

NOKUTs oppsummeringer

Nasjonal deleksamen – et pilotprosjekt og en mulighetsstudie

Oppsummering av erfaringene med å gjennomføre nasjonal deleksamen i tre profesjonsutdanninger

Mars 2017



NOKUT 

NOKUTs arbeid skal bidra til at samfunnet har tillit til kvaliteten i norsk høyere utdanning og fagskoleutdanning, samt godkjent høyere utenlandsk utdanning. Med rapportserien «NOKUTs oppsummeringer» vil vi bidra til økt kunnskap om temaer knyttet til fagskole, høyere utdanning og godkjenning av utenlandsk utdanning i Norge. Data til rapportene får NOKUT gjennom arbeidet med akkreditering og godkjenning.

Vi håper at resultatene våre kan være nyttige i arbeidet med godkjenning av utenlandsk utdanning og for lærestedene i arbeidet med å kvalitetssikre og videreutvikle utdanningstilbudene.

| | |
|--------------------|--|
| Tittel: | Nasjonal deleksamen – et pilotprosjekt og en mulighetsstudie |
| Forfattere: | Kjersti Tokstad, Stephan Hamberg |
| Dato: | 15. mars 2017 |

Sammendrag

NOKUT har på oppdrag fra Kunnskapsdepartementet gjennomført en mulighetsstudie og et pilotprosjekt om nasjonale deleksamener i tre profesjonsutdanninger: bachelorgradsstudiet i regnskap og revisjon, bachelorgradsstudiet i sykepleie og i begge grunnskolelærerutdanningene. Syv eksamener er avviklet per 1. mars 2017, og det gjenstår ytterligere tre eksamener før hele prosjektet er ferdig. Analysene og resultatene som er presentert så langt, viser at nasjonale deleksamener kan gi oss viktig informasjon om studentenes kunnskapsnivå, at det gir fagmiljøer mulighet til å sammenligne seg selv med andre, og at det kan bidra til økt tillit til utdanningen. Prosjektet har også gitt oss mulighet til å kalibrere karaktersettingen på tvers av institusjoner. Samtidig har diskusjonene om nasjonale deleksamener reist mange viktige problemstillinger, og det er fortsatt flere utfordringer som bør løses før nasjonale deleksamener kan videreføres som en varig ordning i høyere utdanning.

Forutsetningene for studentene og for lærestedene bør være så like som mulig, og det krever at eksamenene kommer på samme tid i studieløpet, at innhold og omfang er likt, og at læringsutbyttebeskrivelsene stort sett er de samme uavhengig av studiested. Det er videre en risiko for at studenter og faglærere bruker mer tid på det emnet studentene testes i til eksamen, enn på andre emner («teaching to the test»). Dette kan unngås ved at en nasjonal eksamen omfatter et større emne, for eksempel som i sykepleierutdanningen hvor den nasjonale deleksamenen dekker hele emnet anatomi, fysiologi og biokjemi.

NOKUTs ad hoc-organisering er sårbar og ressurskrevende, og før ordningen med nasjonale deleksamener videreføres, er det viktig å finne gode løsninger for effektivisering og digitalisering.

Innhold

| | | |
|----------|---|-----------|
| 1 | Innledning | 1 |
| 1.1 | Bakgrunn | 1 |
| 1.2 | Oppdraget | 2 |
| 1.3 | Utvalget | 2 |
| 1.4 | Organisering og planlegging | 3 |
| 1.4.1 | Ansvars- og arbeidsdeling | 4 |
| 1.4.2 | Utfordringer..... | 4 |
| 1.4.3 | Logistikk og administrasjon | 6 |
| 2 | Analyser og resultater | 8 |
| 2.1 | Formål 1: Gi nyttig informasjon om studentenes kunnskapsnivå | 8 |
| 2.2 | Formål 2: Gi fagmiljøene mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner..... | 11 |
| 2.3 | Formål 3 (og 4): Bidra til økt tillit til utdanningene i samfunnet | 14 |
| 3 | Utfordringer som må løses før nasjonale deksamener kan innføres | 17 |
| 4 | Konklusjon | 21 |

1 Innledning

1.1 Bakgrunn

Høsten 2014 fikk NOKUT i oppdrag fra Kunnskapsdepartementet å gjennomføre en mulighetsstudie og et pilotprosjekt om nasjonale deksamener i noen utvalgte profesjonsfag. Formålet med nasjonale deksamener er at de skal gi nyttig informasjon om studentenes kunnskapsnivå, gi fagmiljøene mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner og kunne utvikle seg på grunnlag av det og å gi samfunnet økt tillit til utdanningene.

I regjeringens strategi for lærerutdanningen, *Lærerløftet. På lag for kunnskapsskolen* (2014), er innføring av nasjonale deksamener i grunnskolelærerutdanningene et eksplisitt mål, og departementets oppdrag til NOKUT må ses i sammenheng med det. Hensikten med mulighetsstudiet og pilotprosjektet har vært å undersøke om det er praktisk mulig å gjennomføre nasjonale deksamener i høyere utdanning, om det er mulig å måle studentenes kunnskap på tvers av norske utdanningsinstitusjoner og om nasjonale deksamener kan fungere som et verktøy for sammenligning og kvalitetsutvikling.

I tillegg har NOKUT ønsket å finne ut om nasjonale deksamener kan ha en karakterkalibrerende effekt på sikt. Oppdraget må også ses på bakgrunn av SØF-rapport nr. 03/13¹, som bekreftet det tidligere studier også har vist, at det er systematiske forskjeller i utdanningsinstitusjonenes karakterpraksis. Siden karakterpraksisen varierer mye mellom ulike institusjoner, gir ikke karakterer korrekt informasjon om studentenes kompetanse og ferdigheter. Konklusjonen er at karakterer i seg selv ikke kan anvendes som en indikator eller et mål på faglig kvalitet i utdanningene, noe vi kommer tilbake til i løpet av denne rapporten.

NOKUT har publisert en rapport for hver nasjonal deksamene som vi har gjennomført. Denne rapporten er en beskrivelse og gjennomgang av erfaringene med mulighetsstudien og pilotprosjektet nasjonale deksamener, som har omfattet tre profesjonsutdanninger i perioden 2015–2017, men den representerer ingen bred evaluering av prosjektet. Vi går nærmere inn på resultatene og hva som kan forklare forskjellene på dem, og vi redegjør også for hvilke utfordringer som bør være løst før nasjonale deksamener innføres i høyere utdanning permanent. Foruten Kunnskapsdepartementet er rapportens målgruppe alle som er omfattet av nasjonale deksamener, men også andre som er interessert i utdanningskvalitet.

¹ Strøm, Bjørn, Torberg Falch, Trude Gunnes og Marianne Haraldsvik (2013): «Karakterbruk og kvalitet i høyere utdanning. SØF-prosjekt nr. 2150: Analyse av karakterbruk og kvalitet i høyere utdanning». SØF-rapport nr. 03/13.

1.2 Oppdraget

NOKUT fikk oppdraget 8. september 2014. I oppdragsbrevet står det:

I grunnopplæringen har nasjonale prøver vært benyttet i flere år, og det har bidratt til nyttig kunnskap om tilstanden i norsk skole. Nasjonale deksamener i høyere utdanning vil gi oss nyttig informasjon om studentenes kunnskapsnivå og fagmiljøene får mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner, og kan utvikle seg på grunnlag av dette. Det kan også virke tillitvekkende for samfunnet at man måler kunnskapen gjennom nasjonale deksamener.

NOKUT får i oppdrag å gjennomføre en mulighetsstudie og et pilotprosjekt som fullskala gjennomføring av nasjonale deksamener innenfor enkelte profesjonsfag. Dette gjøres i nært samarbeid med de som har erfaringer med nasjonale prøver i grunnopplæringen. NOKUT skal ta sikte på både å utvikle deksamenerne og gjennomføre piloten i løpet av 2015.

Oppdraget fra Kunnskapsdepartementet innebar både en mulighetsstudie og et pilotprosjekt med fullskala gjennomføring av nasjonale deksamener, noe som er en innbyrdes motsetning. En mulighetsstudie skal undersøke om det er praktisk mulig å gjennomføre det som studien har som mål, altså å teste ut om dette er en studie eller et prosjekt som egner seg for å gjennomføre i større format og eventuelt hva som må til for at et prosjekt skal lykkes. En pilot er en småskalatest av en undersøkelse eller en studie man allerede har bestemt at man skal gjennomføre og som ofte brukes til å teste utformingen av det endelige prosjektet.²

NOKUT fikk altså tre typer oppdrag i ett og samme oppdrag sammen med en forventning om at hele oppdraget skulle gjennomføres og gi resultater allerede året etter. Det hadde vært nærliggende å starte med en mulighetsstudie, for deretter å utvikle et pilotprosjekt og på bakgrunn av det igjen gjennomført nasjonale deksamener i fullskalaformat innenfor et fag eller emne i en profesjonsutdanning. Vi slo fast at det ikke var mulig å løse oppdraget nøyaktig slik det stod formulert i oppdragsbrevet – mulighetsstudie, pilotprosjekt og fullskala gjennomføring på én gang. NOKUT og de involverte fagmiljøene måtte derfor bruke litt tid på – i dialog med departementet – å definere prosjektet og finne løsninger på hvordan vi skulle kombinere mulighetsstudien med et pilotprosjekt, og ikke minst hvordan vi skulle kombinere pilotprosjekt og fullskala. Vi kommer tilbake til det underveis i rapporten.

