   | (5) I stor grad |
|------------------------------------------------------------------------|------------------|-------|-------|-------|-----------------|
| Var eksamen i samsvar med læringsutbyttene denne eksamen skal vurdere? | (1) ○            | (2) ○ | (3) ○ | (4) ○ | (5) ○           |
| Hvor godt testet eksamen juridisk metode?                              | (1) ○            | (2) ○ | (3) ○ | (4) ○ | (5) ○           |
| Var arbeidsmengden tilpasset tiden studentene hadde til rådighet?      | (1) ○            | (2) ○ | (3) ○ | (4) ○ | (5) ○           |

Eventuelle kommentarer om eksamen

---



---



---



---



---



---

| Om caset                                                                      | (1) I liten grad | (2)   | (3)   | (4)   | (5) I stor grad |
|-------------------------------------------------------------------------------|------------------|-------|-------|-------|-----------------|
| Var caset formulert på en klar og forståelig måte?                            | (1) ○            | (2) ○ | (3) ○ | (4) ○ | (5) ○           |
| Gikk det tydelig frem hvilke hovedproblemstillinger som skulle identifiseres? | (1) ○            | (2) ○ | (3) ○ | (4) ○ | (5) ○           |

Hvis du har vært sensor tidligere: Hva syns du om dette caset sett i lys av tidligere eksamener?

---

---

---

---

---

---

Har du innspill til hvordan eksamensoppgaven kan utvikles?

---

---

---

---

---

---

Eventuelle kommentarer til caset

---

---

---

---

---

---

Opplevde du at kandidater fant andre problemstillinger enn de som er definert i vurderingsrubrikkene/sensorveiledningen? Og i hvilken grad var dette en utfordring i sensureringen?

---

---

---

---

---

---

Om vurderingsrubrikkene:

| I hvilken grad:                                                                                  | (1) I liten grad          | (2)                       | (3)                       | (4)                       | (5) I stor grad           |
|--------------------------------------------------------------------------------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Var det nyttig å bruke vurderingsrubrikker som vurderingsverktøy                                 | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Bidro vurderingsrubrikkene til at studentene ble vurdert etter samme kriterier                   | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Ga vurderingsrubrikkene god informasjon for sensureringen                                        | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Er du enig i poengsettingen i rubrikkene                                                         | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Ga rubrikkene gode rammer for den skjønnsmessige vurderingen                                     | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Ga vurderingsrubrikkene et riktig bilde av studentenes prestasjon på eksamen                     | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Svarte poenggrensen for bestått/ikke bestått til din vurdering av kandidaten                     | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |
| Anser du at rubrikkene gir studentene god tilbakemelding på sin eksamen/begrunnelse for karakter | (1) <input type="radio"/> | (2) <input type="radio"/> | (3) <input type="radio"/> | (4) <input type="radio"/> | (5) <input type="radio"/> |

Andre kommentarer til vurderingsrubrikkene (utdyp gjerne svarene ovenfor):

---



---



---



---



---



---

Hvordan opplevde du følgende:

|                                  |                                                 |
|----------------------------------|-------------------------------------------------|
| Innledende sensormøte            | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Kalibreringsmøtet                | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Samsensur-ordningen              | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Kommunikasjon med NOKUT          | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Arbeidsmengde ved ordinær sensur | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Proessen rundt klagesensur       | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |

Har du andre kommentarer til eksamen, sensurprosessen eller tips til neste gjennomføring?

---

---

---

---

---

---

Til slutt har vi noen spørsmål om oppfølging og kvalitetsutvikling med nasjonal deleksamen.

|                                                                                                                                              |                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| På hvilken måte har det vært nyttig for deg å være sensor på nasjonal eksamen?                                                               | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Gjøres det analyser av eksamensresultatene ved din institusjon, og hvilke?                                                                   | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Hvordan benytter dere erfaringene fra eksamen og/eller enkeltoppgaver for å justere undervisningen?                                          | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Diskuterer dere med andre campus og institusjoner om mulige årsaker til forskjeller i resultater?                                            | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |
| Hvordan får erfaringer fra nasjonal deleksamen betydning for hvordan du selv utvikler eksamener (eks. format, oppgaver og sensurveiledning)? | <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> |

Med bakgrunn i data og analyser fra nasjonal deleksamen ønsker NOKUT å legge til rette for erfaringsdeling, seminarer og analyser. Er det tema eller problemstilling du ønsker vi ser nærmere på? Kom gjerne med innspill til kunnskapsoppsummering eller annet nyttig i tilknytning til emnet og nasjonal deleksamen.

---

---

---

---

---

---

---

Tusen takk for at du tok deg tid til å svare på undersøkelsen. Det setter vi pris på.





DRAMMENSVEIEN 288 | POSTBOKS 578, 1327 LYSAKER | T: 21 02 18 00 | [NOKUT.NO](https://www.nokut.no)