

NOKUTs oppsummeringer

Nasjonal deleksamen i matematikk for grunnskolelærerutdanningen

Delrapport 1

Mars 2016



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 i matematikk for grunnskolelærerutdanningen
Forfatter(e):	Stephan Hamberg Kjersti Tokstad (prosjektleder) Andreas Snildal
Dato:	09.03.2016
Rapportnummer:	2016 – 1
ISSN-nr	[ISSN-nr]

Sammendrag

1. desember 2015 arrangerte NOKUT i samarbeid med Nasjonalt råd for lærerutdanningen (NRLU) og 15 høyskoler og universiteter en nasjonal deleksamen i matematikk for studenter på grunnskolelærerutdanning 1–7 (GLU 1–7) og grunnskolelærerutdanning 5–10 (GLU 5–10). Totalt 992 studenter fullførte eksamen og fikk sensur. Eksamenen omfattet alle GLU-studenter som fulgte undervisning i høstsemesteret hvor eksamenstemaet undervisningskunnskap i brøk, desimaltall og prosentregning inngikk. GLU-studenter som har undervisning i dette temaet først i vårsemesteret, avlegger nasjonal deleksamen i vårsemesteret. Denne første delrapporten omhandler eksamenen som ble avholdt 1. desember 2015, og vil suppleres når resultater fra fremtidige deksamener i matematikk foreligger.

Det ble opprettet en prosjektgruppe som bestod av representanter fra NRLU, arbeidsutvalget for matematikk og nasjonalt nettverk for matematikk i lærerutdanningen. Denne gruppen (heretter omtalt som NRLUs prosjektgruppe eller bare prosjektgruppen) har, gjennom tett kontakt med fagmiljøene, hatt det overordnede ansvaret for utviklingen av en nasjonal deleksamen i matematikk i grunnskolelærerutdanningene. NOKUT har hatt administrativt og logistisk ansvar i forbindelse med eksamensgjennomføringen.

Nasjonal deleksamen i matematikk for grunnskolelærerutdanningene var i all hovedsak vellykket. Gjennomføringen av eksamenen og sensuren fungerte godt. Analysene viser at eksamenssettet og sensuren er reliable og valide, og resultatene gir oss en mulighet til å si noe om studentenes kunnskapsnivå.

Resultatene indikerer at kunnskapsnivået i brøk, desimaltall og prosentregning er relativt høyt. Hele 69,7 prosent av studentene fikk karakteren C eller bedre, og kun 10,4 prosent strøk. Det er noen forskjeller på institusjonsnivå, men alle institusjonene hadde et karaktersnitt på D eller bedre, og dersom man fjerner kandidater som strøk fra gjennomsnittsmålingen, ligger nesten alle institusjonene med ett snitt på C eller bedre.

Sammenligningen med resultatene fra de ordinære matematikkeksamenene kandidatene avla samme semester, viser også at det ikke er slik at man på lokalt plan gir enklere eksamener eller er «snillere» med sensuren. Resultatene på de ordinære eksamenene var noe svakere enn på den nasjonale deleksamenen, men her er det viktig å poengtere at de ordinære eksamenene dreier seg om hele pensumet, og at det derfor ikke er unaturlig at resultatene er noe svakere.

Innhold

1	Innledning	1
1.1	Roller og ansvarsfordeling.....	2
1.2	Forskriftsendring	2
1.3	Utfordringer i startfasen.....	2
1.4	Organisering av nasjonal deleksamen	3
1.4.1	Oppgaveutvikling	3
1.4.2	Eksamensdagen	4
1.4.3	Sensur.....	4
1.5	Utvalget	4
2	Resultater	6
2.1	Nasjonale tall	7
2.2	Nasjonal deleksamen sammenlignet med ordinær matematikkeksamen	10
3	Detaljert institusjonsoversikt	12
3.1	Høgskolen i Bergen	13
3.2	Høgskolen i Buskerud og Vestfold.....	14
3.3	Høgskolen i Nord-Trøndelag.....	15
3.4	Høgskolen i Oslo og Akershus	16
3.5	Høgskolen Stord/Haugesund	17
3.6	Høgskolen i Sør-Trøndelag	18
3.7	Høgskolen i Telemark	19
3.8	Høgskolen i Østfold.....	20
3.9	Høgskolen i Sogn og Fjordane	21
3.10	Høgskolen i Volda	22
3.11	NLA Høgskolen.....	23
3.12	Universitetet i Nordland	24
3.13	Universitetet i Stavanger	25
3.14	Universitetet i Tromsø – Norges arktiske universitet.....	26
4	Kvalitetssikring av oppgavesett og sensur	27
4.1	Sensurreliabilitet.....	27
4.2	Validitet og reliabilitet av eksamenssettet	30
4.2.1	Eksamenssettets validitet.....	30

4.2.2	Eksamenssettets reliabilitet	30
5	Konklusjon.....	31

1 Innledning

NOKUT fikk 8. september 2014 i oppdrag av Kunnskapsdepartementet å gjennomføre en mulighetsstudie og et pilotprosjekt med nasjonale deksamener i høyere utdanning. Oppdraget er formulert slik:

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 sammenlikne 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 må 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.

Tre profesjonsutdanninger ble valgt ut til å delta i prosjektet, som finner sted i 2015 og 2016. Disse er grunnskolelærerutdanningene (GLU 1–7 og GLU 5–10, sistnevnte for studenter som velger matematikk), bachelorgradsstudiet i sykepleie og bachelorgradsstudiet i regnskap og revisjon.

Denne rapporten handler om nasjonal deksamnen i matematikk i GLU 1–7 og i GLU 5–10 for studenter som har valgt matematikk. Eksamen ble avholdt 1. desember 2015. Formålet med rapporten er å evaluere prosessen og resultatene av nasjonal deksamnen i matematikk i grunnskolelærerutdanningene, beskrive de erfaringene NOKUT og de andre aktørene har dannet seg og å belyse fremtidige utfordringer med nasjonale deksamener.

Innvendinger mot prosjektet

NOKUT tok umiddelbart kontakt med de aktuelle nasjonale rådene/profesjonsrådene etter at vi hadde fått oppdraget. For å kunne lykkes med prosjektet, har det hele tiden vært en forutsetning at det ble forankret i fagmiljøene, men det er ikke til å komme bort fra at det har vært til dels store motforestillinger til prosjektet, både i startfasen og underveis. Disse har vært knyttet til prinsipielle, faglige og praktiske problemstillinger. De vanligste motforestillingene var blant annet at nasjonale deksamener:

- innebærer sterk statlig styring
- bidrar til mindre mangfold og diversitet og kan føre til ensretting av utdanningene
- bryter med institusjonenes autonomi og fagmiljøenes integritet
- mangler hjemmel i eksisterende lovverk
- griper inn i eksisterende studie- og undervisningsplaner ved den enkelte institusjon og skaper praktiske problemer for undervisningspersonale og kandidater
- forstyrrer prosesser og tiltak som allerede er satt i gang for å heve kvaliteten i lærerutdanningene generelt og i matematikk spesielt – tiltak som KD allerede har bedt

Nasjonalt råd for lærerutdanning (NRLU) om å sette i gang, samt tiltak som er forankret i UHR

En annen viktig innvending mot prosjektet var at det ikke lot seg gjøre å utvikle eksamener av høy kvalitet og som oppfyller krav til validitet og reliabilitet innenfor tidsrammene departementet har satt. Vi kommer tilbake til disse problemstillingene senere i rapporten.

1.1 Roller og ansvarsfordeling

Til tross for disse motforestillingene klarte NOKUT og Nasjonalt råd for lærerutdanning (NRLU) å opprettholde et konstruktivt samarbeid, og å komme frem til løsninger på flere utfordringer underveis. Det ble opprettet en prosjektgruppe som bestod av representanter fra NRLU, arbeidsutvalget for matematikk og nasjonalt nettverk for matematikk i lærerutdanningen. Denne gruppen (heretter omtalt som NRLUs prosjektgruppe eller bare prosjektgruppen) har, gjennom tett kontakt med fagmiljøene, hatt det overordnede ansvaret for utviklingen av en nasjonal deleksamen i matematikk i grunnskolelærerutdanningene. I tillegg har en bredt sammensatt eksamensgruppe utgått fra matematikknettverket hatt ansvaret for den faglige utformingen av eksamenssettet, og utarbeidet temaplan, eksempeloppgaver, eksamensoppgaver, og sensorveiledning. NOKUT har hatt administrativt og logistisk ansvar i forbindelse med eksamensgjennomføringen, som bestilling og fordeling av eksamenspapir, kandidatnumre til eksamen, fordeling av besvarelser til sensorkorpsset samt formidling av eksamensresultatene til institusjonene. NOKUT har også tatt ansvar for å avklare en del av problemstillingene som har dukket opp underveis med Kunnskapsdepartementet.

1.2 Forskriftsendring

Parallelt med at fagmiljøene innenfor de utvalgte profesjonsutdanningene startet arbeidsprosessen med å utvikle nasjonale deleksamener, utarbeidet Kunnskapsdepartementet et høringsutkast om endring av forskrift om rammeplaner i henholdsvis begge grunnskolelærerutdanningene, bachelorgradsstudiet i sykepleie og bachelorgradsstudiet i regnskap og revisjon. Endringene ble fastsatt 13. juli 2015, og med dette fikk departementet hjemmel til å pålegge institusjonene å avholde en nasjonal deleksamen. Departementet sendte ut brev 11. august om at gjennomføring av nasjonal deleksamen høsten 2015 var obligatorisk innenfor grunnskolelærerutdanningene og sykepleierutdanningen, og at det var en forutsetning at eksamenene ble gjennomført i tråd med det NOKUT og de nasjonale rådene hadde blitt enige om. For at studentene skulle yte sitt beste på denne eksamenen, bestemte departementet at karakteren enten skal føres på vitnemålet eller telle med i endelig karakter på vitnemålet.

1.3 Utfordringer i startfasen

Eksamensutviklingsarbeidet bød på en del praktiske utfordringer i begynnelsen. De ulike lærerutdanningsinstitusjonene har ulike profiler, og det finnes ingen felles læringsutbyttebeskrivelser i matematikk ut over de nasjonale retningslinjene for grunnskolelærerutdanningene. Institusjonene tilbyr matematikk i ulike semestre i utdanningen, slik at det kan variere fra første til tredje semester når studentene tar matematikk. I tillegg var de fleste undervisningsplanene for 2015 allerede lagt da oppdraget om nasjonale deleksamener kom. Disse faktorene gjorde det vanskelig å lage én, felles eksamen basert på felles læringsutbyttebeskrivelser som skulle inngå i grunnlaget for karakterfastsettelsen i faget. Det var en forutsetning at studentene må 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 eksamen våren 2016. Videre bestemte NOKUT og NRLU at den nasjonale deleksamenen skulle avholdes som en isolert deleksamen – altså komme i tillegg til studentenes ordinære eksamen i matematikk – siden fagmiljøene ikke fant det forsvarlig å utsette studentene for en tellende eksamen i et pilotprosjekt med så stramme tidsrammer at det kunne gå på bekostning av eksamenens kvalitet. Når det gjaldt manglende felles læringsutbyttebeskrivelser og eksamenens innhold, valgte eksamensgruppen et tema som var sentralt for alle som skal undervise i femte til syvende klassetrinn, nemlig undervisningskunnskap i brøk og desimaltall. Således fant man et tema som kunne omfatte studentene i begge grunnskolelærerutdanningene, og hvor studentene kunne avlegge eksamen i det semesteret de tar et emne hvor undervisning i brøk og desimaltall inngår, uavhengig av hvor i studieløpet emnet gis ved de forskjellige institusjonene.

En isolert deleksamen, som ikke inngår i den ordinære vurderingen av faget og som derfor ikke kan gi studiepoeng, er ikke tellende. Karakteren skal likevel fremgå av vitnemålet. Det innebærer at en kandidat kan få utstedt vitnemål, så lenge vedkommende har møtt opp og avlagt eksamen, men for å få utstedt vitnemål er det ikke noe krav om at vedkommende må bestå den nasjonale deleksamenen. Det er med andre ord mulig å stryke på den nasjonale deleksamenen og likevel få utstedt vitnemål. Dette er en uheldig konsekvens av at vi valgte å gjennomføre eksamenen som en isolert deleksamen og må tas høyde for når man ser på resultatene. Det kan være en potensiell fare for at noen studenter ikke har ytt sitt beste, siden hverken NOKUT eller institusjonene har noen sanksjonsmidler overfor dem. Analysene vi har gjort viser imidlertid ingen tegn på at studenter aktivt har boikottet eksamen, eller ikke har ytt sitt beste, men vi kan ikke se bort ifra at dette kan skje i fremtiden. Vi anbefaler derfor at man unngår «isolerte deleksamener» i fremtiden.

1.4 Organisering av nasjonal deleksamen

1.4.1 Oppgaveutvikling

Eksamensgruppen har hatt ansvar for utviklingen av en nasjonal deleksamen i matematikk i GLU. Gruppen utviklet først en felles temaplan for begge grunnskolelærerutdanningene, som beskrev det valgte temaet mer konkret enn det som fremgår av læringsutbyttebeskrivelsene, og denne skulle hjelpe institusjonene frem til en felles forståelse av omfang og nivå. I tillegg var temaplanen forsynt med eksempeloppgaver til hjelp for undervisere. Videre har eksamensgruppen utviklet selve eksamensoppgavene og laget en sensorveiledning.

På grunn av den stramme tidsrammen hadde ikke NOKUT mulighet til å utarbeide et rammeverk for den nasjonale deleksamenen eller forsikre oss om at eksamenssettet oppfyller testmetodiske krav i forkant av eksamensdatoen. Eksamensgruppen ble etablert i februar 2015, og den nasjonale deleksamenen skulle i henhold til oppdraget avholdes innen utgangen av høstsemesteret samme år. Til sammenligning bruker Utdanningsdirektoratet to til tre år på å utvikle en nasjonal prøve i grunnskolen. De nasjonale prøvene piloteres på forhånd, og de oppfyller krav til reliabilitet og validitet. Innenfor norsk høyere utdanning er det foreløpig ikke utbredt å lage eksamensoppgaver ut fra testmetodiske krav. For å sikre at de nasjonale deleksamenene, så godt det lar seg gjøre ut fra forutsetningene, svarer på oppdraget, arrangerte NOKUT et testmetodisk seminar for eksamensgruppen. Eksamensoppgavene

burde ha vært testet ut på forhånd, men tidsrammen tillot ikke det. NOKUT har derimot gjort analyser av eksamenssettet og sensuren i etterkant for å vurdere kvaliteten på dette.

1.4.2 Eksamensdagen

Nasjonal deleksamen i matematikk i GLU ble avholdt som en firetimers skriftlig skoleeksamen uten bruk av hjelpemidler 1. desember 2015. Det var 15 lærerutdanningsinstitusjoner som har lagt matematikk til høstsemestret som deltok i prosjektet høsten 2015, og dette gjelder studenter både i første og tredje semester og GLU 1–7 og 5–10. Institusjonene var selv ansvarlige for selve eksamensgjennomføringen, som å holde lokaler, skaffe eksamensvakter m.m. Kommunikasjonen med institusjonene har for det meste forløpt uproblematisk, selv om enkelte institusjoner har lagt avvikende definisjoner til grunn for eksamenen, for eksempel gjennom å omtale den overfor undervisere og studenter som «nasjonal prøve».

Totalt gjennomførte 992 av 1089 oppmeldte kandidater eksamen, som tilsvarer en gjennomføringsgrad på 91 prosent. Institusjonene sendte besvarelsene til ViaScan AS, som skannet alle besvarelsene og lastet dem opp i en skyløsning, slik at de ble tilgjengelige for sensorene.