1.3 Utvalget

Oppdragsbrevet la ingen føringer for hvilke profesjonsutdanninger NOKUT skulle velge ut, men det lå likevel en klar forventning fra departementet om at begge grunnskolelærerutdanningene (GLU 1–7 og GLU 5–10) skulle delta i prosjektet.³ NOKUT ønsket primært å velge ut profesjonsutdanninger som var positivt innstilt til å teste ut en nasjonal deksamener. På dette tidspunktet fantes det ingen lov hjemmel eller forskrift om nasjonale deksamener, så NOKUT kunne ikke pålegge institusjoner eller fagmiljøer å delta i prosjektet før hjemmelen var på plass.⁴ For at vi skulle rekke å være godt i

² Hovdhaugen, Elisabeth, Vibeke Opheim, Jørgen Sjaastad og Rachel Sweetman (2013): «AHELO mulighetsstudie. Oppsummering av erfaringene med å gjennomføre OECDs AHELO mulighetsstudie i Norge». NIFU-rapport 11/2014, s. 7.

³ Departementet spesifiserte i senere brev til NOKUT (henholdsvis 3. og 26. november 2014) at begge grunnskolelærerutdanningene skulle omfattes av prosjektet.

⁴ Forskriftsendringen ble vedtatt 13. juli 2015, og departementet sendte ut brev 11. august om at gjennomføring av nasjonal deksamener høsten 2015 var obligatorisk innenfor grunnskolelærerutdanningene og sykepleierutdanningen.

gang med prosjektet innen utgangen av 2015, var vi både avhengig av frivillighet og at noen fagmiljøer hadde høstet erfaringer som vi kunne dra veksler på.

Vi hadde møter med fire nasjonale råd/profesjonsråd under UHR kort tid etter at vi hadde fått oppdraget: Nasjonalt råd for økonomisk-administrativ utdanning (NRØA), Nasjonalt råd for lærerutdanning (NRLU), Nasjonalt råd for teknologisk utdanning (NRT) og Nasjonalt profesjonsråd for utdanning og forskning innen sykepleie (NPR-sykepleie). Vi hadde også møter med Utdanningsdirektoratet, testeksperter fra CEMO og EKVA og med studentforeningen til Norsk sykepleierforbund, NSF Student.

Fagpersoner innenfor bachelorgradsstudiet i økonomi og administrasjon hadde på forhånd gitt uttrykk for at det kunne være en fordel med nasjonale eksamener for å sikre opptaket til mastergradsstudiet. Økonomi og administrasjon var for øvrig et av fagområdene SØF-rapporten undersøkte. Samtidig gjorde representanter fra NRØA NOKUT oppmerksom på at institusjoner som tilbyr bachelorgradsstudiet i regnskap og revisjon, i en årrekke hadde samarbeidet om frivillige, felles eksamensoppgaver i de fire avsluttende revisoreksamenene. Alle de aktuelle institusjonene hadde frem til da valgt å delta i samarbeidet om eksamenene i «årsregnskap og god regnskapsskikk» og «revisjon», slik at de i praksis ble oppfattet som nasjonale. Det at det allerede eksisterte en nasjonal ordning med et fast opplegg og en etablert infrastruktur, var en svært viktig faktor for at NOKUT ønsket å ta med bachelorgradsstudiet i regnskap og revisjon i prosjektet.

Norsk sykepleierforbund viste interesse for prosjektet, og NSF Student tok i 2011 initiativ til en frivillig kandidateksamen for tredjeårsstudenter i bachelorgradsstudiet. De hadde fått tilbakemeldinger fra medlemmene om manglende testing i viktige fag, store variasjoner i undervisnings- og vurderingsformer og om at studiekvaliteten generelt var for dårlig mange steder. Testen ble arrangert i årene 2012–2014, og testen i 2013 fikk en del oppmerksomhet. 52 prosent (av 530 deltagere) strøk, og 33 studenter valgte svaralternativ som ville ha tatt livet av pasienten. Der fagforeningen og studentene var entusiastiske og ønsket en nasjonal eksamen velkommen, var nok representantene fra NPR-sykepleie mer nøkterne, for det var tross alt mange praktiske forhold som ikke lå til rette for at det allerede året etter skulle kunne arrangeres én, felles eksamen for alle studenter i et fag eller et emne i sykepleierutdanningen. Men de var likevel positive, og det var et godt utgangspunkt for at NOKUT ønsket å inkludere bachelorgradsstudiet i sykepleie i prosjektet.

Representantene fra NRLU og NRT var mindre positive enn representantene fra de to andre rådene, så vi bestemte oss for ikke å gå videre med ingeniørutdanningene. Siden departementet allerede hadde bestemt at grunnskolelærerutdanningene skulle delta, satt vi igjen med tre profesjonsutdanninger som skulle gjennomføre nasjonale deleksamener i følgende emner: årsregnskap i bachelorgradsstudiet regnskap og revisjon, anatomi, fysiologi og biokjemi i bachelorgradsstudiet i sykepleie og undervisningskunnskap i matematikk i begge grunnskolelærerutdanningene.

1.4 Organisering og planlegging

Vi delte prosjektet inn i tre delprosjekter, ett for hver profesjonsutdanning. Arbeidet og erfaringene med å forberede den nasjonale deleksamenen i årsregnskap fungerte siden som en slags mal for forberedelsene til de andre eksamenene og for samarbeidet med fagmiljøene. Den største forskjellen når det gjaldt årsregnskap, var å gå fra lokal til nasjonal sensur, og at NOKUT skulle stå for den overordnede organiseringen. Ellers var svært mye av opplegget på plass før NOKUT kom inn i bildet.

1.4.1 Ansvars- og arbeidsdeling

NOKUT har hatt det overordnede ansvaret for at prosjektet har blitt utført på en faglig forsvarlig måte. Der mandatet har vært uklart og skapt usikkerhet hos partene, har det vært NOKUTs ansvar å drøfte det med departementet, så har det vært opp til departementet og/eller politisk ledelse å fatte en beslutning. NOKUT har videre hatt ansvar for å koordinere den praktiske og logistiske delen av eksamensgjennomføringen, fra å bestille og distribuere eksamensark i forkant til å fordele resultatene av sensur og behandling av klager på eksamenskarakterer i etterkant. Det er NOKUT som har analysert og presentert resultatene av de nasjonale deleksamenene, og dette er gjort i dialog med fagmiljøene som har utviklet eksamensoppgavene.

Fagpersonene som har utviklet eksamensoppgavene i grunnskolelærerutdanningene og sykepleierutdanningen, har blitt valgt ut og fått mandatene sine av NRLU og NPR-sykepleie. NOKUT har ikke tatt del i utviklingen av eksamensoppgavene, men har lagt til rette for eksamensgruppens arbeid, som formelt har jobbet på oppdrag for NOKUT. Eksamensgruppene har også fungert som et mellomledd mellom rådene og NOKUT når viktige spørsmål har meldt seg. Fagmiljøet i regnskap og revisjon hadde allerede lang erfaring i å utvikle og arrangere felles eksamener, og siden opplegget skulle være mest mulig likt det de allerede hadde, krevde det noe mindre organisering fra NOKUTs side. NRØA oppnevnte likevel en prosjektgruppe, og både lederen for den og NRØA har vært viktige støttespillere og diskusjonspartnere for NOKUT, særlig i begynnelsen, siden bachelorgradsstudiet i regnskap og revisjon var den første profesjonsutdanningen som gjennomførte en nasjonal deleksamen i NOKUTs regi.

1.4.2 utfordringer

Både da vi startet og underveis var utfordringene og problemstillingene mange. Hvem skulle utvikle eksamensoppgavene, og hvem skulle kvalitetssikre dem? Skulle eksamenene telle på vitnemålet, og hva med studentenes klageadgang og øvrige rettssikkerhet? Skulle eksamenene gjennomføres digitalt eller på papir? Dette er bare noen av problemstillingene vi slet med i begynnelsen, og vi klarte å finne midlertidige løsninger både på disse og flere andre underveis, men vi går likevel litt nærmere inn på dem her, siden vi kommer tilbake til dem igjen i slutten av rapporten (se kapittel 3).

Eksamensoppgavenes kvalitet

Flere testeksperter NOKUT hadde møter med, uttrykte bekymring over at vi fikk så kort tid til å utvikle eksamenene og gjennomføre prosjektet. Storskalatester av høy kvalitet fordrer at testene er utviklet i henhold til et godt fundert rammeverk og oppfyller krav til validitet og reliabilitet. Med den korte tidsplanen ville det ikke være tid til å engasjere testeksperter som kunne bidra inn i utviklingen av rammeverk og eksamensoppgaver, og det ville heller ikke være tid til å pilotere oppgavene før de ble gitt som eksamensoppgaver til studentene. Dette var ikke uten risiko. Dersom eksamensoppgavene som ble utviklet, senere skulle vise seg ikke å gi et adekvat bilde av studentenes kunnskap, ville studentene kunne få en tellende karakter på vitnemålet som ikke var representativ for det læringsutbyttet de hadde oppnådd i det respektive faget eller emnet.

Selv om fagmiljøet i regnskap og revisjon hadde lang erfaring i å utvikle og arrangere felles eksamener, hadde hverken de, eller norsk høyere utdanning for øvrig, tradisjon for å utvikle eksamensoppgaver ut fra testmetodiske krav til validitet og reliabilitet. De ble med i prosjektet

frivillig⁵, og det var derfor et viktig hovedprinsipp at den nasjonale deleksamenen skulle avvike minst mulig fra den normen som allerede var satt for dette emnet. Det var både av faglige og tidsmessige grunner derfor uaktuelt å pålegge den etablerte fagkomiteen som utviklet de felles eksamenene, å endre noe eller å la eksterne fagmiljøer/testeksperter kvalitetssikre eksamensoppgavene. Likeledes ønsket vi også at fagmiljøene i sykepleierutdanningen og grunnskolelærerutdanningene skulle stå for oppgaveutviklingen selv, slik at rammene ble så like som mulig for de tre profesjonsutdanningene, og at de skulle få mest mulig eierskap til prosjektet ved at de fikk utforme premissene selv. NOKUT har i stedet arrangert testmetodiske seminarer for arbeidsgruppene som har utviklet eksamensoppgavene i grunnskolelærerutdanningene og i sykepleierutdanningen, så har vi latt testeksperter analysere noen av eksamensoppgavene i grunnskolelærerutdanningene og i sykepleierutdanningen, samt sensurreliabiliteten, i etterkant av sensuren. Dette vil bli nærmere beskrevet i kapittel 2.

Grunnskolelærerutdanningene: tidspunktet for eksamen, tellende eller ikke tellende?

De ulike lærerutdanningsinstitusjonene har ulike profiler, og det fantes ingen felles læringsutbyttebeskrivelser i matematikk ut over de nasjonale retningslinjene for grunnskolelærerutdanningene. Institusjonene tilbød matematikk i ulike semestre i utdanningen, slik at det kunne variere fra første til tredje semester når studentene tok matematikk. I tillegg var de fleste undervisningsplanene for 2015 allerede lagt da oppdraget om nasjonale deleksamener kom. Disse faktorene ville gjøre det vanskelig å lage én felles eksamen basert på felles læringsutbyttebeskrivelser som skulle inngå i grunnlaget for karakterfastsettelsen i faget. Fagmiljøene fant det heller ikke forsvarlig å utsette studentene for en tellende eksamen i et pilotprosjekt med så stramme tidsrammer at det kunne gå på bekostning av eksamenens kvalitet. Det var videre en forutsetning at studentene måtte ha hatt undervisning i matematikk før de skulle opp til eksamen, og det var en klar fordel at undervisningen ble gitt i samme semester som de skulle avlegge eksamen. For å få gjennomført en fullskala nasjonal deleksamen bestemte NOKUT og NRLU at det skulle avholdes to nasjonale deleksamener, én høsten 2015 og én våren 2016. Faggruppen begynte med å velge ut noen sentrale emner i matematikk, for så å lage en egen delemneplan som den nasjonale deleksamenen kunne knyttes opp til. Emnet som studentene skulle prøves i, undervisningskunnskap i brøk, desimaltall og prosentregning, ville ikke gi studiepoeng, og dette utløste spørsmålet om eksamenen kunne telle på studentenes vitnemål. Det har vært en etablert praksis at et emne må være studiepoenggivende for å kunne telle på vitnemålet. Siden denne eksamenen ikke gav studiepoeng, kunne den heller ikke være tellende, men for å få studentene til å yte sitt beste på eksamen, bestemte departementet at karakteren uansett skulle stå på vitnemålet. Det lå likevel i sakens natur at man ikke kunne kreve at studentene måtte bestå den nasjonale deleksamenen for å få utstedt vitnemål. Dette punktet skulle vise seg å by på noen utfordringer underveis, noe vi kommer tilbake til.

Sykepleierutdanningen: store variasjoner til tross for felles, obligatorisk emne

For sykepleierutdanningen eksisterte imidlertid allerede emnet/faget som det skulle arrangeres en nasjonal deleksamen i, men variasjonene mellom utdanningsinstitusjonene i forståelse og utforming av læringsutbyttebeskrivelser, undervisningsmetoder, vurderingsformer, pensum, valg av karakterskala m.m. var stor. I tillegg varierte antallet studiepoeng i anatomi og fysiologi fra åtte til tolv, og

⁵ Forslaget til forskriftsendring var ennå ikke sendt på høring da NRØA vedtok at de ønsket å la den planlagte, felles eksamenen i årsregnskap bli en nasjonal deleksamen.

undervisningen kunne også være slått sammen med andre emner. På dette tidspunktet var det 28 institusjoner som tilbød bachelorgradsstudiet i sykepleie. NPR-sykepleie så det derfor som lite realistisk å få til en fullskala nasjonal deleksamen i anatomi, fysiologi og biokjemi allerede høsten 2015. Løsningen ble å gå bort fra fullskala ved første gjennomføring og heller arrangere en frivillig nasjonal deleksamen for de institusjonene som hadde like nok forutsetninger for å kunne delta. Ved å teste ut konseptet på noen få institusjoner og finne ut hvilke forutsetninger som måtte være på plass for at nasjonale deksamener skulle kunne fungere etter hensikten, kunne dette delprosjektet både bli et reelt pilotprosjekt og en reell mulighetsstudie. For de institusjonene som skulle delta i pilotprosjektet høsten 2015, ville den nasjonale deksamenen erstatte den ordinære eksamenen i anatomi, fysiologi og biokjemi, og den ville derfor bli en tellende eksamen for de studentene som skulle gjennomføre den. Når det gjaldt vektingen av studiepoeng, ble løsningen at man ikke standardiserte antallet studiepoeng for den nasjonale deksamenen, men at institusjonene gav studentene uttelling for antallet studiepoeng som allerede var fastsatt i deres egne studieplaner. Det var ingen optimal løsning, men det var for sent å endre studieplanene underveis i semesteret; dessuten ville det å endre antallet studiepoeng i dette emnet få konsekvenser for resten av emnestrukturen i sykepleierutdanningen.

1.4.3 Logistikk og administrasjon

For NOKUTs del har gjennomføringen av de nasjonale deksamenerne bestått i 75–80 prosent administrering og koordinering og 20–25 prosent analysearbeid. En nasjonal eksamen krever nasjonal sensur, og dette har generert mye administrasjon både for NOKUT og for eksamenskontorene ved institusjonene. En viktig grunn til at arbeidsmengden har blitt så stor, er at alle eksamenene har blitt avvirket på papir. Ikke alle utdanningsinstitusjoner hadde programvare eller infrastruktur for digitale eksamener, og det eksisterte heller ikke en felles IKT-løsning hvor dataflyt kunne foregå på tvers av sektoren, og som NOKUT nærmest bare kunne koble seg til. Studenter og ansatte ved institusjoner som hadde innført digitale eksamener, opplevde det som et tilbakesteg å gå tilbake til penn og papir.

Institusjonene kunne ikke bruke sitt eget eksamenspapir, fordi sensorene da ville se hvilken institusjon den enkelte kandidat kom fra, og full anonymitet har hele tiden vært en viktig forutsetning. NOKUT har derfor bestilt eget eksamenspapir og fordelt til institusjonene før hver nasjonale deleksamen, samt generert og fordelt egne kandidatnummerserier. Institusjonene har sendt alle besvarelsene til et skannefirma som NOKUT har inngått avtale med, og skannefirmaet har så skannet alle besvarelsene og fordelt dem tilfeldig til sensorene i en elektronisk skyløsning. Det er kun NOKUTs prosjektmedarbeidere og de som har utført skanningen, som har kunnet identifisere alle kandidatenes institusjonstilhørighet. All kontakt mellom institusjonene og sensorene har gått via NOKUT.

For å synliggjøre størrelsen på og omfanget av dette arbeidet, viser vi en oversikt over antallet kandidater, sensorer og sensorpar til hver nasjonale deleksamen som har blitt gjennomført så langt.

| Tidspunkt | Eksamen | Antall kandidater | Antall sensorer | Antall sensorpar |
|-------------------|-----------------------------|-------------------|-----------------|------------------|
| 6. mai 2015 | Regnskap og revisjon | 535 | 7 | 21 |
| 1. desember 2015 | Grunnskolelærerutdanningene | 992 | 48 | 92 |
| 17. desember 2015 | Sykepleierutdanningen | 1761 | 40 | 114 |
| 4. mai 2016 | Regnskap og revisjon | 549 | 7 | 21 |
| 11. mai 2016 | Grunnskolelærerutdanningene | 997 | 48 | 92 |
| 1. desember 2016 | Grunnskolelærerutdanningene | 1057 | 50 | 100 |
| 16. desember 2016 | Sykepleierutdanningen | 4838 | 79 | 156 |

Tabell 1.1: Oversikt kandidater og sensorer

Vi har delt inn sensorene i grupper og i sensorpar, med unntak av i regnskap og revisjon, hvor alle de syv sensorene har sensurert med hverandre i til sammen 21 sensorpar. I de andre to profesjonsutdanningene har sensorene inngått i grupper på fire eller fem, hvor alle i gruppen har sensurert med hverandre parvis. For eksamenene i grunnskolelærerutdanningene og i sykepleierutdanningen i desember 2015 var antallet sensorpar henholdsvis 92 og 114. For høsten 2016, da 4838 kandidater avla den nasjonale deleksamenen i sykepleierutdanningen, delte vi inn sensorene i 156 sensorpar. Utover det arbeidet som følger med å være oppdragsgiver for sensorene (oppnevning, utbetaling av honorarer og øvrig kontakt m.m.), har NOKUT tatt imot karakterprotokoller fra hvert sensorpar og videreformidlet alle resultatene til institusjonene. Ønsker om begrunnelser og ny sensur har også blitt behandlet med full anonymitet, og også her har NOKUT fungert som mellomledd mellom institusjonene og sensorene. Samarbeidet og arbeidsdelingen mellom NOKUT og eksamenskontorene og mellom NOKUT og sensorene har fungert greit, men har neppe vært optimal for noen av partene. NOKUT har ingen tidligere erfaring med eksamens- og studieadministrasjon, og vi har vært nødt til å finne ad hoc-løsninger underveis. Slike ad hoc-systemer er ikke uten sårbarhet, og risikoen for å gjøre feil er åpenbart til stede.