1.4.3 Sensur

En nasjonal eksamen forutsetter nasjonal sensur. For å sikre likebehandling og full anonymitet ble det satt sammen et nasjonalt sensorkorps på 48 personer fra 19 utdanningsinstitusjoner. Antallet sensorer var basert på en forventning om ca. 1100 innleverte besvarelser. Sensorkorpset ble satt sammen av eksamensgruppen, og formelt oppnevnt av alle institusjonene som gjennomførte den nasjonale deleksamenen.

Hver besvarelse ble vurdert av to sensorer. For å få til en karakterkalibrering og en institusjonell spredning av besvarelsene delte NOKUT inn sensorene i til sammen 92 sensorpar, hvor hver sensor, med unntak av seks, sensurerte sammen med fire andre sensorer. De seks øvrige sensorene sensurerte sammen med tre andre. Hvert sensorpar sensurerte 11 besvarelser, som var tilfeldig fordelt, og hver enkelt sensor vurderte i gjennomsnitt 40 tilfeldig fordelte besvarelser fra alle de 15 institusjonene.

1.5 Utvalget

Tabell 1.1 viser hvilke institusjoner som deltok og hvilke programmer studentene gikk på.

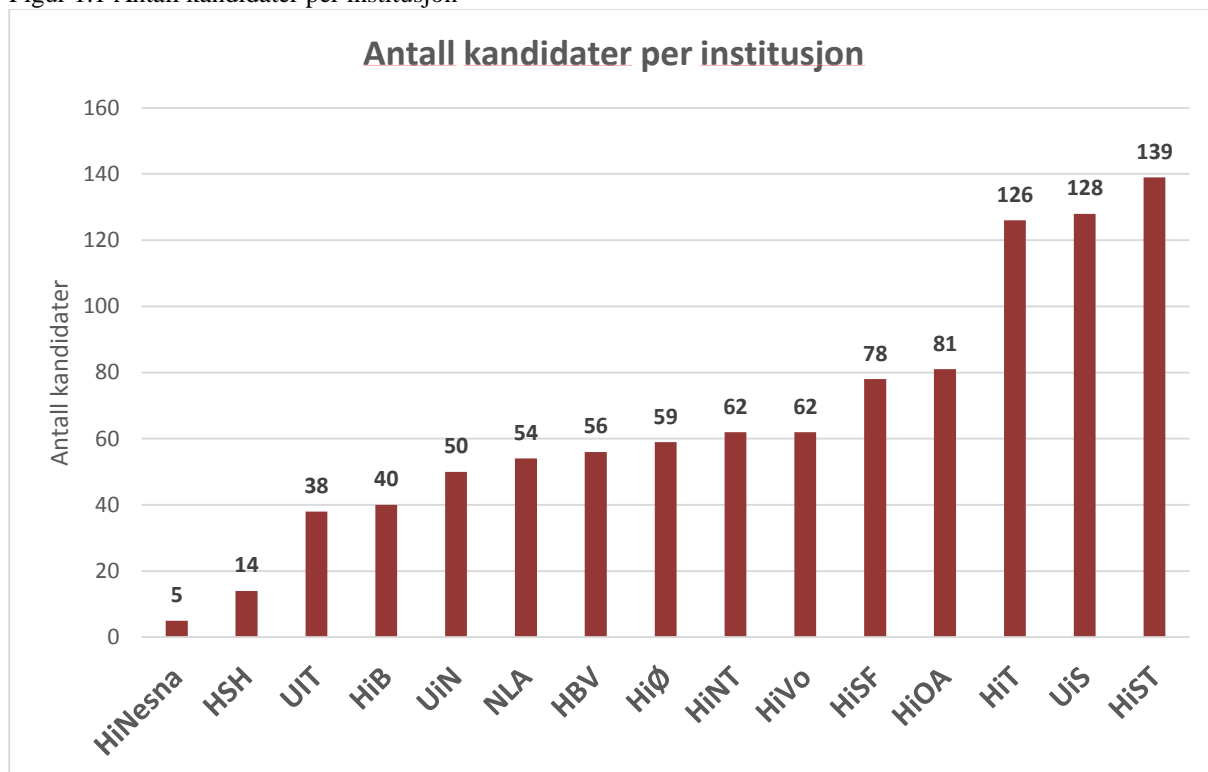
Tabell 1.1 Institusjoner og programmer¹

Institusjon	Studie-program 1	Studie-program 2	Semester
Høgskolen i Bergen	GLU 5–10		1. semester
Høgskolen i Buskerud og Vestfold	GLU 5–10		1. semester
Høgskolen i Nesna	GLU 5–10		1. semester
Høgskolen i Nord-Trøndelag	GLU 1–7		1. semester
Høgskolen i Oslo og Akershus	GLU 5–10		1. semester
Høgskolen i Sør-Trøndelag	GLU 5–10		1. semester
Høgskolen i Telemark	GLU 5–10	GLU 1–7	1. og 3. semester
Høgskolen i Østfold	GLU 5–10		1. og 3. semester
Høgskolen Stord/Haugesund	GLU 5–10		3. semester
Høgskolen i Sogn og Fjordande	GLU 5–10	GLU 1–7	1. og 3. semester
Høgskolen i Volda	GLU 5–10	GLU 1–7	1. semester
NLA Høgskolen	GLU 5–10	GLU 1–7	1. semester
Universitetet i Nordland	GLU 5–10	GLU 1–7	1. semester
Universitetet i Stavanger	GLU 5–10	GLU 1–7	1. semester
Universitetet i Tromsø – Norges arktiske universitet	GLU 5–10	GLU 1–7	1. og 3. semester

I tillegg til at det er studenter ved forskjellige institusjoner som kommer fra forskjellige programmer og fra forskjellige semestre, er det også stor forskjell på hvor mange studenter som tok eksamen ved de forskjellige studiestedene. Figur 1.1 viser antallet studenter som tok den nasjonale deleksamenen ved hver institusjon.

¹ I analysene under har vi ikke skilt mellom studenter som går på GLU 5–10 eller GLU 1–7. Vi har heller ikke skilt mellom 1. og 3. semester. Grunnen til dette er at NOKUT på nåværende tidspunkt ikke har denne informasjonen på individnivå.

Figur 1.1 Antall kandidater per institusjon



Som vi ser av figuren, over var det kun fem kandidater ved Høgskolen i Nesna som tok eksamen. For å bevare anonymiteten til disse kandidatene, samt å være sikre på relativt robuste tall, vil vi ikke vise noen enkeltanalyse fra Høgskolen i Nesna. Som tabellen over (tabell 1.1) viser, har studenter fra GLU 1–7 og 5–10 tatt den nasjonale deleksamenen. Ved sju institusjoner har studenter fra begge programmene deltatt. I analysene under har vi ikke tatt hensyn til dette. Det vil si at resultatene for alle studentene, ved for eksempel Høgskolen i Telemark, er slått sammen. Vi har mottatt tilbakemeldinger fra institusjonene som påpeker at dette kan være uheldig. Det er sannsynlig at studenter ved ett av programmene har gjort det bedre eller dårligere enn studenter ved det andre programmet. Dessverre har ikke NOKUT informasjon til å skille ut studenter basert på studieprogram, kun institusjon. I den avsluttende rapporten som kommer i august 2016 vil NOKUT ta hensyn til disse forholdene.

2 Resultater

I dette kapittelet presenterer vi resultatene fra den nasjonale deleksamenen i matematikk som ble gjennomført 1. desember 2015. Vi sammenligner også resultatene på den nasjonale deleksamenen med resultatene fra den ordinære matematikkeksamenen som studentene hadde senere i desember. Først presenterer vi tall for hele populasjonen. Deretter presenterer vi resultatene for hver institusjon. Vi avslutter med en analyse av sensuren.

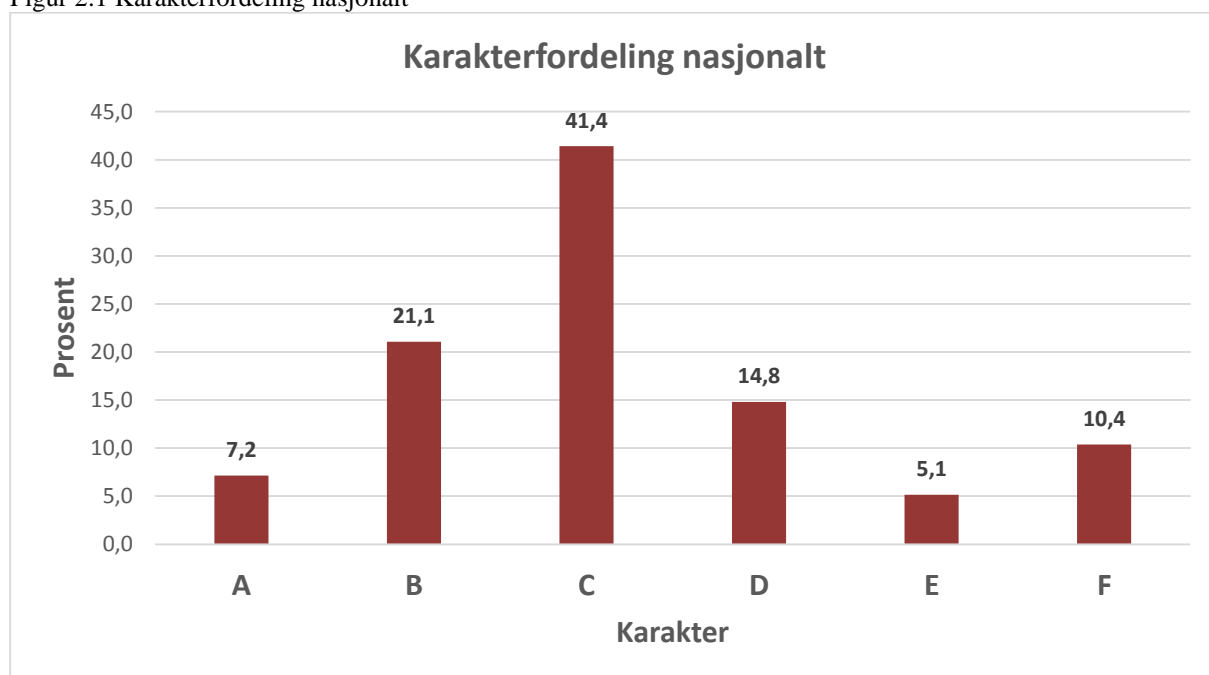
I figurene under diskuterer vi karakterer både som tall og som bokstaver. Dvs. at vi har i noen tilfeller gjort bokstavkarakterene om til tall for bedre å illustrere den variasjonen som eksisterer i datamaterialet (bokstavkarakteren A=5, B=4, C=3, D=2, E=1, F=0). Vi inkluderer også de studentene

som stryker i alle gjennomsnittsoversiktene (i motsetning til institusjonene selv som rapporterer gjennomsnittskarakterer uten å inkludere studenter som stryker).

2.1 Nasjonale tall

Som nevnt tidligere, var det totalt 992 kandidater som tok den nasjonale deleksamenen. Dersom vi ser på resultatene for alle kandidatene, kan vi si at det ser ut som at norske lærerstudenter gjør det relativt godt i matematikkfaget (brøk, desimaltall og prosent). Gjennomsnittskaracteren var 2,8 (C), og hele 69,7 prosent av studentene fikk karakteren C eller bedre. Figur 2.1 viser hele svarfordelingen nasjonalt.

Figur 2.1 Karakterfordeling nasjonalt



Som vi ser fra figur 2.1, er karakterfordelingen tilnærmet normalfordelt på nasjonalt nivå, og en stor andel av kandidatene gjør det meget godt. Uten sammenlignbare tall er det vanskelig å si om en strykprosent på 10,4 prosent er uforholdsmessig høy, men som vi viser under, er 10,4 prosent en lavere strykprosent enn det man hadde under den ordinære lokale eksamenen i matematikk (15,3 prosent).

Det er viktig å påpeke at eksamensresultatene ikke ble normalfordelt. Med andre ord, det var ikke slik at karakterene ble satt etter sensur. For den nasjonale deleksamenen fikk hver sensor en sensorveiledning og tilgang til et poengskjema. Hver sensor rettet sine besvarelser og oppga poengsummen på hver oppgave i skjemaet. Sensoren diskuterte så med sin medsensor om sensuren på hver oppgave. Begge sensorene sendte så sine skjema til NOKUT. Eksamensgruppen og sensorene ble enige om hvilke poenggrenser som skulle gjelde for bokstavkarakterene den 4. desember 2015, kun få dager etter at de hadde fått tilgang til kandidatenes besvarelser. Hver kandidat kunne oppnå maksimalt 40 poeng, og eksamensgruppen fulgte anbefalingene til Norsk matematisk råd satte følgende poenggrenser for hver enkelt karakter:

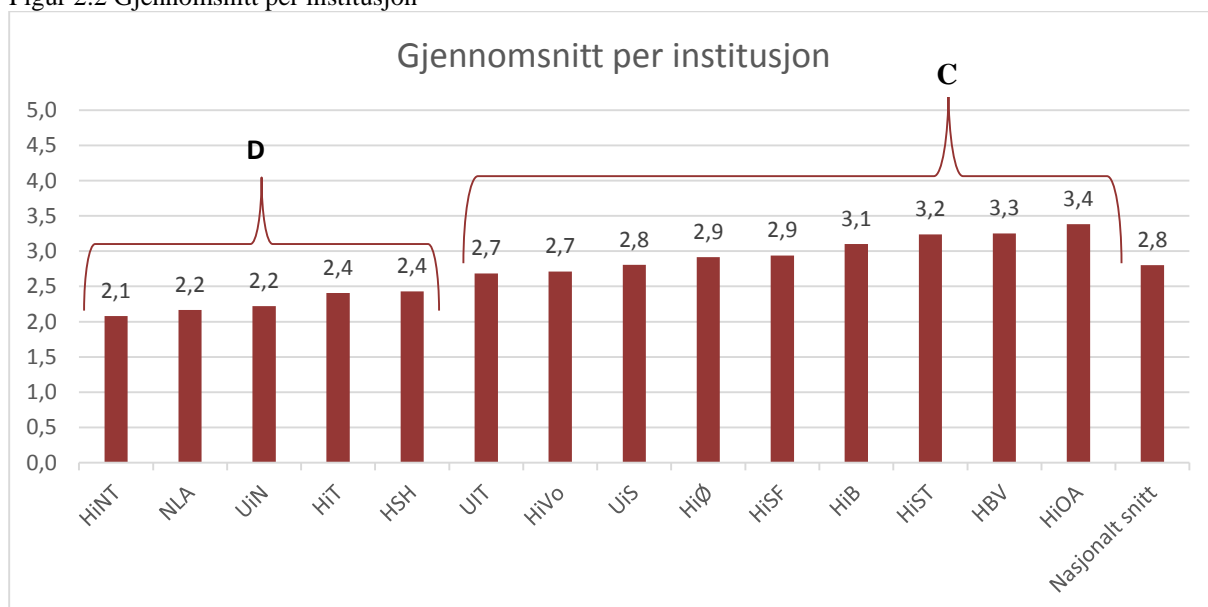
Tabell 2.1 poenggrenser for enkeltkarakterer

Poeng	Karakter
0-15	F
16-17	E
18-22	D
23-30	C
31-35	B
36-40	A

Karakterene er med andre ord absolutte, og karakterfordelingen har ikke blitt endret etter at sensuren er ferdigstilt.

Dersom vi beveger oss ned til institusjonsnivået (figur 2.2), ser vi at det er ganske stor spredning i gjennomsnittskarakter når vi bruker desimaltall (1,3 poeng). Dersom vi ser på bokstavkarakterer, ser vi at fem av institusjonene hadde D i gjennomsnitt, mens de resterende ni hadde C i gjennomsnitt. Vi ser også at ingen institusjoner hadde et gjennomsnitt på mer enn 0,7 fra det nasjonale gjennomsnittet.