Alt i alt har mellom 150 og 160 eksterne fagpersoner jobbet på oppdrag for NOKUT i forbindelse med nasjonal deleksamen. Dette inkluderer arbeidsgruppene som har utviklet eksamensoppgavene, sensorkorpene og noen testeksperter. Hvis vi i tillegg tar med kontaktpersoner for tjenestene vi har kjøpt og studieadministrativt ansatte ved institusjonene, blir antallet personer som NOKUT har involvert i prosjektet, langt høyere.

2 Analyser og resultater

I dette kapitlet vurderer vi i hvilken grad prosjektet har lyktes i å si noe om de eksplisitte formålene Kunnskapsdepartementet beskrev i oppdragsbrevet fra 2014. Kunnskapsdepartementet spesifiserte tre hovedformål. Nasjonal deleksamen skal:

- gi nyttig informasjon om studentenes kunnskapsnivå (formål 1)
- gi fagmiljøene mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner, noe som igjen skal gi miljøene grunnlag for utvikling (formål 2)
- bidra til økt tillit til utdanningene i samfunnet (formål 3)

I tillegg la NOKUT til grunn at det ut fra et kvalitetsperspektiv er et mål i seg selv at en gitt karakter skal reflektere samme kunnskapsnivå, uavhengig av hvilken institusjon eksamenen er avlagt ved, jf. SØF-rapport nr. 03/13, s. 1:

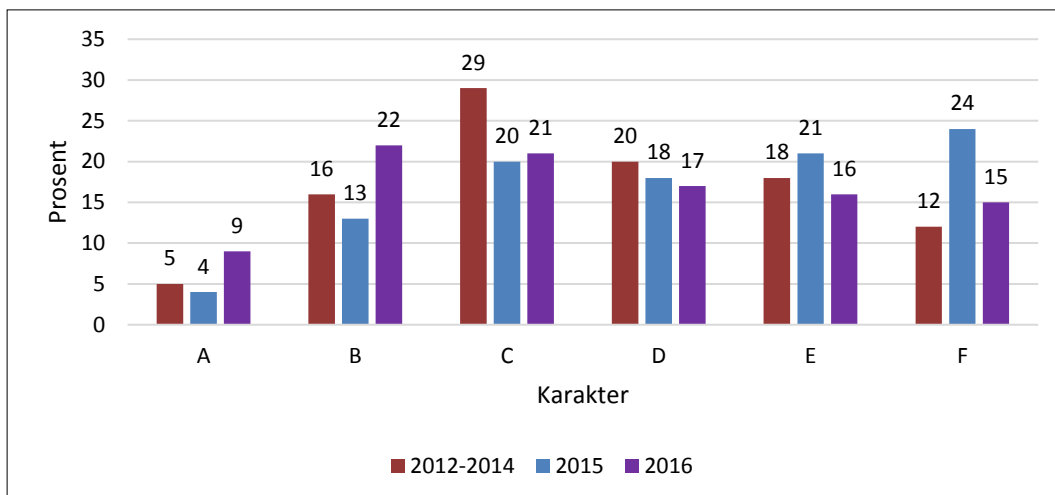
Intensjonen ved dagens karaktersystem er at en gitt karakter skal reflektere samme kunnskapsnivå uavhengig av hvilken institusjon eksamenen er avlagt på. Da vil karakterene gi korrekt informasjon om studentenes ferdigheter og innsats til andre utdanningsinstitusjoner (dersom studentene bytter institusjon under utdanningsløpet) og til potensielle arbeidsgivere.

Det siste formålet er dermed karakterkalibrering (formål 4).

2.1 Formål 1: Gi nyttig informasjon om studentenes kunnskapsnivå

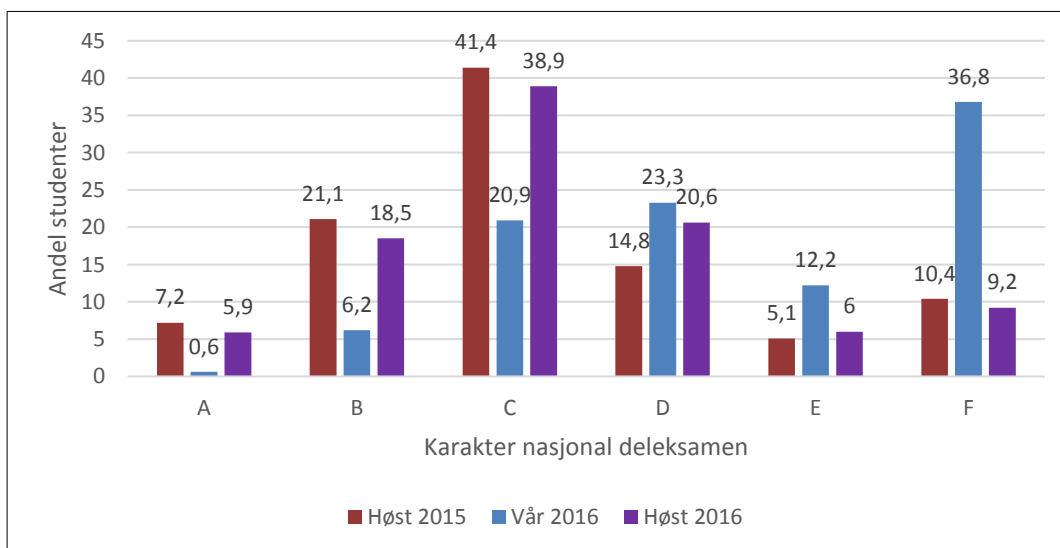
For å kunne si noe nyttig om studentenes kunnskapsnivå er det helt sentralt at sensurordningen sikrer likebehandling av alle studenter. Det vil si at all sensur er blind, at alle studenter vurderes etter de samme kriteriene og at ingen sensorer vet hvilken institusjon studentene tilhører. Basert på de analysene vi har gjort av sensuren, eksamenssettene og studentenes resultater, mener vi at de resultatene vi har presentert i tidligere rapporter, gir oss nyttig informasjon om studentenes kunnskapsnivå.

Figurene under (figur 2.1–2.4) viser karakterfordelingen for hvert fag på de enkelte eksamenene.



Figur 2.1: Karakterfordelinger i årsregnskap

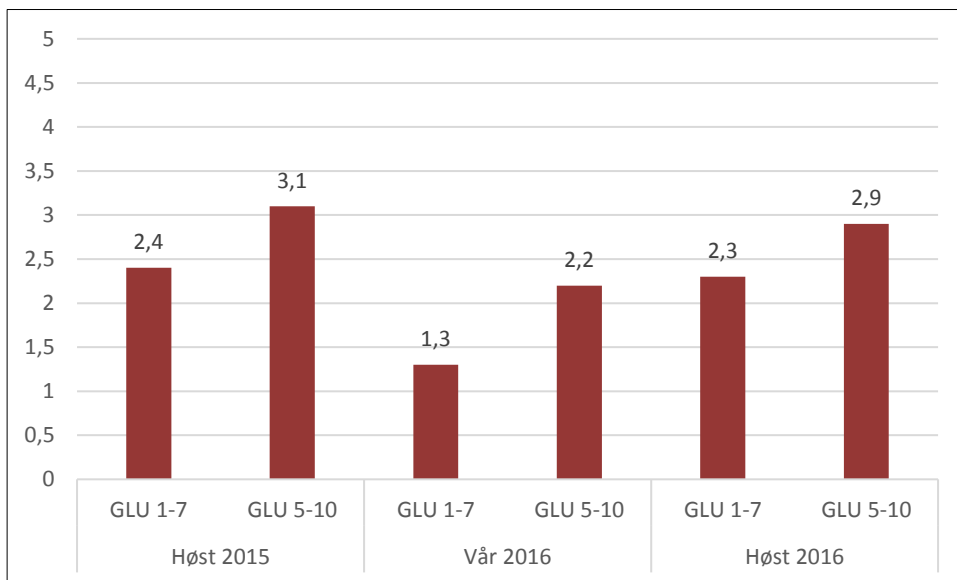
Etter den første nasjonale deleksamenen i årsregnskap (blå stolper) sammenlignet NOKUT resultatene fra tre tidligere eksamener, hvor sensuren var lokal (rød stolper). Denne sammenligningen viste at resultatene fra den første nasjonale deleksamenen var signifikant svakere enn resultatene fra de tre foregående årene. NOKUT rapporterte etter denne eksamenen at det kunne se ut som om lokal sensur var noe «snillere» enn nasjonal sensur. Etter den andre nasjonale deleksamenen (lilla stolper) kan det se ut som om at eksamenen enten var noe vanskeligere i 2015, eller at sensuren i 2015 var noe streng. Resultatene fra 2016 er, som man ser, ganske like resultatene fra årene før man hadde den første nasjonale deleksamenen i årsregnskap. Vi henviser til den andre delrapporten for nasjonal deleksamen i årsregnskap for mer detaljerte analyser.