Figur 2.2 Gjennomsnitt per institusjon

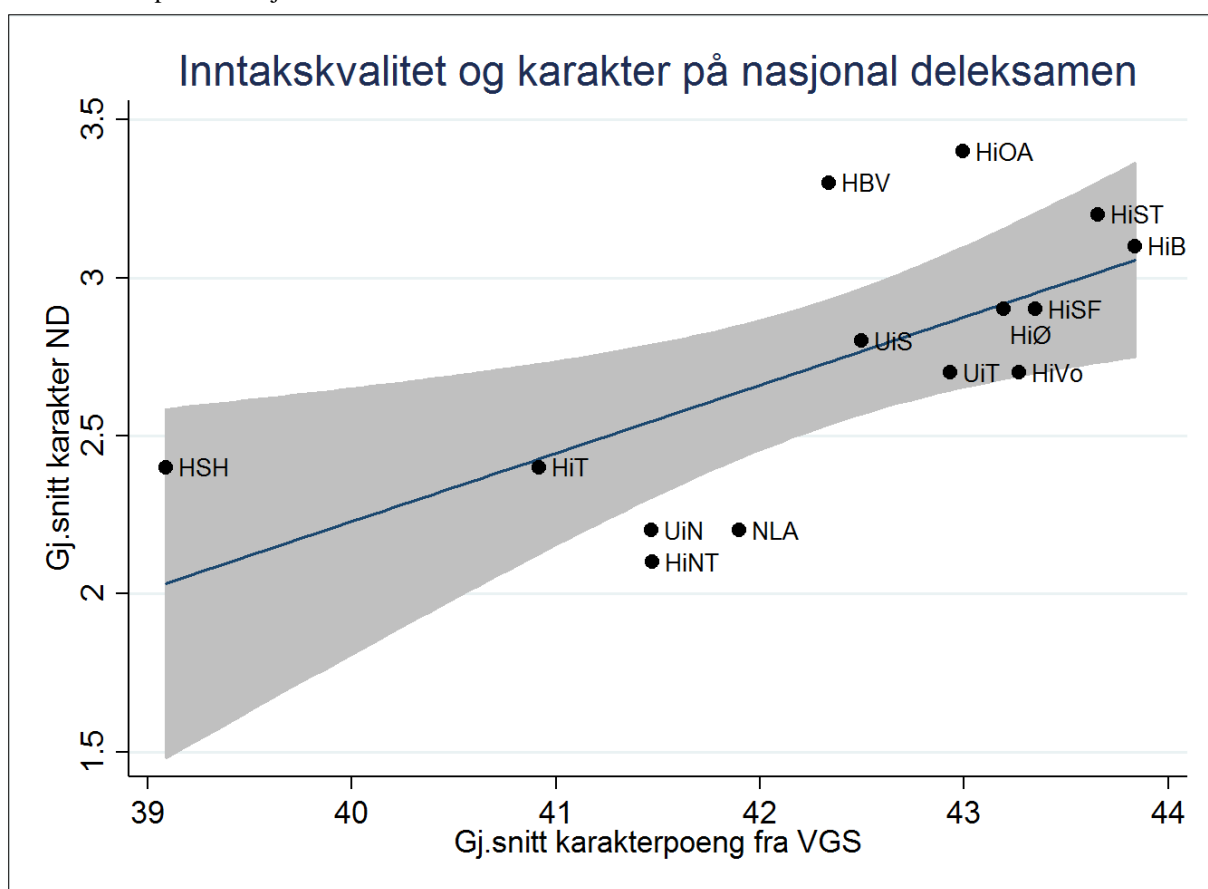


Selv om resultatene er relativt gode ved de fleste institusjonene, er det interessant å se at alle de store profesjonshøgskolene (HiOA, HBV, HiST og HiB) har bedre resultater på den nasjonale deleksamenen enn de nye universitetene (UiS og UiN) og Universitetet i Tromsø. Vi har ikke noe datamaterialet som kan belyse årsaken til dette, men det kan være en indikasjon på at det å få universitetsstatus ikke automatisk bedrer undervisningskvaliteten eller inntakskvaliteten ved en institusjon.

Hva kan forklare forskjellen på gjennomsnittskarakterene mellom institusjonene? Som tabell 1.1 viser, er de fleste studentene som har tatt den nasjonale deleksamenen i sitt første semester i utdanningen. Det er derfor rimelig å anta at mye av variasjonen i karakterer på eksamen kan forklares av studentenes forkunnskaper i matematikk. For å teste en slik hypotese burde vi hatt informasjon om studentenes karakterpoeng fra videregående utdanning på individnivå. NOKUT jobber med dette, men har i dag ikke tilgang til disse dataene på individnivå. For å få en indikasjon om sammenhengen

mellom inntakskvalitet og karakterer på den nasjonale deleksamenen bruker vi derfor data fra NOKUTs studiebarometer. Basert på Studiebarometeret har vi regnet ut gjennomsnittlig karakterpoeng på alle programmene som deltok på den nasjonale deleksamenen. Vi har så sett på sammenhengen mellom dette og karaktersnittet på programnivå. Det er to viktige forbehold å ta i denne analysen: 1) Gjennomsnittlig karakterpoeng er basert på studenter som ikke tok den nasjonale deleksamenen, men går på samme studieprogram, og 2) karakterpoeng er kun basert på de studentene som svarte på den nasjonale studentundersøkelsen Studiebarometeret. Vi legger til grunn at gjennomsnittlig karakterpoeng ikke har endret seg nevneverdig fra 2014 til 2015, og at tallene er representative for hele studieprogrammet.² Sammenhengen vises i figur 2.3.

Figur 2.3 Sammenheng mellom karakterpoeng fra videregående utdanning og karakter på den nasjonale deleksamenen per institusjon.



Figuren viser gjennomsnittskarakteren på den nasjonale deleksamen og gjennomsnitt karakterpoeng fra videregående skole per institusjon (studieprogram). Den diagonale linjen er regresjonslinjen som viser den statistiske sammenhengen mellom variablene. Det grå feltet er et 95% konfidensintervall. Forutsatt at den statistiske modellen er god kan man si at UiN, NLA og HiNT sine resultater på den nasjonale deleksamen er dårligere enn man burde forventet basert på studentenes inntakskvalitet. På den andre siden ser vi at HBV og HiOA gjorde det bedre enn forventet basert på inntakskvaliteten. Det vi ser er en klar sammenheng mellom inntakskvalitet og karakter på den nasjonale deleksamenen.³ Dette er ikke overraskende når man vet at de fleste studentene har tatt den nasjonale deleksamenen i

² Nye tall fra DBH viser at tallene vi har brukt i denne delen av analysen er representative.

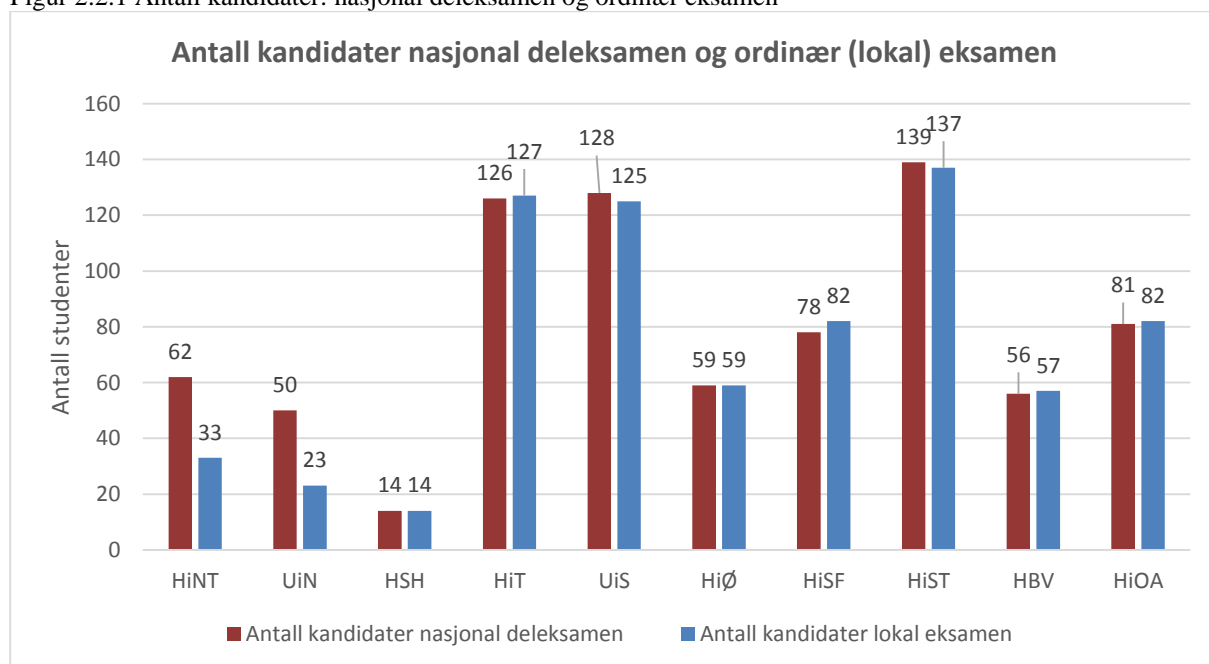
³ Korrelasjonen mellom gjennomsnittet på den nasjonale deleksamen og karakterpoeng fra VGS er 0,65 og resultatet er statistisksignifikant.

sitt første semester. Det er naturlig å anta at studentenes forkunnskaper i matematikk påvirker deres resultat på en eksamen som fokuserer på brøk, desimaltall og prosentregning. Samtidig er det viktig å påpeke at vi her ikke har kontrollert for variabler som også kan påvirke sammenhengen, som for eksempel: alder, kjønn, om studenten har vært i praksis, størrelsen på studieprogrammet, etc. Det figuren viser er en mulig sammenheng som bør undersøkes nærmere. Dersom sammenhengen holder på individnivå, kan man si at studentenes forkunnskaper matematikk i stor grad forklarer resultatet på den nasjonale deleksamen.

2.2 Nasjonal deleksamen sammenlignet med ordinær matematikkeksamen

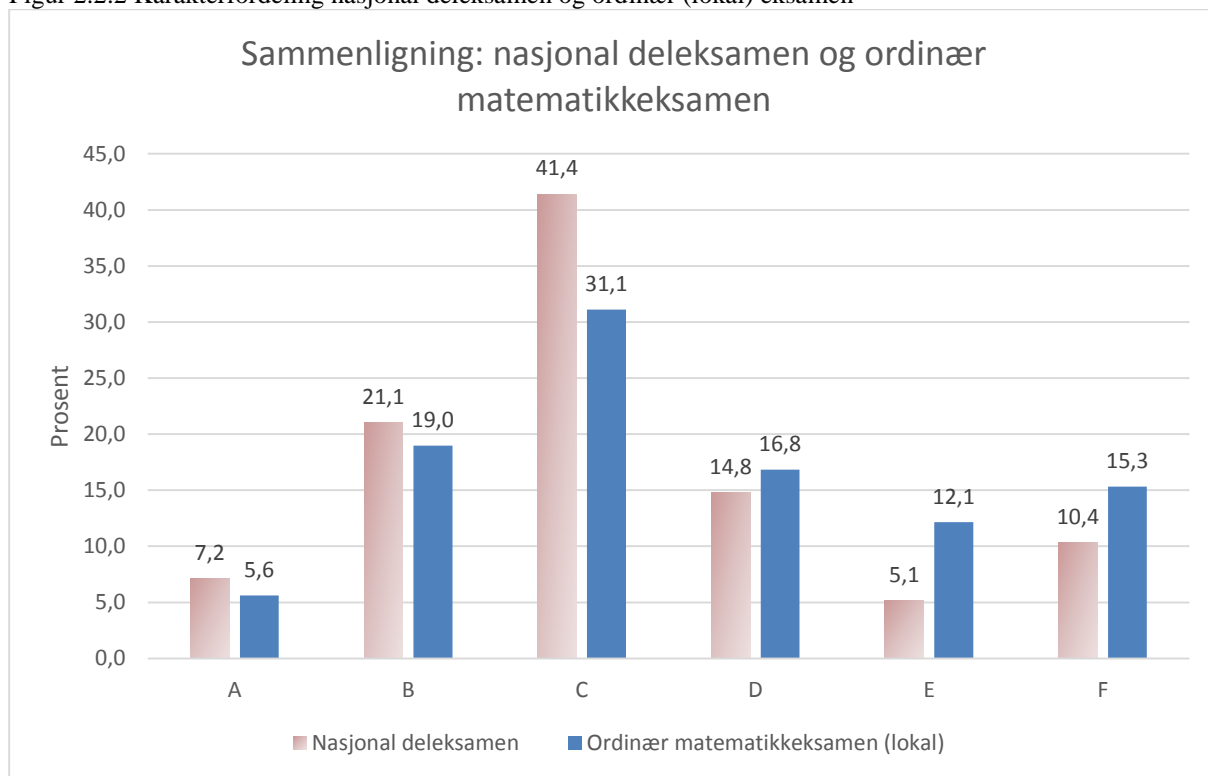
De fleste institusjonene som deltok i den nasjonale deleksamenen, avholdt også en ordinær eksamen i matematikk i desember 2015 som de enkelte institusjonenes deleksamenkandidater også avla. Denne ordinære eksamenen blir skrevet, sensurert og arrangert lokalt på de enkelte institusjonene. Institusjonene som avholdt ordinær skriftlig eksamen i matematikk, sendte NOKUT resultatene fra denne eksamenen. I de påfølgende figurene sammenligner vi resultatene fra den nasjonale deleksamenen med resultatene fra den ordinære eksamenen. Den første figuren viser hvilke institusjoner NOKUT har fått resultater fra og hvor mange studenter ved hver institusjon som har tatt de to eksamenene.

Figur 2.2.1 Antall kandidater: nasjonal deleksamen og ordinær eksamen



Som vi ser, er det noen institusjoner hvor flere studenter har tatt den nasjonale eksamenen enn den lokale eksamenen. Årsaken til dette er som regel at et eller flere studieprogram ikke har hatt ordinær lokal eksamen ennå, eller at karakterene i emnet er basert på både mappevurdering og eksamen. I denne rapporten ønsker vi kun å sammenligne de skriftlige eksamenene med hverandre. Som over hadde det vært bedre å gjøre denne sammenligningen på individnivå, men NOKUT har ennå ikke fått tilgang til disse dataene. Vi er likevel ganske sikre på at det er de samme kandidatene som i all hovedsak har tatt begge eksamenene. Å sammenligne på programnivå er derfor mindre problematisk. I figur 2.2.2 viser vi karakterfordelingen på de to eksamenene på nasjonalt nivå.

Figur 2.2.2 Karakterfordeling nasjonal deleksamen og ordinær (lokal) eksamen

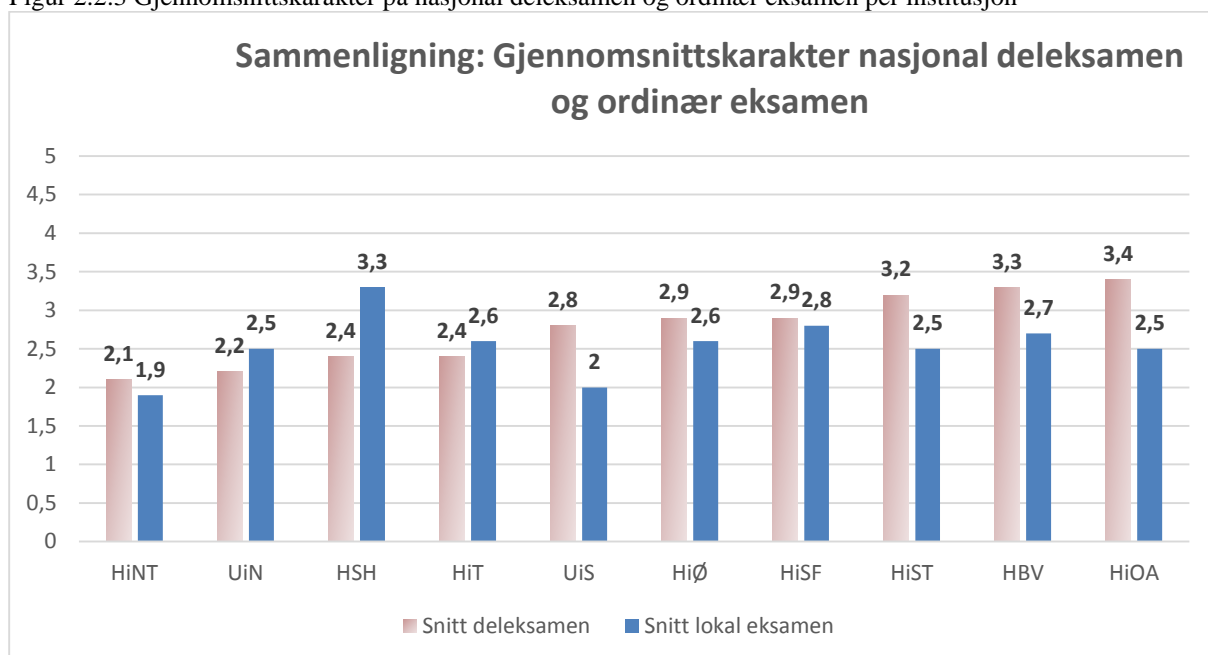


Som vi ser, er resultatene ved den ordinære lokale eksamenen noe dårligere enn på den nasjonale deleksamenen. Gjennomsnittskarakteren på den ordinære lokale eksamenen var 2,4 (D), mens det på den nasjonale deleksamenen var 2,8 (C).⁴ Hvis vi ser på institusjonsnivå (figur 2.2.3), ser vi at kun tre institusjoner har en høyere gjennomsnittskarakter på den ordinære eksamen enn på den nasjonale deleksamenen.

Det er interessant at resultatet på den nasjonale deleksamen er noe bedre enn på de ordinære eksamenene. En mulig årsak er at den nasjonale deleksamen kun dreide seg om en del av et større pensum, mens de ordinære eksamenene dreide seg om hele pensumet i matematikk. En annen mulig årsak er at institusjonene har brukt ekstra ressurser og tid på å undervise spesielt mye i brøk, prosent og desimaltall på grunn av den nasjonale deleksamen. Med andre ord, det finnes en mulighet for at de gode resultatene kan være et resultat av såkalt «teach to the test». NOKUT har ingen mulighet til å sjekke dette, men på grunn av meget korte tidsfrister og en helt ny eksamen har vi liten grunn til å tro at dette er årsaken til de gode resultatene.

⁴ Dersom vi skulle ekskludert studentene som fikk F i gjennomsnittsberegningen, ville snittet vært C på begge eksamenene.

Figur 2.2.3 Gjennomsnittskaracter på nasjonal deleksamen og ordinær eksamen per institusjon



NOKUT har ikke kvalitetssikret de lokale eksamenene eller sensuren, men resultatene fra de to eksamenene viser at det ikke er slik at institusjonene tilpasser nivået på lokale eksamener eller sensur for å bedre studentenes karakterer. Med andre ord, det er ikke slik at institusjonene er «snille» med karakterene når de settes lokalt.

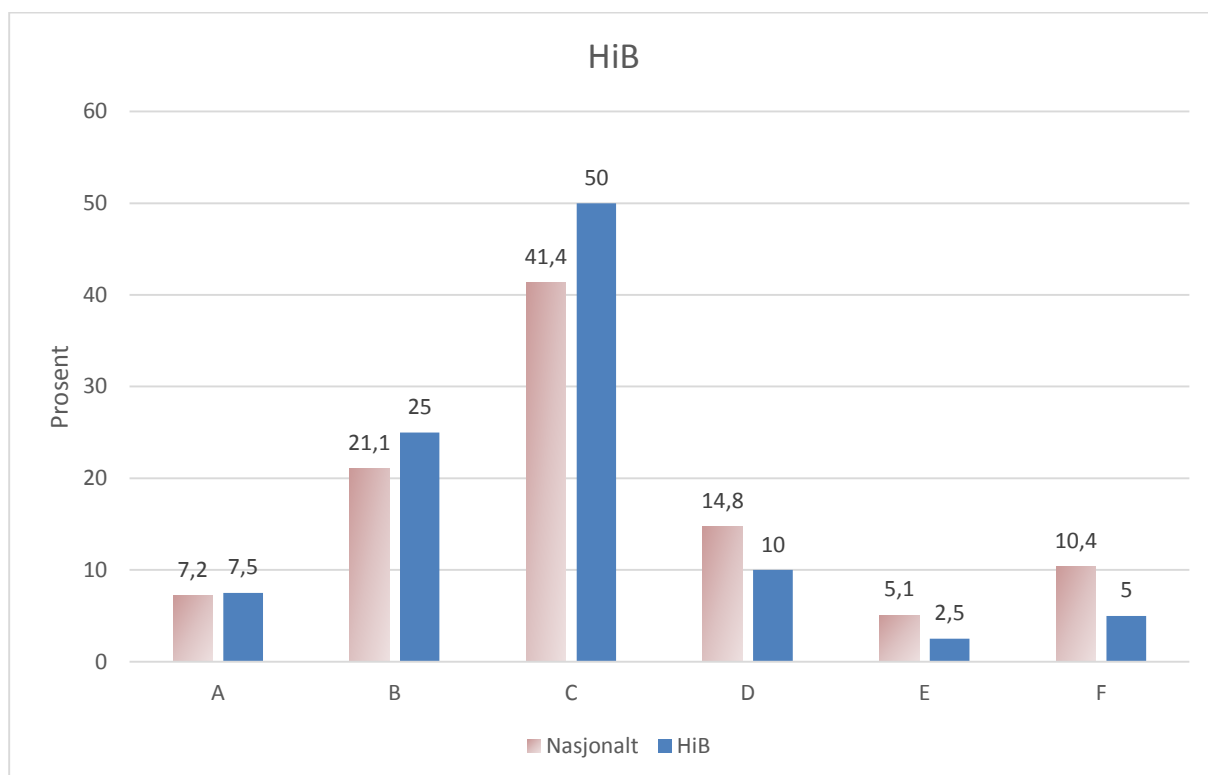
3 Detaljert institusjonsoversikt

I dette kapitlet går vi gjennom resultatene til hver av institusjonene som deltok på den nasjonale deleksamenen i matematikk. Vi viser karakterfordelinger på den nasjonale deleksamenen, samt den ordinære eksamenen der vi har resultater. NOKUT har gitt den enkelte institusjon mulighet til å kommentere resultatene vi presenterer for å informere om mulige årsaker til resultatene eller om forbehold som bør tas. Vi inkluderer disse der vi mener det er hensiktsmessig.

3.1 Høgskolen i Bergen

Gjennomsnittskarakteren ved HiB var 3,1 (C), noe som også var bedre enn det nasjonale gjennomsnittet. Ved HiB var det 40 studenter som tok den nasjonale deleksamenen. Disse studentene kommer fra tre forskjellige studieprogrammer, og sju av studentene var i sitt 7. semester i høst. De 33 andre studentene var i sitt første semester, men skal ikke ha ordinær matematikkeksamen før våren 2016. På bakgrunn av dette har vi ikke noen resultater fra ordinær eksamen å sammenligne med for HiB.

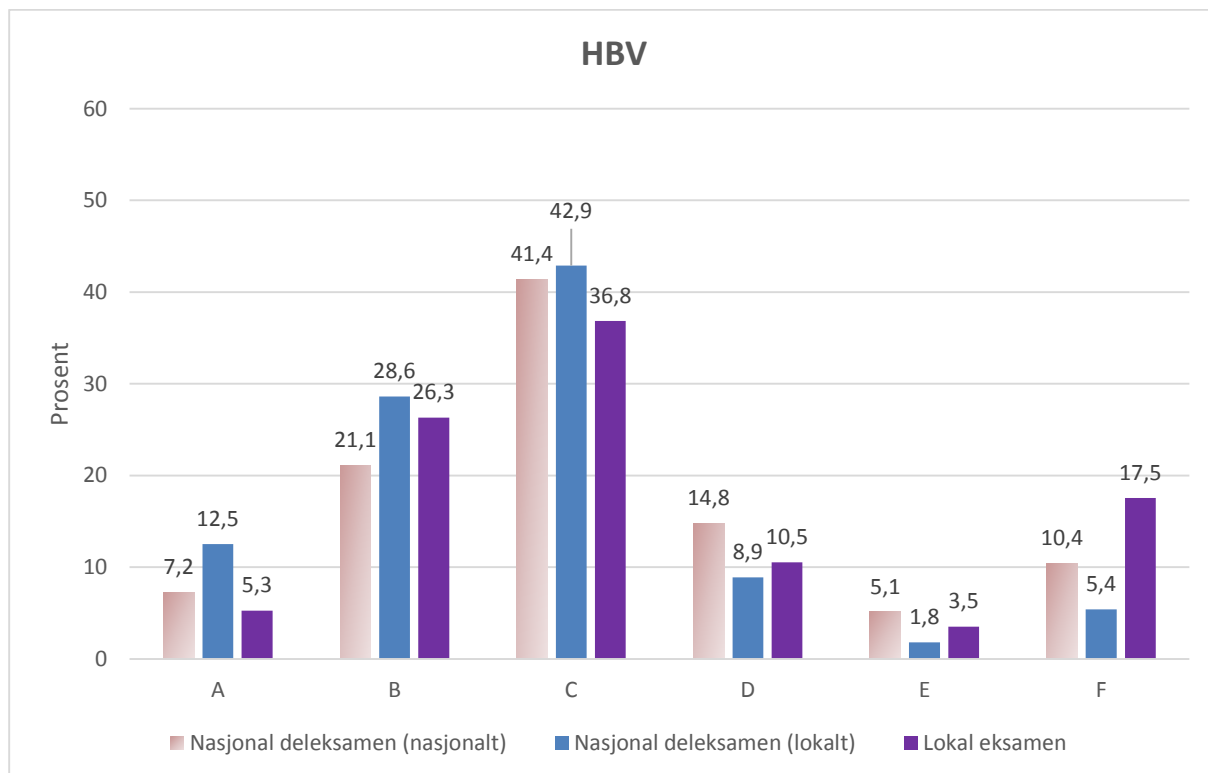
Karakterfordelingen på den nasjonale deleksamenen viser at HiBs fordeling er rimelig lik den nasjonale, men de har en litt større andel C og ca. 5 prosentpoeng mindre andel D og F.



3.2 Høgskolen i Buskerud og Vestfold

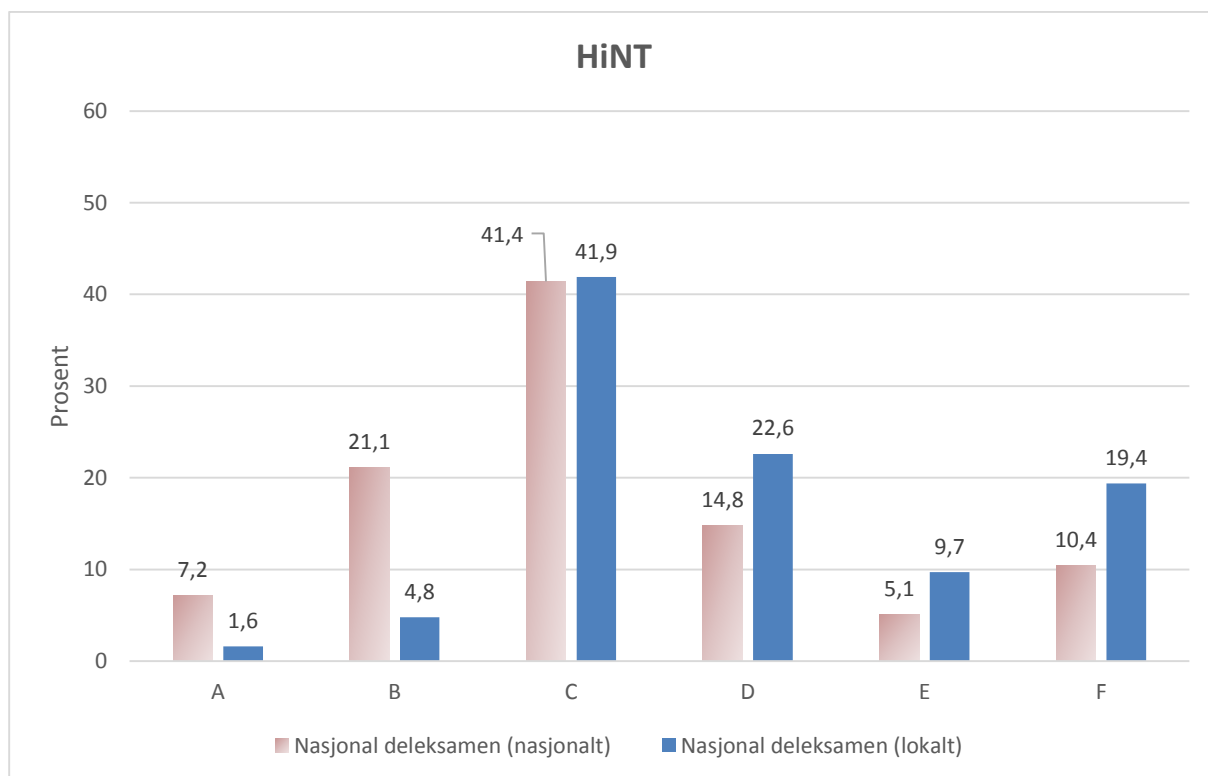
HBV gjorde det også meget godt på den nasjonale deleksamenen, og gjennomsnittet her var 3,3, altså en halv karakter bedre enn det nasjonale gjennomsnittet. På den ordinære eksamenen var gjennomsnittet også en C, men 0,6 karakterpoeng svakere (2,7). Karakterfordelingen viser at det på den ordinære eksamenen var en betydelig nedgang i andel A og B og en relativt stor økning i andel strykkarakter (ca. 12 prosentpoeng).

Sammenlignet med den nasjonale karakterfordelingen på den nasjonale deleksamenen hadde HBV en større andel A og B, samt fem prosentpoeng færre F.



3.3 Høgskolen i Nord-Trøndelag

62 studenter fra to forskjellige studieprogrammer tok den nasjonale deleksamenen i matematikk, men kun studenter ved det ene programmet hadde ordinær eksamen i desember 2015. Vi sammenligner derfor ikke resultatene fra den nasjonale eksamenen og ordinær eksamen. Gjennomsnittskarakteren ved HiNT var 2,1 (D), den laveste av de institusjonene som deltok, og en av fem institusjoner med karakteren D som gjennomsnitt. Karakterfordelingen viser at en veldig liten andel fikk A eller B (6,4 prosent), mens strykprosenten var dobbelt så stor ved HiNT som nasjonalt. Det er verdt å merke seg at HiNT nå har fusjonert med Universitetet i Nordland, som også hadde relativt svake resultater.