Figur 2.2: Karakterfordeling i matematikk for GLU-studenter

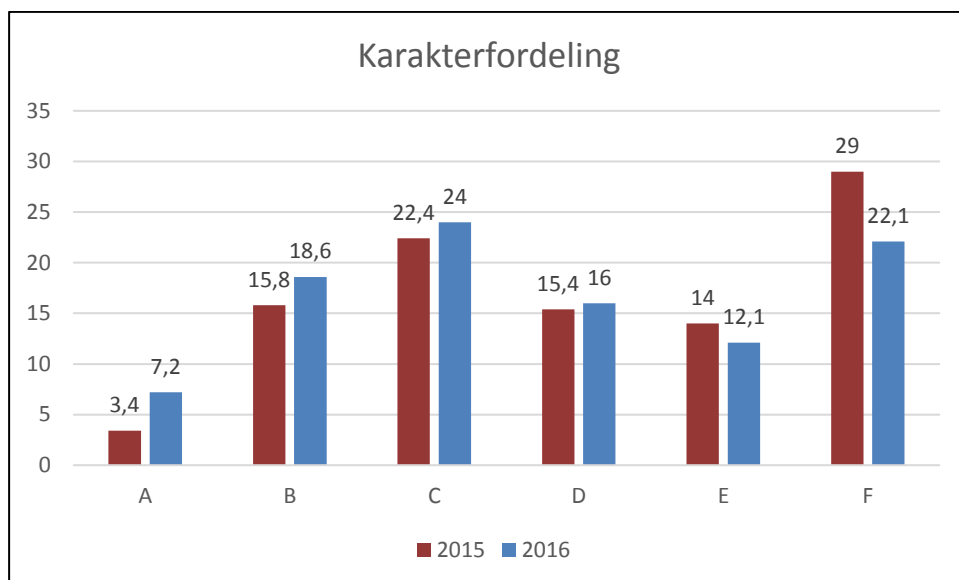
Resultatene fra høsten 2015 og høsten 2016 er veldig like. Resultatene fra våren 2016 var meget svake, og NOKUT argumenterte for at den viktigste årsaken til dette var at eksamenen ikke var tellende for studentene. Høsten 2016 bestemte Kunnskapsdepartementet at nasjonal deleksamen i matematikk for grunnskolelærerstudenter skulle være tellende. Den store forbedringen i resultatene indikerer at vårens analyse var korrekt. En kritikk mot vårens analyse var at majoriteten av studentene på vårens eksamen gikk på GLU 1–7, og at disse studentene tar matematikk som et obligatorisk fag,

mens majoriteten av studenter som tar eksamen på høsten, går på GLU 5–10 og tar matematikk som et valgfag. Dette er et sterkt motargument, men dersom vi skiller 1–7- og 5–10-studentene, ser vi at begge gruppenes resultater er signifikant bedre på høstens eksamen enn på vårens eksamen, noe som igjen støtter opp om argumentet om at studentene nedprioriterte eksamenen våren 2016.



Figur 2.3: Gjennomsnittskarakter i matematikk for GLU-studenter

Sykepleierstudentene har hatt to nasjonale deksamener i anatomi, fysiologi og biokjemi. Resultatene fra høsten 2015 var, som fagmiljøene forventet, svake. Resultatene fra eksamenen høsten 2016 var en del bedre. Høsten 2015 fikk ca. 42 prosent av studentene karakteren C eller bedre, mens hele 29 prosent strøk. Høsten 2016 fikk ca. 46 prosent av studentene karakteren C eller bedre, mens strykprosenten gikk ned til 22 prosent. Det er vanskelig for NOKUT å finne årsaken til denne forbedringen. Høsten 2015 var det kun 12 institusjoner (fordelt på 14 studiesteder) som deltok i pilotprosjektet, mens alle institusjonene som tilbyr bachelorgradsstudiet i sykepleie (32 studiesteder), deltok høsten 2016. Den andre delrapporten for nasjonal deksamener i anatomi, fysiologi og biokjemi går mer i detalj og sammenligner blant annet de institusjonene som har vært med begge årene.



Figur 2.4: Karakterfordeling i anatomi, fysiologi og biokjemi for sykepleierstudenter

2.2 Formål 2: Gi fagmiljøene mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner

NOKUT har i tidligere delrapporter oppgitt resultater og karakterfordeling på institusjonsnivå, og vi gjør derfor ikke dette her. Oversikten over resultater og karakterfordeling på institusjonsnivå gir institusjonene gode muligheter til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner. Man skal likevel være forsiktig med å sammenligne resultatene uten å ta hensyn til andre faktorer som kan påvirke studentenes resultater. Dette gjelder for eksempel studentenes startkompetanse, alder og kjønn. På institusjonsnivå kan resultatene også påvirkes av når i studiet studentene avlegger nasjonal deleksamen. NOKUT har gjennom en databehandleravtale med Kunnskapsdepartementet fått tilgang til bakgrunnsdata på individnivå fra CERES (tidligere FSAT). Disse dataene gjør det mulig for NOKUT å vurdere hva som forklarer variasjonen i studentenes resultater, samt å bruke disse resultatene til å måle institusjonenes bidrag til studentenes resultater. Som man vil se av analysene under, har regresjonsmodellen begrenset forklaringskraft. Det vil si at analysene kan gi oss noen indikasjoner på viktige forklaringsvariabler, samt institusjonenes bidrag. Denne typen analyser kan likevel hjelpe institusjonene og fagmiljøene til å få en bedre forståelse av hvor de står og gi dem viktig informasjon om hvilke institusjoner det kan være hensiktsmessig å samarbeide med for heve kvaliteten på utdanningen.

I analysen nedenfor ser vi spesifikt på nasjonal deleksamen i matematikk for GLU-studentene høsten 2015.

| Avhengig variabel: Totalt antall poeng (0–40) | | |
|--|-------------------|--------------------|
| | Modell 1 | Modell 2 |
| Matematikk karakter fra VGS | 2,7*** (0,24) | |
| Karakterpoeng fra VGS | 0,29*** (0,07) | 0,45*** (0,04) |
| Antall semester | -0,1 (0,24) | -0,40* (0,24) |
| Alder | 0,23 (0,05) | 0,18*** (0,05) |
| Kvinne | -0,78 (0,69) | -1,52*** (0,54) |
| GLU 5–10 | 2,18*** (0,65) | 3,57 (0,05) |
| Konstant | -4,94 (2,91) | 2,03 (2,33) |
| N | 501 | 740 |

***p > .01, **p > .05, *p > .1

Tabell 2.1: Regresjonsanalyse nasjonal deleksamen GLU høsten 2015

Regresjonsmodellene over viser at inntakskvaliteten, både med tanke på matematikkarakter fra videregående skole og karakterpoeng fra videregående skole, har positive effekter på studentenes resultater. I modell 1 ser vi at for hver karakter bedre en student har i matematikk fra videregående skole, gir det 2,7 poeng mer i uttelling på den nasjonale deleksamenen. I modell 2 har vi fjernet matematikkarakteren fra regresjonsanalysen, fordi vi mangler matematikkarakter for mange av studentene. Ved å fjerne dette fra modellen øker vi antallet observasjoner.

Når vi fjerner matematikkarakteren i modell 2, ser vi at studentenes starkompetanse (målt som karakterpoeng fra videregående skole) er en viktig forklaringsvariabel. En økning på 1 karakterpoeng fører til ca. 0,5 poeng mer på den nasjonale deleksamenen. Studenter med 35 karakterpoeng fra videregående skole får altså rundt 5 poeng færre (av 40) enn en student med 45 karakterpoeng fra videregående skole. Modell 2 viser også at kvinner får ca. 1,5 poeng lavere enn menn, men én av grunnene til dette er at kvinneandelen er høyere i GLU 1–7, og at GLU 1–7-studenter skårer ca. 3,5 poeng lavere enn GLU 5–10-studenter.

Institusjonsbidrag

I regresjonsmodellene over så vi på hvordan individspesifikke faktorer påvirket studentenes resultater. Regresjonsmodellene kan også brukes til å vurdere institusjonenes bidrag til studentenes resultater på de nasjonale deleksamenene. Institusjonsbidragsmodellen som presenteres her, er en enkel modell som gir indikasjoner på institusjonenes bidrag, men den gir ikke noe fasitsvar.

For å «måle» institusjonsbidraget bruker man regresjonsanalysen på individnivå til å predikere hva gjennomsnittspoengscoren per institusjon skulle ha vært basert på studentenes startkompetanse, kjønn, alder, hvilket semester de var i da de tok eksamen og om de gikk på GLU 1–7 eller 5–10. Man ser så på forskjellen mellom den predikerte gjennomsnittsscoren og den faktiske gjennomsnittsscoren.

Institusjoner med en faktisk gjennomsnittsscore som er høyere enn den predikerte gjennomsnittsscoren, har et positivt institusjonsbidrag, mens institusjoner med en faktisk gjennomsnittsscore som er lavere enn den predikerte gjennomsnittsscoren, har et negativt institusjonsbidrag. I tabellen nedenfor illustrerer vi metoden basert på resultater og data fra den nasjonale deleksamenen i matematikk for GLU høsten 2015.

| Institusjon | Antall studenter | Faktisk gj.snittscore | Predikert gj.snittscore | Differanse |
|-------------|------------------|-----------------------|-------------------------|------------|
| HiSF | 78 | 26,3 | 23,0 | 3,3 |
| HBV | 56 | 28,4 | 26,9 | 1,5 |
| UiS | 128 | 25,7 | 24,5 | 1,2 |
| HiT | 63 | 25,6 | 24,6 | 1,0 |
| HiOA | 81 | 29,1 | 28,6 | 0,5 |
| HiVolda | 62 | 25,6 | 25,2 | 0,4 |
| HiST | 139 | 28,2 | 28,0 | 0,3 |
| HiØ | 59 | 26,7 | 27,7 | -1,0 |
| UiT | 38 | 24,3 | 25,3 | -1,0 |
| HiNT | 117 | 21,9 | 23,7 | -1,8 |
| NLA | 54 | 21,9 | 25,1 | -3,2 |

Tabell: 2.2: Institusjonsbidrag poeng GLU høsten 2015

Tabellen over viser institusjonsbidraget til 11 institusjoner hvor mer enn 30 studenter avla eksamen i desember 2015. Vi bruker Høgskulen i Sogn og Fjordane som eksempel. 78 studenter tok eksamen, og gjennomsnittsscoren var 26,3 poeng (man kan oppnå maksimalt 40 poeng). Basert på startkompetanse, kjønn, alder, hvor lenge studentene har studert og om de går på GLU 1–7 eller 5–10, predikerer modellen at gjennomsnittsscoren er 23 poeng. Studentene fikk altså 3,3 poeng mer i gjennomsnitt enn det man kunne forvente basert på modellen. Ser man på NLA Høgskolen, fikk studentene der 3,2 poeng mindre i gjennomsnitt enn modellen predikerer.