Høgskolen i Nord-Trøndelag er den eneste av institusjonene hvor kun studenter fra GLU 1–7 har tatt den nasjonale deleksamen. Fra HiNT rapporterer de at det sannsynligvis er signifikante forskjeller på resultatene til studenter fra GLU 1–7 og GLU 5–10. Årsaken til dette er at studenter på GLU 5–10 selv har valgt matematikk som fagretning og at disse studentene gjennomgående har:

- Klart større bevissthet om at de skal undervise i matematikk
- En slags forståelse for at temaene på prøven er viktige
- En bedre bakgrunn i matematikk fra videregående skole
- Perspektiv på at de skal studere faget over lenger tid slik at faget dermed er viktigere for dem
- Større sannsynlighet for å prestere godt på en eksamen av denne typen

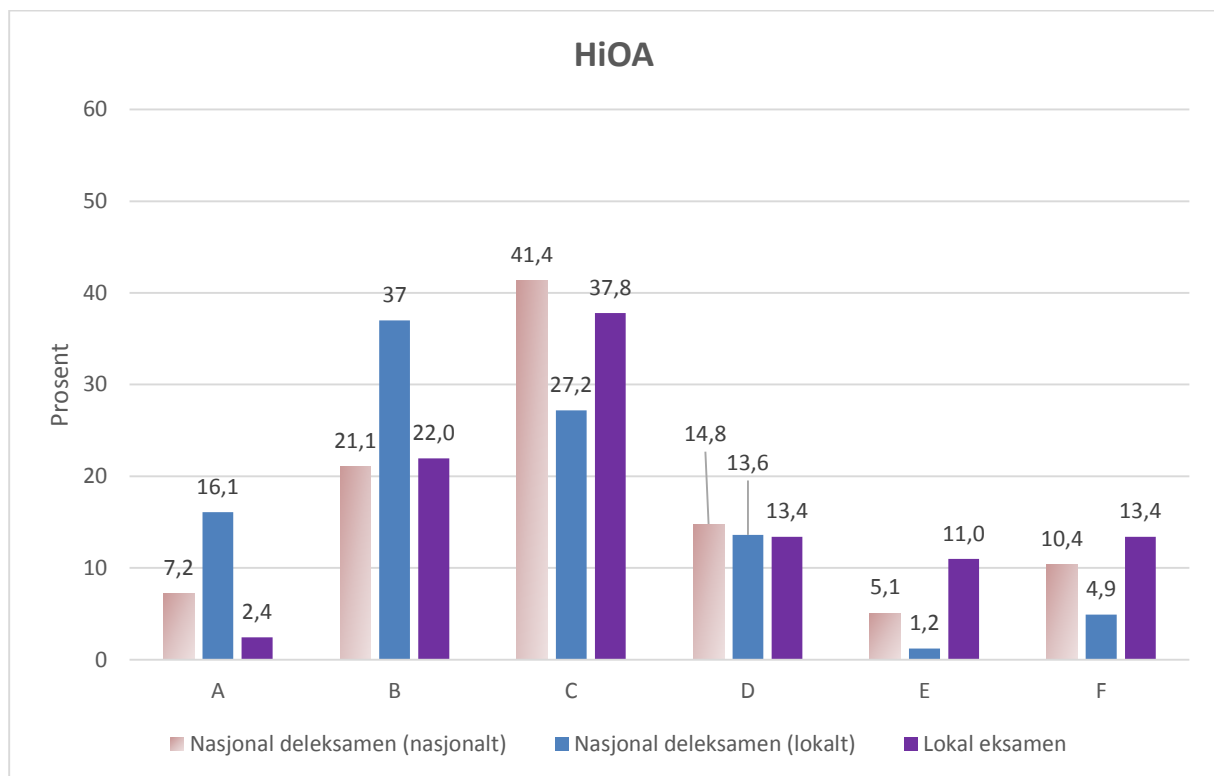
Siden HiNT kun har studenter fra GLU 1–7 kan dette derfor forklare hvorfor resultatene er svake sammenlignet med de andre institusjonene.

NOKUT er enige i at det er uheldig at vi på nåværende tidspunkt ikke kan skille resultatene per studieprogram, og vil gjøre dette i vår endelige rapport senere i 2016.

3.4 Høgskolen i Oslo og Akershus

Studentene ved HiOA var de som fikk høyest gjennomsnittskarakter på den nasjonale deleksamenen med et snitt på 3,4. Studentene hadde med andre ord nesten B i snittkarakter. På den lokale ordinære eksamenen var resultatene svakere, og gjennomsnittskarakteren var 2,5. Forskjellen i snittkarakter på de to eksamenene er allikevel mindre enn én bokstavkarakter, og snittet på begge to var C. Hvis vi ser på karakterfordelingen ser vi at det er en stor reduksjon i andelen studenter som fikk A og B, og en relativt stor økning i andelen studenter som fikk C, E og F.

Sammenlignet med den nasjonale karakterfordelingen på den nasjonale deleksamenen hadde HiOA en mye større andel A og B, en mye mindre andel C og kun halvparten så mange stryk.

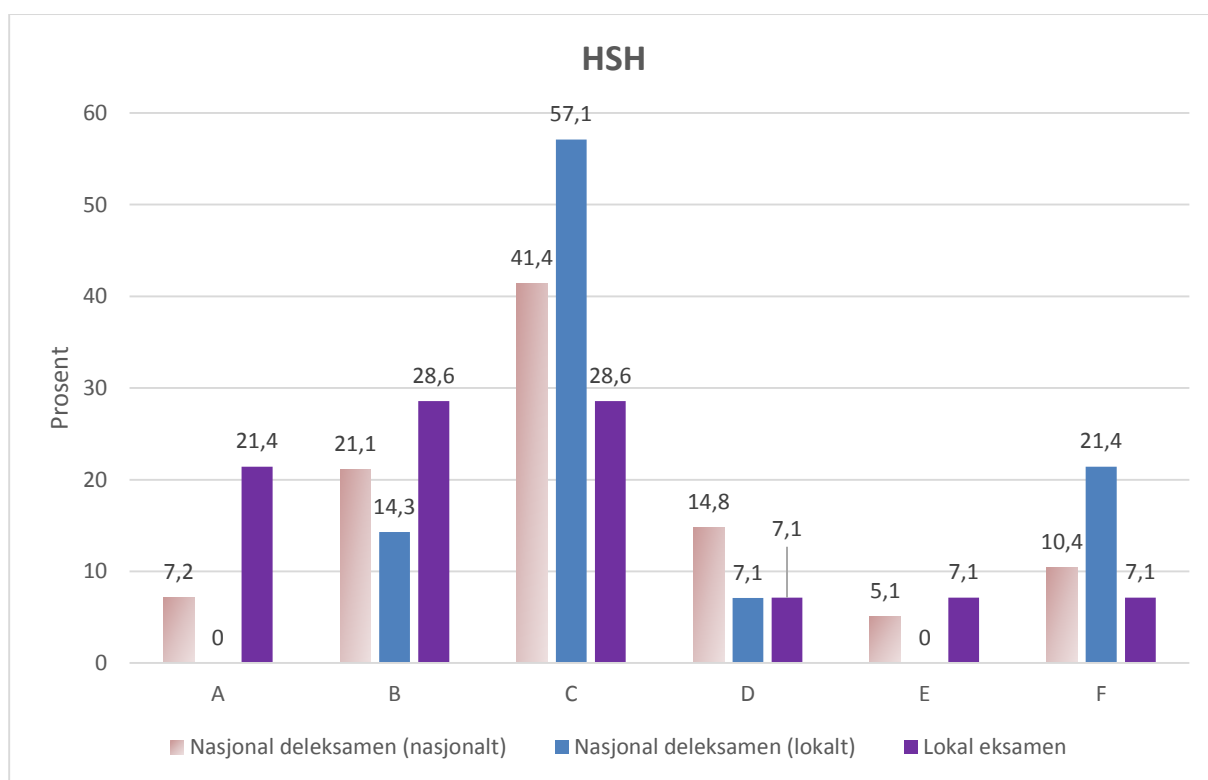


HiOA rapporterer at deres studenter hadde vært i 2 ukers praksis på ungdomstrinnet i forkant av dette delemnet. De var orientert om emneplan for nasjonal deleksamen og om eksempeloppgavene. Studentene fikk velge minst 4 av disse eksempeloppgavene som en del av et individuelt skriftlig arbeidskrav. Tilbakemelding på dette ble gitt i forkant av nasjonal deleksamen. Undervisningen i brøk, prosent og desimaltall ble avsluttet i uke 45.

Den ordinære eksamen i uke 50 var en skriftlig eksamen (tall, tallære, algebra og funksjoner). De siste tre ukene før nasjonal deleksamen ble studentene våre undervist blant annet i algebra og funksjoner i tillegg til en tverrfaglig skriveuke.

3.5 Høgskolen Stord/Haugesund

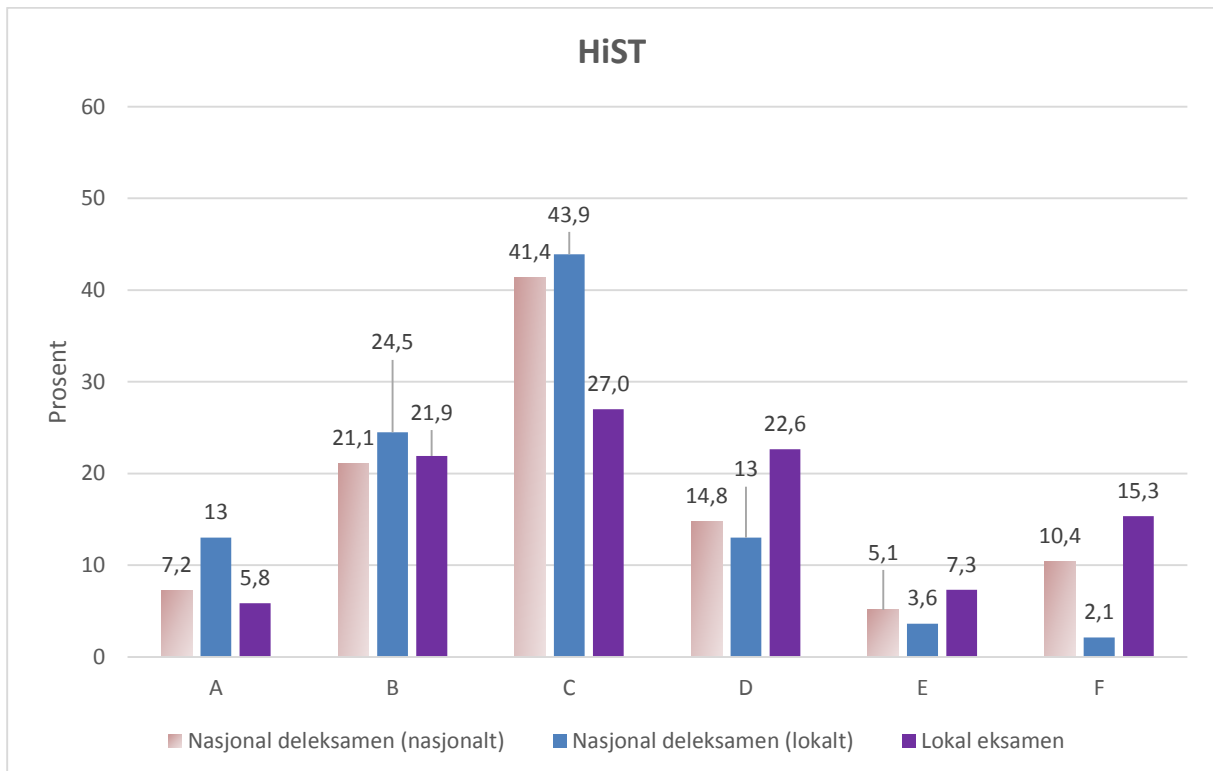
Det var kun 14 studenter ved HSH som tok den nasjonale deleksamenen, og gjennomsnittskaracteren var 2,4 (D). Det var også 14 studenter som tok den ordinære eksamenen i matematikk i desember 2015, og her var gjennomsnittet 3,3 (C). Med kun 14 studenter mangler tallene robusthet, men det er allikevel påfallende at man som eneste institusjon hadde en snittkaracter på ordinær eksamen som var en hel bokstavkaracter høyere enn på den nasjonale deleksamenen. Ser vi på karakterfordelingen, ser vi at HSH på den ordinære eksamenen hadde 50 prosent A eller B, mens på den nasjonale deleksamen fikk ingen studenter A, og kun 14 prosent fikk B. Det er også påfallende at strykprosenten var 14 prosentpoeng lavere på ordinær eksamen enn på den nasjonale deleksamenen. Selv med kun 14 studenter bør fagmiljøet vurdere om den ordinære eksamenen var for enkel, eller om sensuren var for «snill».



3.6 Høgskolen i Sør-Trøndelag

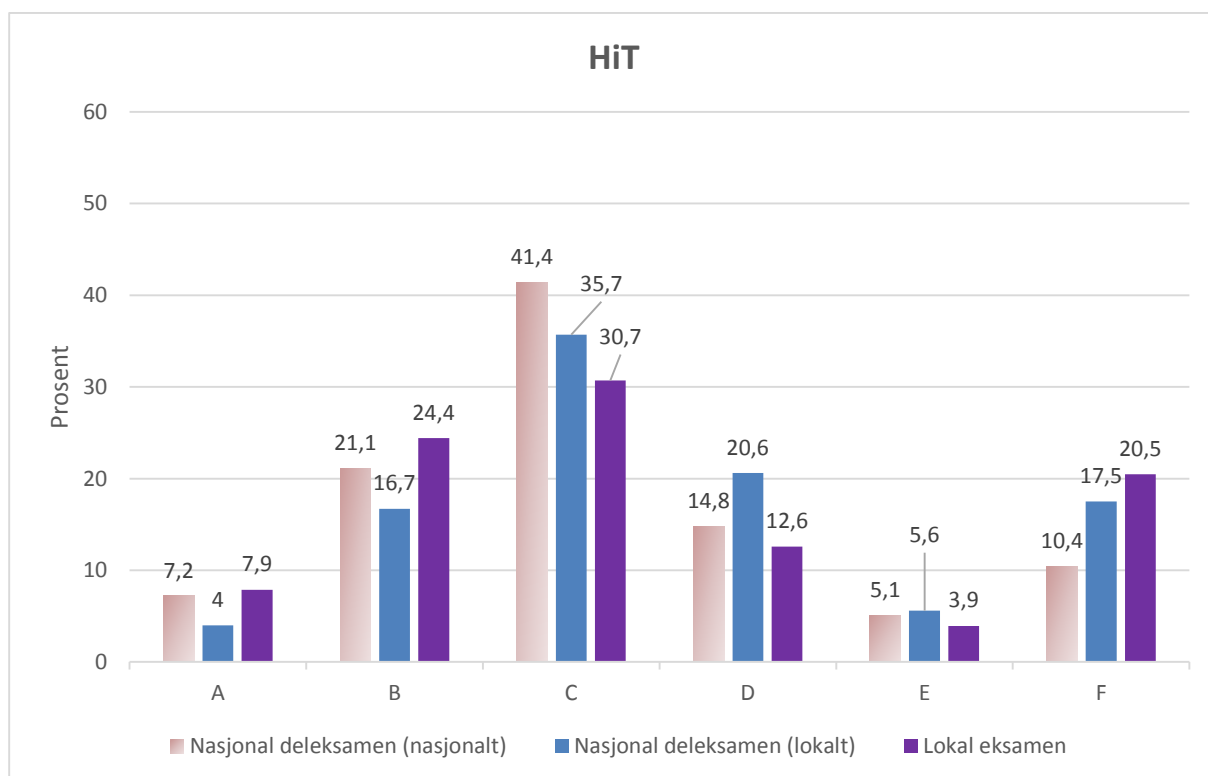
HiST var den institusjonen med flest kandidater på den nasjonale deleksamenen, og de gjorde det meget bra med et gjennomsnitt på 3,2. På den ordinære eksamenen var snittet mer enn en halv karakter lavere (2,5), men fortsatt innenfor en bokstavkarakter. Den største forskjellen ved HiST er den store nedgangen i C (16,9 prosentpoeng) og den store oppgangen i D (ca. 10 prosentpoeng) og F (ca. 13 prosentpoeng).

Sammenlignet med den nasjonale karakterfordelingen på den nasjonale deleksamenen var det spesielt andelen F som skiller stort. HiST hadde kun 2,1 prosent stryk mot 10,4 prosent stryk på nasjonalt nivå.



3.7 Høgskolen i Telemark

Ved HiT var det hele 126 studenter som tok den nasjonale deleksamenen, og 127 studenter som tok den ordinære eksamenen i matematikk. På den nasjonale deleksamenen var snittet 2,4 (D), mens på den ordinære eksamenen var snittet 2,6 (C). HiT gjorde det altså litt dårligere enn det nasjonale snittet på den nasjonale deleksamenen. Ved HiT er det hele fire forskjellige programmer som deltok på den nasjonale deleksamenen, to campus-programmer og to nettbaserte utdanninger. Det er mulig at det er forskjeller mellom disse programmene, men per dags dato har ikke NOKUT informasjon om dette. Karakterfordelingen viser en ganske grei normalkurve, men strykeprosenten ligger noe høyt på både den ordinære eksamenen og på den nasjonale deleksamenen.



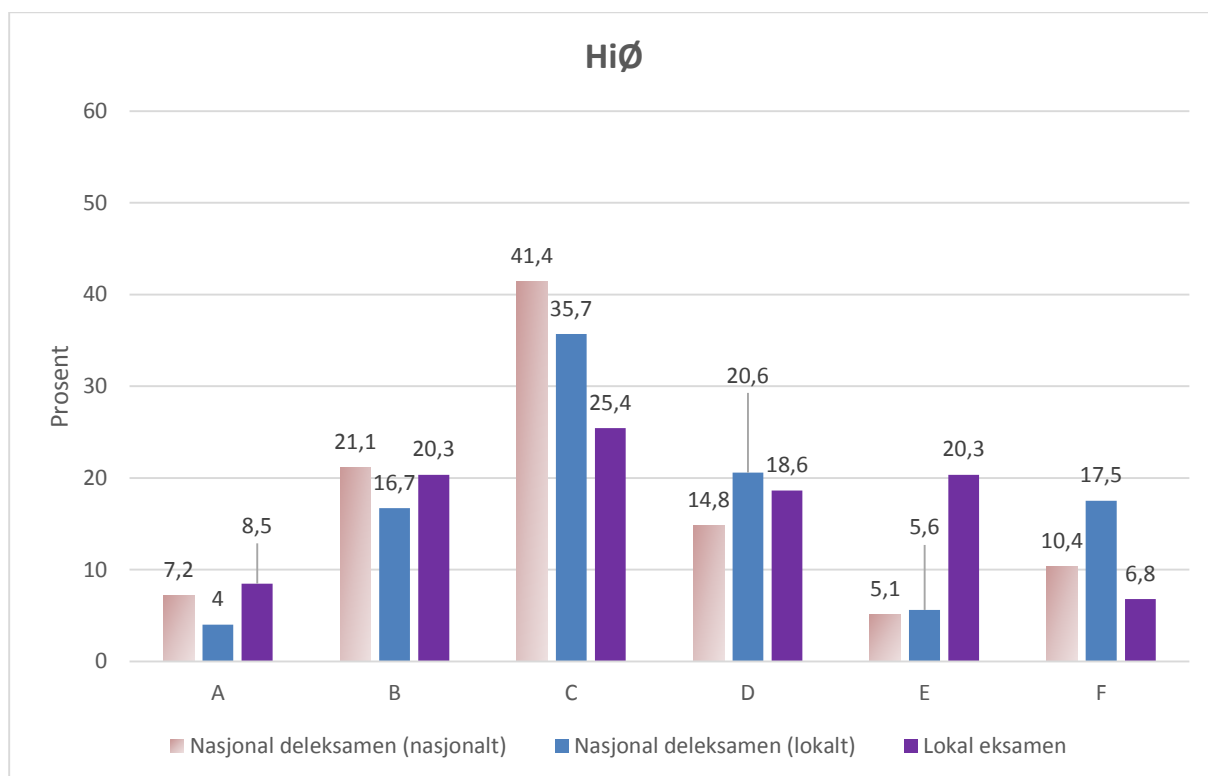
Høgskolen i Telemark er en av institusjonene hvor studenter fra både GLU 1–7 og GLU 5–10 har tatt eksamen. HiT rapporterer at resultatene sannsynligvis forskjellige på de to studieprogrammene. For eksempel er det slik at studentene i ett av programmene tok den nasjonale deleksamen tidligere i sitt studieløp enn andre studenter. Det er også slik at disse studentene ikke hadde hatt fokus på fagdidaktikk eller praksis.

NOKUT er enig at dette er faktorer som sannsynligvis har påvirket resultatene til HiT og vi vil forsøke å analysere resultater per studieprogram i vår hovedrapport senere i 2016.

3.8 Høgskolen i Østfold

Også Høgskolen i Østfold gjorde det bedre enn det nasjonale gjennomsnittet på den nasjonale deleksamenen. Gjennomsnittet for HiØ var 2,9 (C). På den ordinære matematikkeksamenen var gjennomsnittet 2,6, noe som er noe svakere, men fortsatt innenfor samme bokstavkarakter.

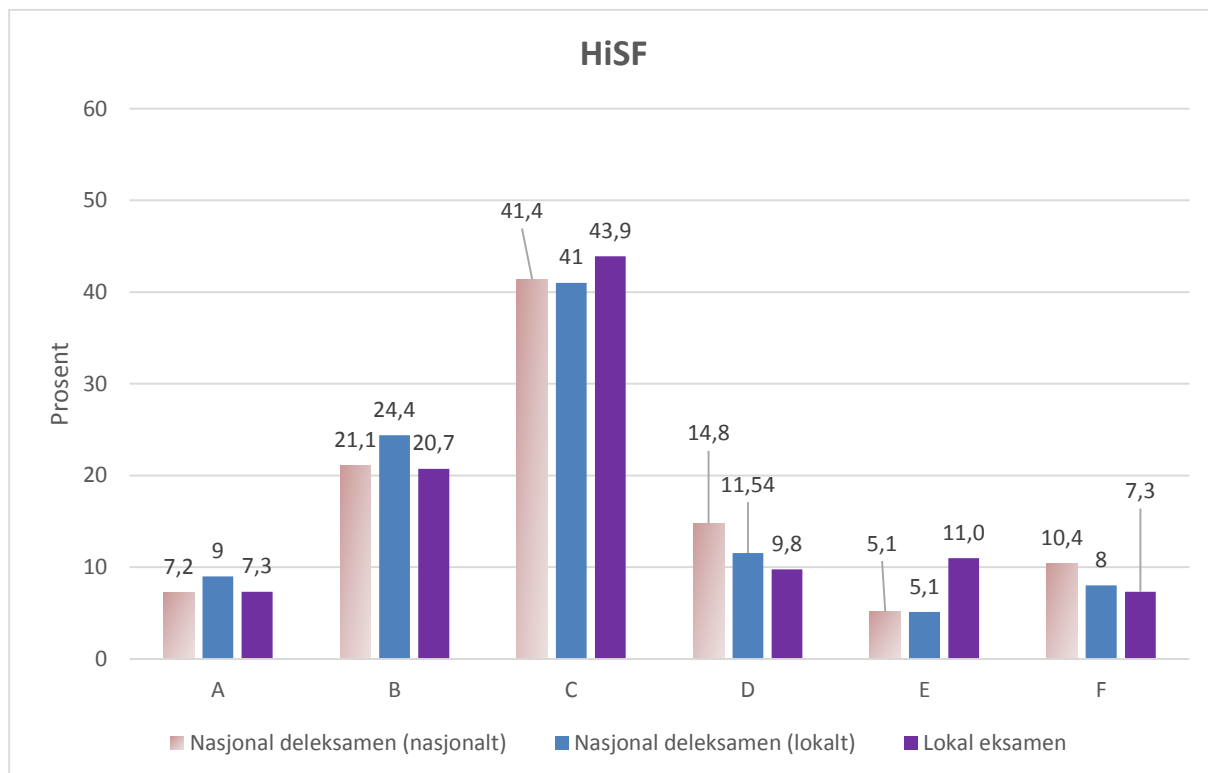
Karakterfordelingen indikerer noen interessante forskjeller: På den nasjonale deleksamen fikk kun 5,6 prosent av studentene E, og hele 17,5 prosent strøk. Disse tallene er nesten motsatt på den ordinære eksamenen, hvor 20,3 prosent fikk E, og kun 6,8 prosent strøk. NOKUT vet ikke årsaken til dette, men det er rimelig å anta at man vurderte skillet mellom E og F veldig ulikt på de to eksamenene. Vi anbefaler HiØ å se nærmere på hva som kan forklare denne forskjellen.



3.9 Høgskulen i Sogn og Fjordane

HiSF er den av de mindre profesjonshøgskolene som gjorde det best på den nasjonale deleksamenen i matematikk, med et snitt på 2,9 (C). Selv om høgskolen er blant de mindre profesjonshøgskolene, var det nesten like mange studenter ved HiSF som deltok som det var ved HiOA (78 mot 81 studenter).⁵ Det var 82 studenter som tok den ordinære matematikkeksamenen ved HiSF i desember 2015. På den ordinære eksamenen var gjennomsnittet 2,8, altså nesten identisk med snittet på den nasjonale deleksamenen.

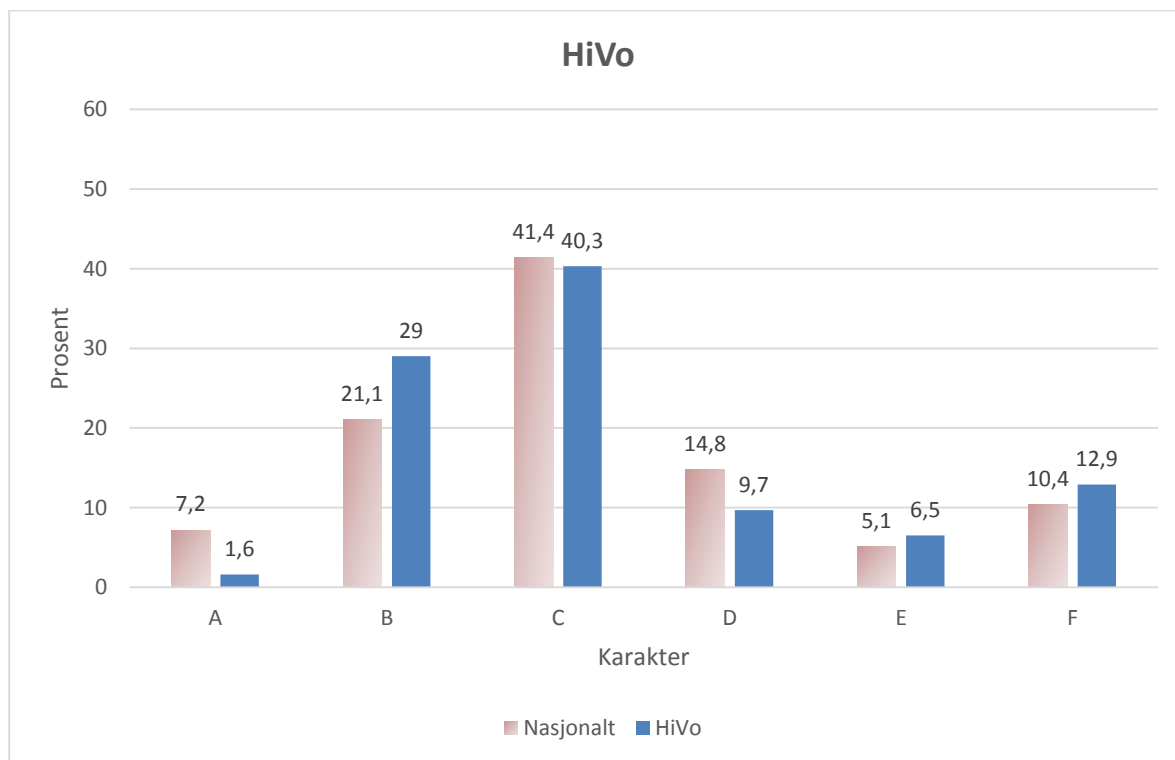
Ser vi på karakterfordelingene hos HiSF, ser vi at de i stor grad følger den nasjonale fordelingen.



⁵ De fleste lærerstudentene ved HiOA skal ta den nasjonale deleksamenen i mai 2016.

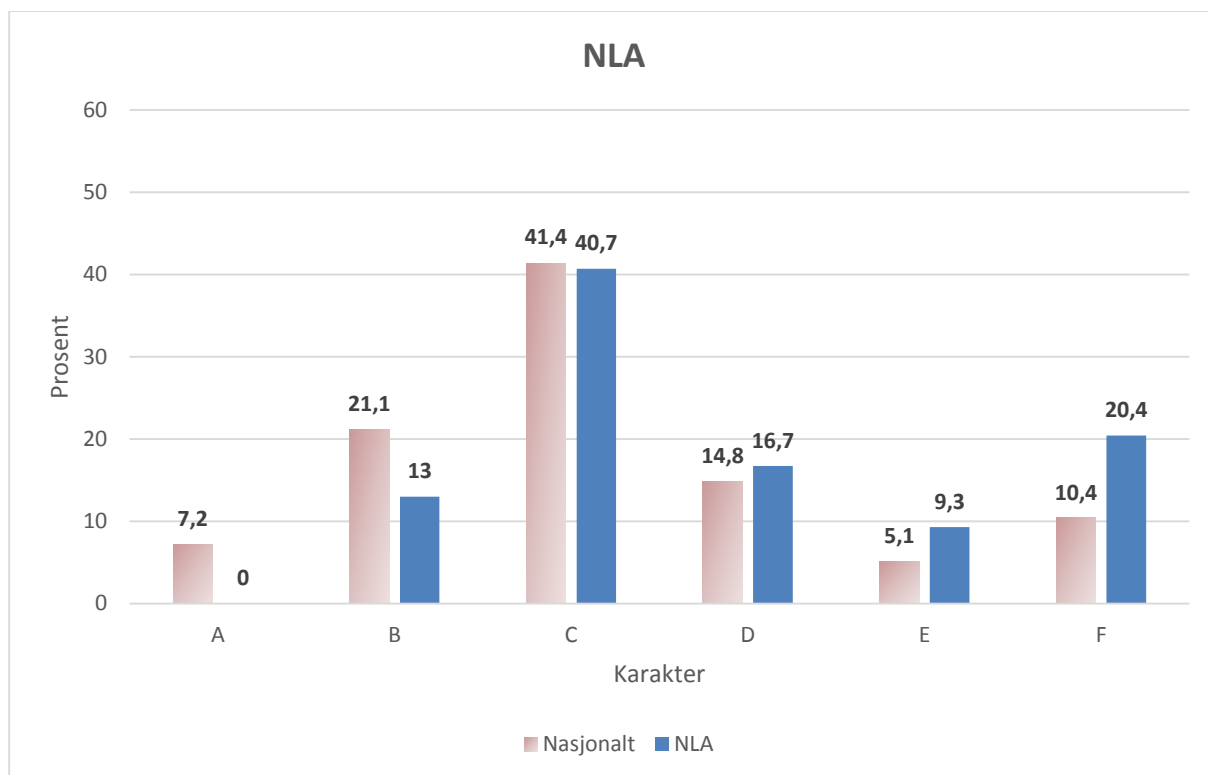
3.10 Høgskulen i Volda

Det var totalt 62 studenter fra HiVo som tok den nasjonale deleksamenen i desember 2015, og gjennomsnittskarakteren var 2,7 (C). Ingen studenter ved HiVo hadde ordinær matematikkeksamen i desember 2015, så vi kan ikke sammenligne de to eksamenene for HiVo. Som vi ser fra karakterfordelingen, var det få studenter som fikk A, men nesten 30 prosent fikk B. For de andre karakterene var det bare små forskjeller fra den nasjonale karakterfordelingen.