Hva betyr +/- 3 poeng karaktermessig? I tabellen under transformerer vi råscoren til bokstavkarakterer. Som vi ser, er den faktiske gjennomsnittskarakteren for alle institusjonene, bortsett fra Høgskolen i Nord-Trøndelag og NLA Høgskolen, den samme som den predikerte karakteren. For Høgskolen i Nord-Trøndelag og NLA Høgskolen er den faktiske gjennomsnittskarakteren én karakter lavere enn den predikerte karakteren (men, den faktiske karakteren på begge høgskolene lå tett opp til C).

| Institusjon | Antall studenter | Faktisk gj.snitt | Predikert gj.snitt | Differanse |
|-------------|------------------|------------------|--------------------|------------------|
| HiSF | 78 | C | C | Ingen differanse |
| HBV | 56 | C | C | Ingen differanse |
| UiS | 128 | C | C | Ingen differanse |
| HiT | 63 | C | C | Ingen differanse |
| HiOA | 81 | C | C | Ingen differanse |
| HiVolda | 62 | C | C | Ingen differanse |
| HiST | 139 | C | C | Ingen differanse |
| HiØ | 59 | C | C | Ingen differanse |

| | | | | |
|------|-----|---|---|---------------------|
| UiT | 38 | C | C | Ingen differanse |
| HiNT | 117 | D | C | Én karakter svakere |
| NLA | 54 | D | C | Én karakter svakere |

Tabell 2.3: Institusjonsbidrag karakter GLU høsten 2015

Hva forklarer disse forskjellene? Vi har ikke nok, eller riktig type data per i dag til å svare på dette spørsmålet, men én mulighet er at forhold knyttet til institusjonen, for eksempel undervisningsformer, pensum, eller andre institusjonsspesifikke faktorer forklarer forskjellen.

Denne typen analyser gir fagmiljøene mulighet til å sammenligne seg selv med andre, og forhåpentligvis samarbeide med hverandre for å finne ut av hva man kan gjøre for å løfte studentenes ferdigheter og kunnskaper innen faget.

Analysen over er altså en enkel modell basert på én eksamen og et relativt lite antall studenter. Det betyr at resultatene ikke er særlig robuste og at tilfeldigheter kan forårsake en del av variasjonen. Dersom man skal fortsette med nasjonale deleksamener, bør det legges til rette for at man kan gjøre lignende analyser over tid, og at fagmiljøene selv får tilgang til dataene, slik at de selv både kan gjøre denne typen analyse og mer avanserte analyser.

2.3 Formål 3 (og 4): Bidra til økt tillit til utdanningene i samfunnet

En viktig forutsetning for å øke tilliten til utdanningene er at samfunnet har tillit til at karakterene studentene oppnår, reflekterer det samme kunnskaps- og ferdighetsnivået uansett hvor en kandidat har studert. Dette formålet er derfor knyttet tett til NOKUTs eget formål om å bidra til karakterkalibrering.

For å sørge for at alle studentene blir vurdert ut fra de samme kriteriene har eksamensgruppene laget meget detaljerte sensorveiledninger. All sensur har vært blind, og alle kandidater har fått eksamensbesvarelsene sine sensurert av to sensorer.

Sensorordningen har også gitt NOKUT mulighet til å vurdere reliabiliteten på oppgavene og sensuren. I samarbeid med eksamensgruppen ba NOKUT alle sensorene om å rapportere detaljert sensur. Det vil si at vi ba hver sensor om å oppgi hvor mange poeng hver kandidat fikk på hver enkelt eksamensoppgave. Sensorene fikk et poengskjema hvor de skulle fylle inn kandidatnummer og poengsummen kandidaten oppnådde på oppgave 1a, 1b, 1c, osv. Siden hver oppgave ble rettet av to sensorer, kan vi vurdere hvor enige sensorene var på hver enkelt oppgave, samt på totalscoren til hver enkelt kandidat. Sensorene ble så enige seg imellom om hvor mange poeng kandidaten skulle få på hver oppgave og på totalscoren. På denne måten har NOKUT mulighet til å vurdere sensurreliabiliteten, samt reliabiliteten på eksamensoppgavene. Tabellen under viser hva NOKUT mottok fra hver sensor.

| Oppgave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | Totalt |
|-----------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|--------|
| Sensor 1 | 2 | 2 | 1 | 1 | 2 | 2 | 0 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 20 |
| Sensor 1 | 2 | 2 | 1 | 1 | 1 | 2 | 0 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 19 |
| Samsensur poeng | 2 | 2 | 1 | 1 | 2 | 2 | 0 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 20 |

Tabell 2.4: Karakterskjema

Hver sensor fylte inn «sin» rad, og sensorene ble enige om hvor mange poeng hver kandidat skulle få på hver deloppgave, samt totalscoren. I eksempelet over ser vi at begge sensorene ga samme antall poeng på alle oppgaver, unntatt oppgave 5. Der ble de enige om hvor mange poeng kandidaten skulle få.

Denne sensurordningen har mange fordeler. Den gir oss mulighet til å kvalitetssikre oppgavene og sensorveiledningen, vurdere reliabiliteten på sensuren, gi fagmiljøene tilbakemeldinger om hvordan studentene gjør det på forskjellige type oppgaver, og den gir oss mulighet til å gi studentene detaljerte tilbakemeldinger om deres resultater.

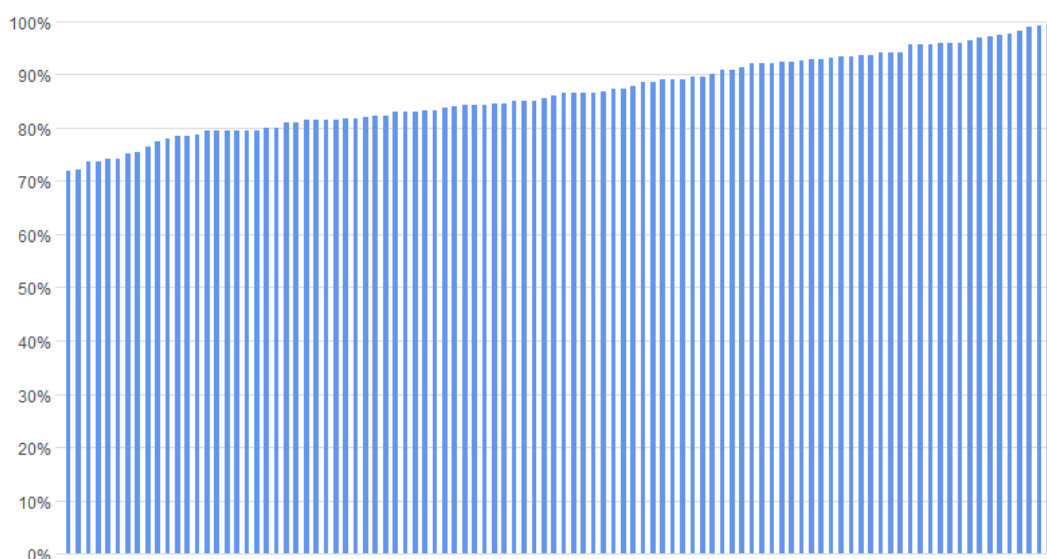
Kvalitetssikring av oppgavene

NOKUT har ikke gjort statistiske analyser for å vurdere hvor valid eksamenssettene var. Eksamensgruppene er bredt sammensatt med fagansatte fra flere universiteter og høyskoler som både deltok og ikke deltok på eksamen. Disse eksamensgruppene har jobbet seg gjennom emneplaner og pensum og lagd oppgaver de mener måler det de ønsker å måle. Etter sensuren har NOKUT sendt en spørreundersøkelse til sensorene og bedt dem om å vurdere hvor vanskelig eksamenssettet var, hvor dekkende det var og hvor klart formulert det var. NOKUT og eksamensgruppene er totalt sett relativt sikre på at validiteten av eksamenssettene har vært relativt god.

Sensurreliabilitet

NOKUT har selv vurdert reliabiliteten på sensuren for alle tre GLU-eksamenene og eksamenen for sykepleierstudentene høsten 2016. Nedenfor viser vi et eksempel fra GLU-eksamenen høsten 2016.

Som beskrevet over var det totalt 50 sensorer som sensurerte den nasjonale deleksamenen i matematikk for GLU høsten 2016. Disse ble delt inn i grupper på fem, og i hver gruppe sensurerte alle sensorene sammen med hverandre. Til sammen var det 100 sensorpar, og hvert sensorpar samsensurerte 10 eller 11 eksamensbesvarelser. Hvert eksamenssett inneholdt 20 spørsmål, så i gjennomsnitt sensurerte hvert sensorpar mellom 200 og 220 enkeltoppgaver sammen. Figur 2.5 viser hvor mange prosent av disse oppgavene hvert sensorpar har vurdert likt, der hver stolpe representerer ett sensorpar, og høyden på stolpen viser hvor mange prosent av oppgavene sensorene har vurdert likt. Hver enkelt sensor vurderte i gjennomsnitt 42 besvarelser.



Figur 2.5: Sensurreliabilitet: Prosentvis enighet.

Resultatene er svært gode. En stor majoritet av sensorparene gir samme poengsum til kandidatene på nesten alle oppgavene. Når man i tillegg vet at kandidatenes karakterer er basert på samsensur, er det en meget liten sannsynlighet for at sensuren på noen måte er tilfeldig. Det er i all hovedsak to årsaker til at sensurreliabiliteten er så høy. For det første vitner det om at oppgavesettet gjør det mulig for sensorene enkelt å vurdere hvor godt et svar er, og for det andre har sensorveiledningen vært god og tydelig. Basert på sensuranalysen kan vi si at det er lite sannsynlig at kandidatene hadde fått et annet resultat på eksamen dersom de hadde blitt sensurert av et annet sensorpar.

Resultatene av klagesensuren bekrefter den høye sensurreliabiliteten. Etter matematikkeksamnen i grunnskolelærerutdanningene høsten 2016 var det totalt 37 studenter som klagde og fikk ny sensur. Av disse gikk seks kandidater opp én karakter, mens én kandidat gikk opp mer enn én karakter.

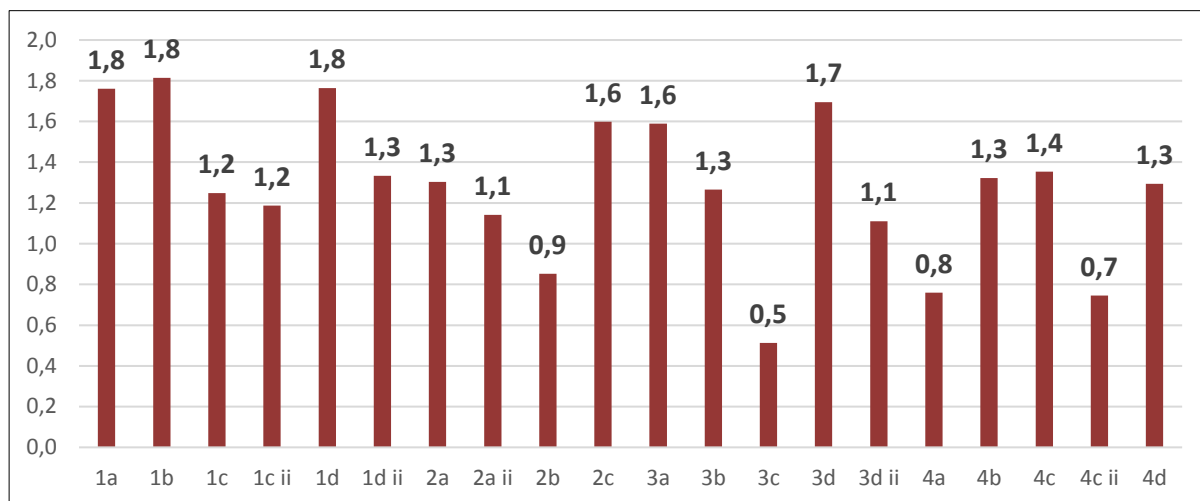
| Opprinnelig karakter | Antall | Antall forbedret karakter | Antall dårligere karakter | Opp 1 karakter | Ned 1 karakter | Opp mer enn 1 karakter |
|----------------------|--------|---------------------------|---------------------------|----------------|----------------|------------------------|
| F | 30 | 4 | 0 | 3 | 0 | 1 |
| E | 2 | 1 | 0 | 1 | 0 | 0 |
| D | 1 | 1 | 0 | 1 | 0 | 0 |
| C | 3 | 0 | 0 | 0 | 0 | 0 |
| B | 1 | 0 | 0 | 0 | 0 | 0 |
| SUM | 37 | 6 | 0 | 5 | 0 | 1 |

Tabell 2.5: Klagesensur GLU høsten 2016

Detaljerte tilbakemeldinger til institusjonene

Detaljert sensur gir NOKUT og fagmiljøene mulighet til å se hvordan studentene presterer på forskjellige type oppgaver. Nasjonal deleksamen i matematikk for GLU-studenter inneholder forskjellige typer med oppgaver. Noen er fokusert på matematikk-didaktikk, mens andre går på matematikkferdigheter. I figur 2.6 kan vi se det nasjonale gjennomsnittet per deloppgave. Denne typen

analyse kan enkelt gjøres på institusjonsnivå, slik at fagmiljøene også kan sammenligne seg med andre for å se hvor andre institusjoner gjør det bedre eller svakere.



Figur 2.6: Nasjonalt gjennomsnitt per oppgave GLU høsten 2015

Detaljerte tilbakemeldinger til studentene

Sensurordningen gir oss også en mulighet til å gi studentene detaljerte tilbakemeldinger. I de eksamenene hvor sensorene har oppgitt detaljert samsensur, har NOKUT gitt studenter som ønsker karakterbegrunnelse, en oversikt over hvor mange poeng de fikk på hver oppgave, samt sensorveiledningen. Dette gir studentene en god oversikt over hvordan de gjorde det på hver enkelt oppgave, samtidig som de kan vurdere hva de skulle ha gjort for å oppnå flere poeng.

Oppsummering analyser og resultater

Analysene og resultatene som er presentert i dette kapittelet, samt i de individuelle delrapportene, viser at nasjonale deleksamener kan gi oss viktig informasjon om studentenes kunnskapsnivå, at det gir fagmiljøer mulighet til å sammenligne seg selv med andre, og at det kan bidra til økt tillit til utdanningen. Prosjektet har også gitt oss mulighet til å kalibrere karaktersetningen på tvers av institusjoner. Samtidig er det viktig å påpeke at dersom man skal nå alle målene man har satt seg, så er det mange forutsetninger som må være på plass. I det neste kapittelet diskuterer vi disse forutsetningene mer i detalj.

3 utfordringer som må løses før nasjonale deleksamener kan innføres

I skrivende stund har NOKUT i samarbeid med institusjonene og fagmiljøene gjennomført til sammen syv nasjonale deleksamener i tre rammeplanstyrte utdanninger: tre i grunnskolelærerutdanningene, to i bachelorgradsstudiet i sykepleie og to i bachelorgradsstudiet i regnskap og revisjon. Som vi har påpekt i de første delrapportene etter hver enkelt eksamen, har vi møtt på mange utfordringer. Dette være seg

juridiske, prinsipielle, faglige og praktiske. Selv om vi har klart å løse mange av dem underveis, er det fortsatt en rekke utfordringer som bør løses før nasjonale deleksamener innføres som en fast ordning i høyere utdanning.

Avklaring av faglig frihet versus politisk og samfunnsmessig kontroll

En del av kritikken mot nasjonale deleksamener har handlet om at nasjonale deleksamener går på bekostning av akademisk frihet og institusjonell autonomi. Isolert sett behøver ikke en nasjonal deleksamen i en rammeplanstyrt utdanning å representere en trussel mot akademisk frihet og institusjonell autonomi, siden disse utdanningene allerede i stor grad er standardiserte. Likevel har institusjonene en viss frihet til å velge sin egen profil på utdanningene. En nasjonal eksamen vil i større eller mindre grad med nødvendighet gripe inn i denne friheten. For at studentene skal kunne vurderes så likt som mulig og på mest mulig like betingelser, krever det for eksempel at eksamenen kommer på samme tid i studieløpet, at innholdet i og omfanget av emnet studentene skal prøves i er likt, og at læringsutbyttebeskrivelsene stort sett er de samme uavhengig av studiested. Departementet varsler i Meld. St. 16 at de på sikt vil redusere bruken av rammeplaner, og at de er enig i mye av kritikken som har blitt rettet mot innholdet i rammeplanene (s. 94). Denne kritikken har bestått i at de begrenser institusjonenes faglige frihet og hemmer nytenkning, en kritikk som også har blitt rettet mot nasjonale deleksamener. Det at departementet på sikt ønsker å begrense bruken av rammeplaner fordi det begrenser faglig frihet, men samtidig ønsker å innføre nasjonale deleksamener, som av mange høringsinstanser oppfattes som en ytterligere innskrenkning av den faglige friheten, kan synes som en motsetning.

Unngå «teaching to the test»

Kritikerne av nasjonale deleksamener har også pekt på risikoen for at både faglærere og studenter bruker mer tid på de emnene/delemnene som studentene testes i til nasjonal deleksamen, enn på andre emner/delemner i faget eller i utdanningen («teaching to the test»). Denne risikoen vil alltid være der, og NOKUT kan heller ikke utelukke at man i prosjektperioden lokalt har brukt mer tid på å forberede seg til nasjonal deleksamen enn til andre eksamener eller annen læring. For å demme opp for denne risikoen bør et helt fag eller et helt emne være gjenstand for eksamen, slik man eksempelvis har forsøkt å få til i sykepleierutdanningen ved at den nasjonale deleksamenen dekker hele emnet anatomi, fysiologi og biokjemi. I grunnskolelærerutdanningene har man valgt å teste studentene i et mindre emne/delemne (undervisningskunnskap i brøk, desimaltall og prosentregning) i stedet for et større emne som tilsvarer for eksempel 15 studiepoeng, fordi tidsplanen ikke tillot noe annet. Hvis en nasjonal deleksamen i matematikk dekker et større emne, vil man unngå at ett delemne får større oppmerksomhet enn de andre, og man slipper å måtte skifte delemne etter en periode. De som utvikler oppgavene, vil kunne etablere og også pilotere en oppgavebank som dekker hele emnet, i stedet for å måtte begynne på nytt med et annet delemne etter en tid. Det vil også gi et bedre sammenligningsgrunnlag over tid, som man ikke vil få dersom man bytter emner/delemner etter en periode.

Eksamenene må være tellende og studiepoenggivende

De to første eksamenene i grunnskolelærerutdanningene var ikke tellende for studentene, og studenter som ikke bestod den nasjonale deleksamenen vil likevel få utstedt vitnemål. Årsaken var at den nasjonale deleksamenen ikke ga uttelling i form av studiepoeng, fordi den ikke var knyttet opp til et

eget emne i studieplanen. Da det ble kjent blant studentene at eksamenen ikke var tellende, påvirket det eksamensresultatene negativt. Eksamener er tids- og ressurskrevende. Å avholde en isolert eksamen som ikke gir studiepoeng er ineffektivt og har blitt dårlig mottatt av både studentene og fagmiljøene. NOKUTs anbefaling er at nasjonale deksamener ikke bare skal telle på vitnemålet, men at de også gir uttelling i form av studiepoeng.