3.11 NLA Høgskolen

NLA høgskolen er den eneste private høgskolen som deltok på den nasjonale deleksamenen, og studentene hadde ikke ordinær eksamen i matematikk i desember 2015. Totalt var det 54 studenter fra NLA som tok den nasjonale deleksamenen, og snittkarakteren var 2,2 (D). Ser vi på karakterfordelingen, ser vi at ingen studenter fikk A, mens det kun var 13 prosent som fikk B. NLA Høgskolen hadde også dobbelt så stor strykprosent som det nasjonale gjennomsnittet.

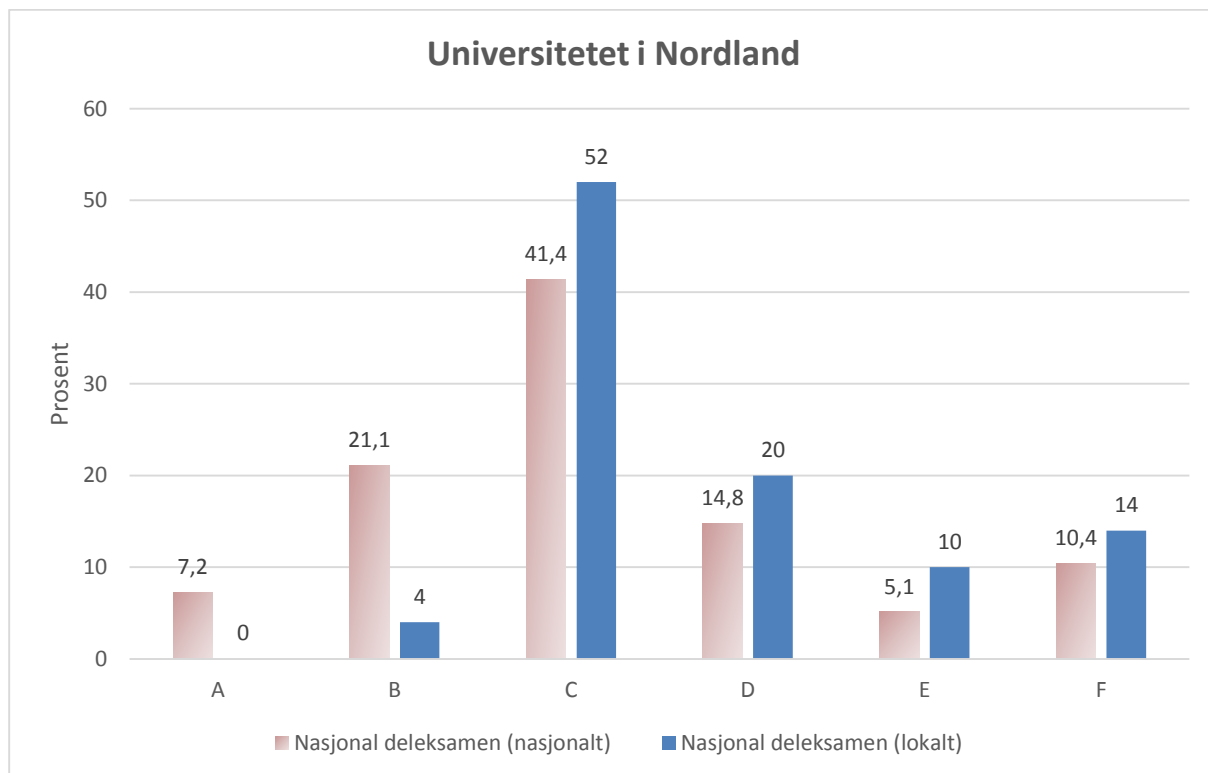


NLA høgskolen rapporterer at matematikkemnet for GLU programmet strekker seg over et helt semester og at studentene kun hadde en forelesning med brøk, desimal og prosent i høstsemesteret og at hoveddelen av undervisningen foregår i vårsemesteret. De påpeker også at den nasjonale deleksamen kom midt i studentenes praksisperiode, og at dette sannsynligvis har gått ut over studentenes eksamensforberedelser. Det at studentene kun hadde hatt en uke med praksis kan også bety at studentene har funnet oppgaver som fokuserer på undervisningskunnskaper utfordrende.

3.12 Universitetet i Nordland

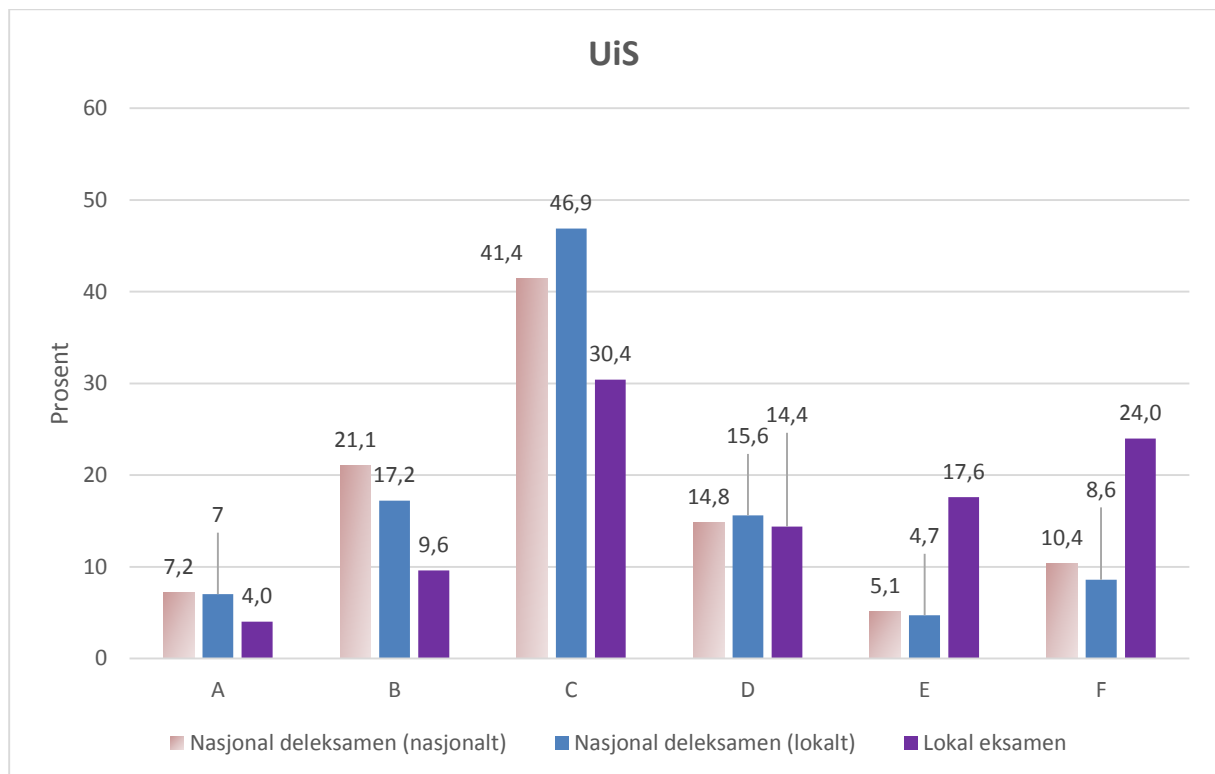
50 studenter ved Universitetet i Nordland tok den nasjonale deleksamenen, og snittet var 2,2 (D). Ved Universitetet i Nordland var det bare studenter ved et av programmene som tok ordinær eksamen i desember. Siden vi ikke vet hvilke studenter som tok begge eksamenene, sammenligner vi ikke resultatene på de to eksamenene.

Dersom vi ser på karakterfordelingen på den nasjonale deleksamenen, ser vi at Universitetet i Nordland ikke hadde noen kandidater som fikk A og kun fire prosent som fikk B. Hele 52 prosent fikk C, mens 44 prosent fikk D, E eller F.



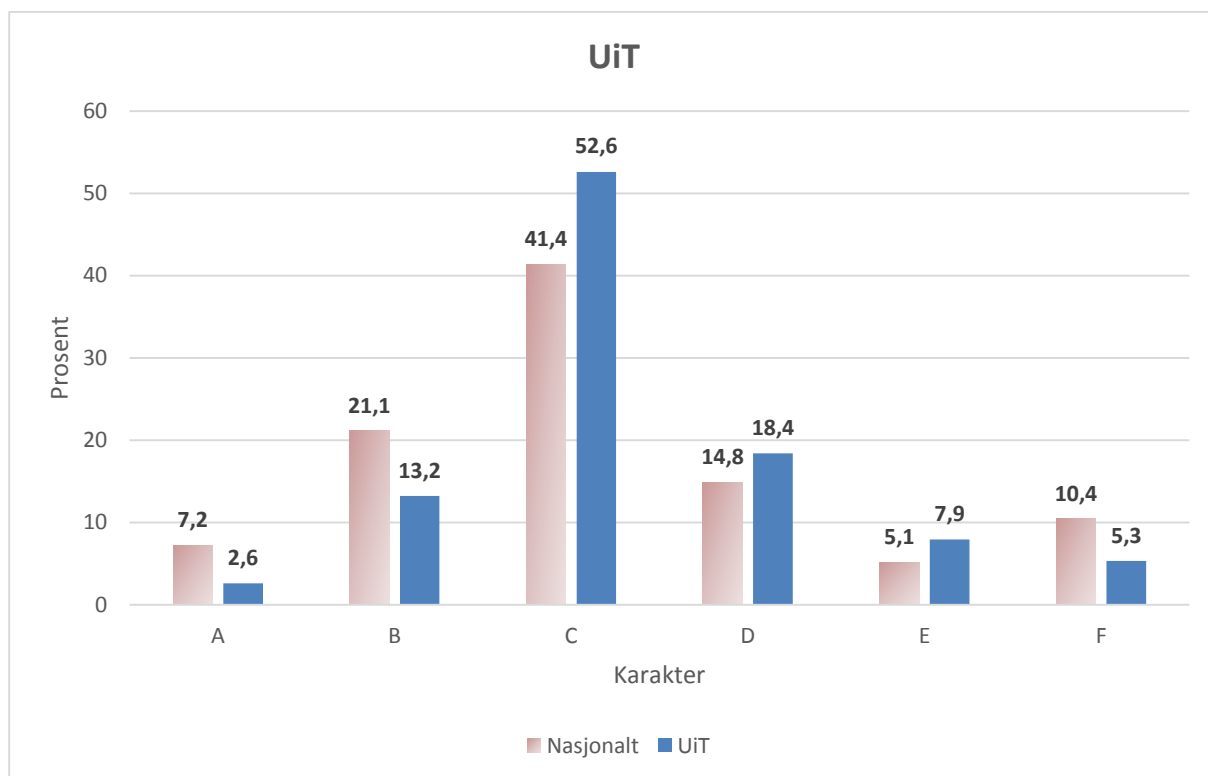
3.13 Universitetet i Stavanger

UiS gjorde det best av universitetene som deltok på den nasjonale deleksamenen, og gjennomsnittet var 2,8 (C), likt med det nasjonale snittet. På den ordinære matematikkeksamenen var gjennomsnittet kun 2,0 (D), nesten lavest av alle institusjonene. Karakterfordelingen viser også at UiS hadde en markant nedgang i andelen kandidater som fikk C (16 prosentpoeng) og en markant økning i andelen kandidater som fikk E (13 prosentpoeng) og F (15 prosentpoeng). Dette er noe overraskende, og selv om NOKUT ikke kan forklare årsaken til sammenhengen, kan det virke som at vanskelighetsgraden på den lokale eksamenen kan være noe høy, eller at sensuren er noe streng. Dette bør fagmiljøet ved UiS vurdere nærmere.



3.14 Universitetet i Tromsø – Norges arktiske universitet

Ved UiT er det studenter ved de 5-årige grunnskoleutdanningene som har tatt den nasjonale deleksamenen. Disse 38 studentene er de eneste av de studentene som har tatt den nasjonale deleksamenen som går på 5-årige studieprogrammer. Karaktergjennomsnittet for UiT var 2,7 (C), altså tett opptil det nasjonale snittet. Studentene ved UiT hadde heller ikke ordinær eksamen i desember 2015. Karakterfordelingen for UiT viser at færre studenter fikk A eller B på den nasjonale deleksamenen enn på den nasjonale karakterfordelingen. Samtidig fikk hele 52,6 prosent C, og kun 5,3 prosent strøk.



4 Kvalitetssikring av oppgavesett og sensur

Hovedformålene med nasjonal deleksamen er som følger:

1. bidra til nyttig informasjon om studentenes kunnskapsnivå
2. gi fagmiljøer mulighet til å sammenligne seg med tilsvarende fagmiljøer ved andre institusjoner
3. virke tillitvekkende for samfunnet

For at disse formålene skal kunne oppnås, må vi vite at eksamensoppgavene måler det de skal måle og at det ikke er tilfeldig hvilken karakter studenten får – altså: oppgavene må være reliable og valide, og sensuren må være reliabel. Som beskrevet over, vil man normalt bruke flere år på å pilotere oppgavesett for å sikre validitet og reliabilitet. Dette var ikke mulig å gjøre i forbindelse med denne nasjonale deleksamenen. For allikevel å kunne si noe om reliabiliteten på oppgavene og sensuren har NOKUT gjort flere analyser i ettertid av både eksamenssettet og sensuren. I dette kapitlet rapporterer vi resultatene av disse analysene.

4.1 Sensurreliabilitet

I samarbeid med eksamensgruppen ba NOKUT alle sensorer om å rapportere detaljert sensur. Det vil si, vi ba hver sensor om å oppgi hvor mange poeng hver kandidat fikk på hver enkelt eksamensoppgave. Sensorene fikk et poengsskjema hvor de skulle fylle inn kandidatnummer og poengene 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 totalskåren til hver enkelt kandidat. Sensorene ble så enige seg imellom om hvor mange poeng kandidaten skulle få på hver oppgave og totalskåren.⁶ På denne måten har NOKUT mulighet til å vurdere sensurreliabiliteten, samt reliabiliteten på eksamensoppgavene (se diskusjon under).

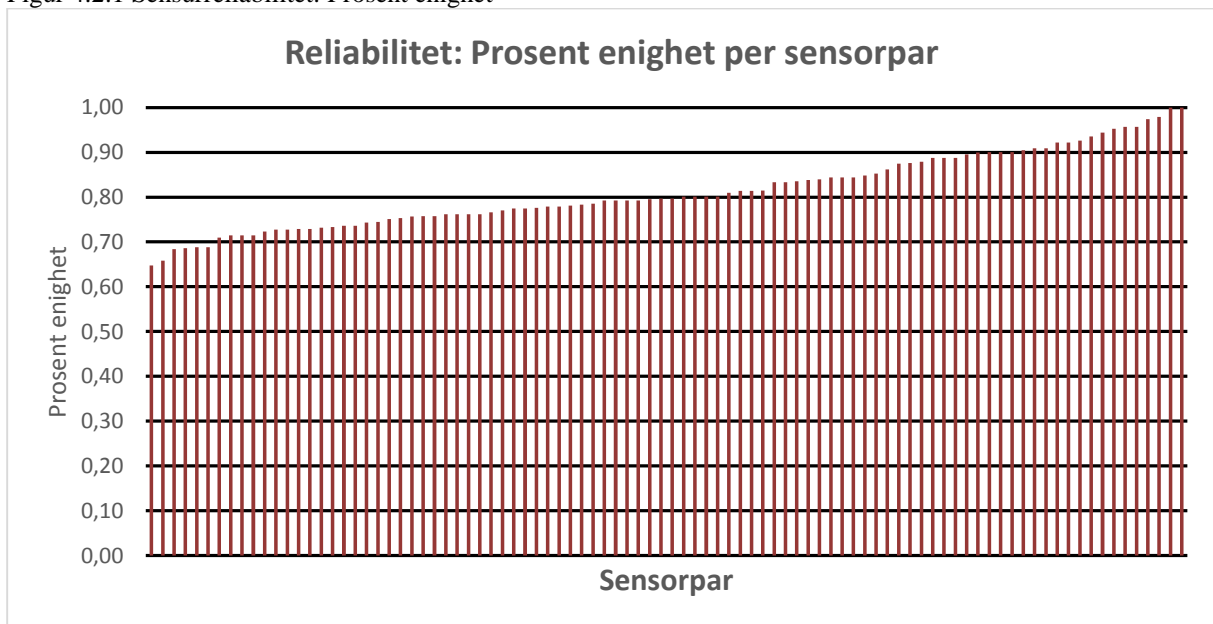
⁶ Alle sensorene ble ikke enige om poeng på hver enkelt oppgave, men de rapporterte totalsummen. I de tilfellene tok NOKUT selv og vurderte den mest sannsynlige samsensuren per oppgave.

Figur 4.1.1 Eksempel på poengsskjema

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Kandidat nr: 1																					
2	1a	1b	1c (i)	1c (ii)	1d (i)	1d (ii)	2a (i)	2a (ii)	2b	2c	3a	3b	3c	3d (i)	3d (ii)	4a	4b	4c (i)	4c (ii)	4d	Totalt	
3	Sensor 1	2	2	1	1	2	2	0	1	2	1	2	2	0	1	0	0	1	0	0	0	20
4	Sensor 2	2	2	1	1	1	2	0	1	2	1	2	2	0	1	0	0	1	0	0	0	19
5	Samsensur poeng	2	2	1	1	2	2	0	1	2	1	2	2	0	1	0	0	1	0	0	0	20
6																					Endelig sum:	20
7	Kandidat nr: 2																					
8	1a	1b	1c (i)	1c (ii)	1d (i)	1d (ii)	2a (i)	2a (ii)	2b	2c	3a	3b	3c	3d (i)	3d (ii)	4a	4b	4c (i)	4c (ii)	4d	Totalt	
9	Sensor 1	2	2	2	2	2	0	2	2	1	2	0	1	0	2	1	1	1	2	0	2	27
10	Sensor 2	2	2	2	2	2	0	2	2	1	2	1	1	0	2	1	1	1	2	0	2	28
11	Samsensur poeng	2	2	2	2	2	0	2	2	1	2	?	1	0	2	1	1	1	2	0	2	27
12																					Endelig sum:	28
13	Kandidat nr: 3																					
14	1a	1b	1c (i)	1c (ii)	1d (i)	1d (ii)	2a (i)	2a (ii)	2b	2c	3a	3b	3c	3d (i)	3d (ii)	4a	4b	4c (i)	4c (ii)	4d	Totalt	
15	Sensor 1	1	2	0	0	2	2	1	0	1	1	2	1	0	2	2	0	1	2	0	2	22
16	Sensor 2	2	2	0	0	2	2	2	0	1	1	2	1	1	2	2	0	1	2	0	2	25
17	Samsensur poeng	2	2	0	0	2	2	2	0	1	1	2	1	1	2	2	0	1	2	0	2	25
18																					Endelig sum:	25
19																						

Som beskrevet over var det totalt 48 sensorer som sensurerte den nasjonale deleksamenen. Disse ble delt inn i grupper på fire og fem, og i hver gruppe sensurerte alle sensorene sammen med hverandre. Hvert sensorpar samsensurerte ca. 11 oppgaver. Til sammen var det 92 sensorpar. Hvert eksamenssett inneholdt 20 spørsmål, så i gjennomsnitt sensurerte hvert sensorpar 220 oppgaver sammen. Figur 4.2.1 viser hvor mange prosent av de 220 oppgavene hvert sensorpar hadde vurdert likt.

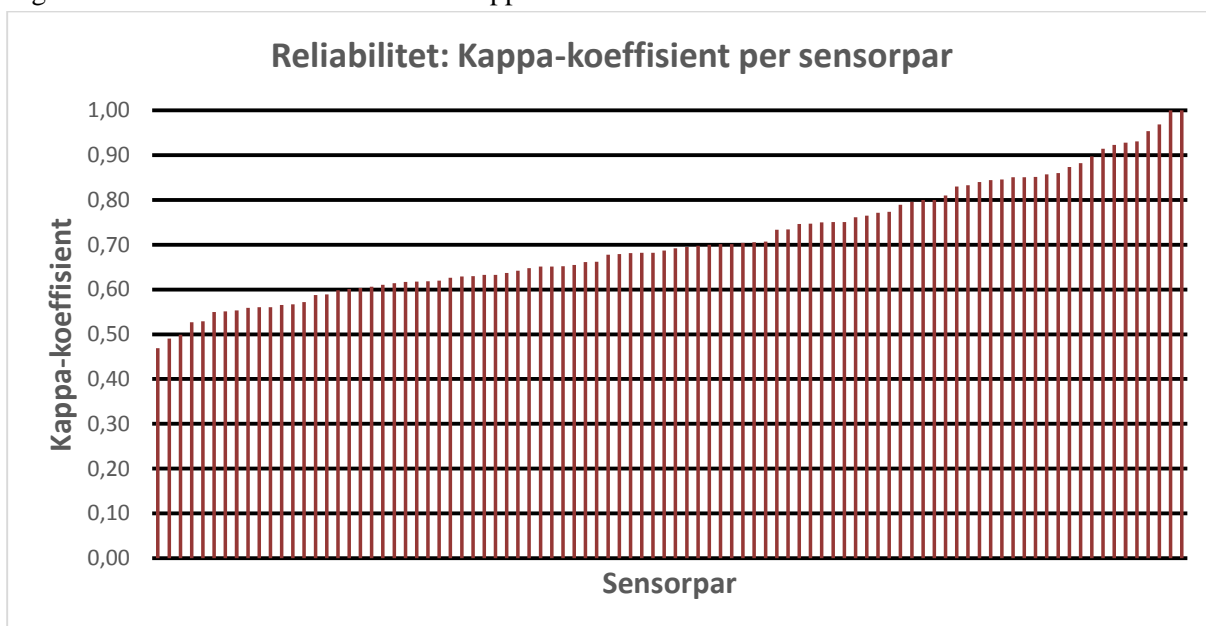
Figur 4.2.1 Sensurreliabilitet: Prosent enighet



Hver stolpe i figuren representerer ett sensorpar, og høyden på stolpen viser hvor mange prosent av oppgavene sensorene hadde vurdert likt. Som vi ser helt til høyre, er det to sensorpar som har vurdert

alle oppgaver likt (100 prosent). Dette er med all sannsynlighet ikke korrekt; her er det nok snakk om feilrapportering fra sensorparene. Ifølge forskningslitteraturen innenfor statistikk og reliabilitetsmålinger er «prosent enighet» ikke et veldig godt mål på sensurreliabilitet. Grunnen til dette er at et slikt mål ikke tar hensyn til den underliggende sannsynligheten for at et sensorpar, basert på ren gjetting, ville vært enig på en rekke av oppgavene.⁷ Et bedre reliabilitetsmål er Cohens kappa, også kalt kappa-koeffisient. Cohens kappa tar hensyn til den underliggende sannsynligheten for gjetting og regner ut en koeffisient som gir oss et bedre reliabilitetsmål. I figur 4.2.2 viser vi kappa-koeffisienten per sensorpar.

Figur 4.2.2. Sensurreliabilitet: Cohens kappa



Kappa-koeffisienten beveger seg fra 0 (der man ikke er enig på noen oppgaver) til 1 (hvor man er enig på alle oppgaver). Av de 92 sensorparene har 76 sensorpar (83 prosent) en kappa-koeffisient på over 0,60. Hvor godt er dette? I forskningslitteraturen opererer man med følgende skala:⁸

- 0,0 – 0,20 svak enighet
- 0,21 – 0,40 rimelig enighet
- 0,41 – 0,60 moderat enighet
- 0,61 – 0,90 betydelig enighet
- 0,81 – 1,0 nesten perfekt enighet

Resultatene er med andre ord meget 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

⁷ Se for eksempel, Krippendorf, K. (2004). «Reliability in content analysis: Some common misconceptions and recommendations» Human Communications Research 30(4): 411-433, og Lombard, M. et al. (2002). «Content analysis in mass communication: Assessment and reporting of intercoder reliability» Human Communication Quarterly 28(4): 587-604.

⁸ Se for eksempel Landis, J. & Koch, G. (1977). «The Measurement of Observer Agreement for Categorical Data» Biometrics 33: 159-174.

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. I neste delkapittel ser vi nærmere på selve oppgavesettet.

4.2 Validitet og reliabilitet av eksamenssettet

Her ønsker vi å vurdere om eksamenssettet måler det vi ønsker å måle (validitet), om eksamenssettet er reliabelt – det vil si: hvis en student hadde fått et lignende, men ikke identisk eksamenssett, hadde da studenten fått en bedre eller dårligere karakter? Og til slutt: hvor vanskelig var eksamenssettet?

4.2.1 Eksamenssettets validitet

Vi har ikke gjort statistiske analyser for å vurdere hvor valid eksamenssettet var. Eksamensgruppen var bredt sammensatt med fagansatte fra flere universiteter og høyskoler som deltok og ikke deltok på eksamen. Denne eksamensgruppen jobbet seg gjennom emneplaner og pensum, og lagde oppgaver de mente måler det de ønsker å måle. Etter sensuren sendte NOKUT en spørreundersøkelse til sensorene og ba dem om å vurdere hvor vanskelig eksamenssettet var, hvor dekkende det var og hvor klart formulert det var. Hvis vi ser på de to siste spørsmålene, ser vi at eksamen dekket godt og var godt formulert. NOKUT og eksamensgruppen er totalt sett relativt sikre på at validiteten av eksamenssettet var høyt.

4.2.2 Eksamenssettets reliabilitet

Til å vurdere eksamenssettets reliabilitet og kvalitet har NOKUT engasjert Rolf Vegar Olsen fra Enhet for kvantitative utdanningsanalyser ved Institutt for lærerutdanning og skoleforskning ved Universitet i Oslo. Olsen har brukt en Rasch-analyse for å vurdere eksamenssettet i sin helhet. Vi presenterer ikke hele analysen her, men gjengir hovedkonklusjonene.

I hovedsak viser analysen at den nasjonale deleksamenen var en eksamen med gode oppgaver. Oppgavene diskriminerte godt. Det vil si at kandidater som får ett poeng på en oppgave, er litt dyktigere enn de som får 0 poeng, og de som får to poeng, er litt dyktigere enn de som får ett poeng. Det er allikevel ikke alle oppgaver som fungerer like godt. Det er spesielt svarene som gir ett poeng, som kan være problematiske. I noen oppgaver er det slik at kandidater får ett poeng uavhengig av deres dyktighet. Årsaken til dette kan ligge i sensorveiledningen – at det ikke er helt tydelig hva som skal gi ett poeng. En annen årsak kan ligge i måten sensorene har blitt enige om poengene for hver enkelt oppgave. Hvis en sensor har gitt to poeng og en annen har gitt null poeng, kan sensorene ha blitt enige om å gi ett poeng, uten at man har gjort en grundig sjekk av om dette samsvarer med sensorveiledningen. En tredje grunn kan være at oppgavene skiller dårlig og at det er vanskelig å si om en student fortjener 0, 1, eller 2 poeng.

Videre viser analysen at hver enkelt oppgave passer meget godt til den underliggende statistiske modellen. Det vil si at vi statistisk sett ønsker at sannsynligheten for at en student skal få to poeng på en enkeltoppgave, bør ha en klar sammenheng med hvor dyktig studenten er (dyktighet betyr hvor

mange poeng studenten fikk totalt). Med andre ord, jo dyktigere en student er, jo større sannsynlighet er det at studenten får 2 poeng på den enkelte oppgaven.

Analysen viser at reliabiliteten er litt lav (Cronbachs alfa=0,83). Et vanlig krav for denne typen prøver er en Cronbachs alfa på over 0,85. Hovedårsaken til dette er at oppgavesettet inneholdt for få oppgaver. Substansielt betyr dette at målefeilen (95 prosent konfidensintervall) på totalskåren ligger på +/- 6 totalpoeng. Med andre ord, hadde en student som fikk 30 poeng på denne eksamenen tatt en lignende eksamen en gang til, er det en stor sannsynlighet for at hennes totalskåre hadde blitt mellom 24 og 36 poeng. Det vil si at karakteren kunne endret seg fra C til B eller C til D.

Til slutt viser analysen at oppgavesettet hadde litt for mange enkle oppgaver. Gjennomsnittet var 64 prosent riktig, som er noe høyt. Dette kan også bedres ved å lage flere oppgaver.

Selv om reliabiliteten er litt lav og eksamenssettet noe enkelt, viser resten av analysen at eksamenssettet var meget godt formulert. Med tanke på at eksamensgruppen ikke hadde tid til å pilotere oppgavene, er dette veldig positivt. Basert på sensuranalysen og analysen av oppgavesettet er det ingen tvil om at den nasjonale deleksamenen i matematikk gir oss god informasjon om studentenes kunnskap i brøk, desimaltall og prosentregning.

5 Konklusjon

Nasjonal deleksamenen i matematikk for grunnskolelærerutdanningene var i all hovedsak vellykket. Gjennomføringen av eksamenen og sensuren fungerte godt. Analysene viser at eksamenssettet og sensuren er reliable og valide, og resultatene gir oss en mulighet til å si noe om studentenes kunnskapsnivå.

Resultatene indikerer at kunnskapsnivået i brøk, desimaltall og prosentregning er relativt høyt. Hele 69,7 prosent av studentene fikk karakteren C eller bedre, og kun 10,4 prosent strøk. Det er noen forskjeller på institusjonsnivå, men alle institusjonene hadde et karaktersnitt på D eller bedre, og dersom man fjerner kandidater som strøk fra gjennomsnittsmålingen, ligger nesten alle institusjonene med ett snitt på C eller bedre.

Sammenligningen med de ordinære eksamenene viser også at det ikke er slik at man på lokalt plan gir enklere eksamener eller er «snillere» med sensuren. Resultatene fra de ordinære eksamenene var noe svakere enn fra nasjonal deleksamen, men her er det viktig å poengtere at de ordinære eksamenene dreier seg om hele pensumet og at det derfor ikke er unaturlig at resultatene er noe svakere.

For å måle i hvilken grad institusjonene bidrar til studentenes kunnskap må NOKUT ha tilgang til bakgrunnsinformasjon om studentene. Dette har ikke NOKUT på nåværende tidspunkt, men vi får tilgang til dette i løpet av kort tid. Det er allikevel sannsynlig at institusjonsbidraget vil vise seg å være relativt lite. Hovedgrunnen til dette er at denne eksamenen kommer på slutten av studentenes første semester i studieprogrammet. Det er med andre ord sannsynlig at studentenes resultater på den nasjonale deleksamenen i hovedsak kan forklares av hvor mye matematikk de kunne da de startet på studieprogrammet. NOKUT foreslår derfor at man i fremtiden vurderer å avholde nasjonal deleksamenen i et senere semester.

En annen utfordring var at dette var en isolert deleksamen. Det vil si at karakteren kommer på vitnemålet, men at den ikke er tellende. Studentene kan altså møte opp på eksamen, stryke på eksamen og allikevel få en godkjent utdanningen. NOKUT foreslår derfor at man i fremtiden kun avholder integrerte nasjonale deksamener. Altså at de nasjonale deksamenene blir tellende. Dette kan gjøres ved at resultatet på den nasjonale deksamenen teller som en viss prosentandel på en avsluttende karakter i et emne, eller at karakteren på den nasjonale deksamenen er den eneste vurderingen studentene får i et emne.