Tidspunkt i studieløpet: hva er det vi måler?

Tidspunktet for en nasjonal deksamener er heller ikke uvesentlig. De nasjonale deksamenerne i grunnskolelærerutdanningene og i sykepleierutdanningen har kommet tidlig i studieløpet, primært i første semester for de fleste. Som vi har påvist, er det en stor sammenheng mellom studentenes startkompetanse og resultatene på nasjonal deksamener. For å unngå at vi ender opp med å måle startkompetansen i stedet for studentenes læringsutbytte og institusjonenes bidrag til studentenes læring i løpet av studiet, vil det være en fordel å legge de nasjonale deksamenerne til et senere tidspunkt. Det vil på den annen side gjøre det vanskelig i sykepleierutdanningen, siden anatomi, fysiologi og biokjemi er et grunnleggende fag og derfor gis tidlig i studiet. En grundig vurdering av tidspunkt i studieløpet bør likevel gjøres før nasjonale deksamener innføres fast.

Tidspunktet for nasjonal deksamener bør være det samme for studentene

I henhold til den nye rammeplanen for femårig grunnskolelærerutdanning kan tidspunktet for matematikk i GLU 1–7 i teorien variere fra første til tredje studieår. Dersom det skal variere fra første til tredje studieår når studentene gjennomfører nasjonal deksamener, vil det gi svært usikre resultater. Slik matematikk er et modningsfag, er likeledes det å studere også en modningsprosess. Desto senere studentene gjennomfører nasjonal deksamener, desto flere fordeler vil de kunne ha sammenlignet med dem som tar nasjonal deksamener tidlig i studieløpet, som høyere grad av modenhet, bedre studieteknikk og lenger erfaring i å besvare eksamensoppgaver generelt. Dersom man på mest mulig objektivt grunnlag skal kunne måle studentenes kunnskapsnivå og institusjonenes bidrag, bør tidspunktet for en nasjonal deksamener være det samme for alle. Det har ikke vært mulig å legge eksamen til samme tidspunkt for grunnskolelærerstudentene i prosjektperioden. Hvis nasjonale deksamener skal videreføres, bør denne problemstillingen vurderes nøye.

GLU 1–7 og GLU 5–10 bør holdes fra hverandre

I prosjektperioden har studenter fra begge grunnskolelærerutdanningene, GLU 1–7 og GLU 5–10, gjennomført den samme eksamenen. Som et eksperiment har det vært nyttig og interessant, fordi vi har kunnet påvise signifikante resultatforskjeller mellom de to studentgruppene. Dersom nasjonale deksamener skal videreføres i grunnskolelærerutdanningene, bør disse studentgruppene imidlertid holdes adskilt. De stiller med ulike forutsetninger og ulik motivasjon, men siden det er anonym sensur, blir de likevel vurdert etter de samme kriteriene. Som vi allerede har vært inne på bør studentene testes på mest mulig likt grunnlag og med mest mulig like forutsetninger.

Eksamensoppgavene må kvalitetssikres

Før nasjonale deksamener innføres, må fagmiljøene få tilstrekkelig med tid til å utvikle og pilotere eksamenene og kunne integrere dem i emnestrukturen og i studiet som helhet. Det bør være bred

konsensus i fagmiljøene om hvilke deler av læringsutbyttet en nasjonal deleksamen skal teste studentene i. Dette er viktig av hensyn til testenens validitet, men også av hensyn til testenens legitimitet blant de fagansvarlige. Hvis de som skal undervise i faget ikke har tiltro til at nasjonale deksamener kan gi et riktig bilde av studentenes kunnskapsnivå, vil de heller ikke ha tiltro til at resultatene kan brukes til kvalitetsutvikling.

Sensurfrist og sensoroppnevning bør ses nærmere på

Sensorene har gitt tilbakemelding om at det er ekstra krevende å sensurere nasjonale deksamener. For noen eksamener er det mange besvarelser som skal vurderes, de skal sensurere sammen med flere andre, hvor alle skal fylle ut et detaljert skjema for samsensur, og mange synes det er krevende å forholde seg til en omfattende og detaljert sensorveiledning som de selv ikke har hatt noen innvirkning på. Lovens krav om at sensuren skal foreligge innen tre uker, er ikke tilpasset nasjonale deksamener og skaper høyt tidspress for sensorene og for NOKUT. Sensurfristen kan gjøre det vanskelig å skaffe tilstrekkelig antall sensorer, samtidig som den korte tidsfristen skaper behov for flere sensorer. Dersom sensurfristen hadde vært lenger, kunne man ha klart seg med færre sensorer og dermed fått en enda høyere sensurreliabilitet. Ordningen med at hver institusjon på oppfordring fra NOKUT må oppnevne hver enkelt sensor i sensorkorpsene i forkant av hver nasjonale deleksamen er heller ikke optimal og bør ses nærmere på. For å skape forutsigbarhet for sensorene og for institusjonene bør hvert sensorkorps oppnevnes for lenger tid av gangen.

Nasjonale deksamener må innbefatte kontinuasjonseksamener

Av hensyn til tid og logistikk har det ikke vært mulig å inkludere kontinuasjonseksamener i pilotprosjektet. Studentene har like fullt krav på å få gå opp til en kontinuasjonseksamen, dessuten må sykepleierstudentene bestå basisfagene for å kunne gå videre i studiet. I regnskap og revisjon, hvor felles eksamener har vært praktisert lenge, avholder institusjonene kontinuasjonseksamenene lokalt. Vi besluttet derfor å følge samme ordning for sykepleierutdanningen, selv om den nok ikke ble oppfattet som rettferdig av studentene og fagmiljøet. For grunnskolelærerutdanningene har det blitt avholdt to nasjonale deksamener årlig, én om høsten og én om våren. Studenter som ikke har bestått eksamen, eller som har ønsket å forbedre karakteren sin, har da fått en ny mulighet påfølgende semester. Den foreløpige ordningen med å kunne gå opp til en ny, nasjonal deleksamen neste semester har fungert greit, siden den nasjonale deleksamenen i matematikk ikke har vært studiepoenggivende og heller ikke tellende i begynnelsen. Så fort eksamenen gir uttelling i studiepoeng og det stilles krav om progresjon i studiet, må en kontinuasjonseksamen komme mye tidligere. For at nasjonale deksamener skal oppfattes som rettferdig og for å unngå spekulasjoner om «taktisk stryk», fordi studenter frykter at vurderingen er strengere ved en nasjonal deleksamen, må kontinuasjonseksamener også avholdes nasjonalt.

Eksamensavviklingen bør effektiviseres og digitaliseres

NOKUTs ad hoc-organisering som bindeledd mellom institusjonene, sensorene og fagmiljøene er ikke ideell. Vi har ikke tilgang til FS, og vi må derfor kommunisere med det lokale eksamenskontoret ved hvert enkelt studiested. Dette skaper også mer arbeid for eksamenskontorene ved at det bryter med deres eksisterende rutiner, og at de må etablere nye, særskilte rutiner kun for nasjonale deksamener. Dersom det oppstår feil, kan det gå utover studentenes rettssikkerhet, og vi har hatt flere tilfeller hvor

studenter har klaget på eksamenskarakteren og bedt om ny sensur, uten at dette har blitt fanget opp av institusjonene i tide. Videre er løsningen med nasjonale eksamener på papir sårbar, tungvint og ressurskrevende og i utakt med det arbeidet som foregår i sektoren med digitalisering av eksamen (eCampus). NOKUT har deltatt på erfaringsseminarer i regi av Uninett, og vi har hatt et arbeidsmøte sammen med Uninett og representanter fra sektoren for å kartlegge hva som må være på plass for å kunne gjennomføre de nasjonale deleksamenene digitalt. NOKUT har imidlertid ikke hatt kapasitet til å fortsette arbeidet med digitale eksamener parallelt med pilotprosjektet, eller å inkludere det i pilotprosjektet. Dette arbeidet må imidlertid gjøres, og den digitale infrastrukturen bør være på plass og ferdig utprøvd før nasjonale deleksamener innføres i fullskala-format. Digitale eksamener vil også gi de som utvikler eksamenssettene langt flere muligheter.

Tilrettelegging for forskning

NOKUT har en databehandleravtale med Kunnskapsdepartementet som gir oss tillatelse til å innhente og bruke individdata i resultatanalysene, men vi har ikke tillatelse til å utlevere data til forskningsformål, og dataene må slettes idet pilotprosjektet avsluttes. For at fagmiljøene skal kunne sammenligne seg med hverandre og kunne bruke resultatene til så vel kvalitetsforbedring som til forskning, må fagmiljøene få tilgang til disse dataene. Det vil kreve samtykke fra studentene, noe vi ikke har innhentet til dette prosjektet. Dersom nasjonale deleksamener skal videreføres, bør man i forkant avklare hvordan man deler data, slik at de enkelte fagmiljøene og andre forskere kan benytte seg av materialet. Digitale eksamener vil for øvrig forenkle prosessen med å innhente samtykke fra studentene.

4 Konklusjon

Gjennomføringen av mulighetsstudiet og pilotprosjektet nasjonale deleksamener har vist at vi greier å oppnå formålene som er satt for det, men NOKUTs midlertidige løsning er sårbar, ressurskrevende og lite effektiv. Dersom ordningen skal videreføres og videreutvikles, bør det finnes løsninger på de utfordringene som er beskrevet i kapittel 3.