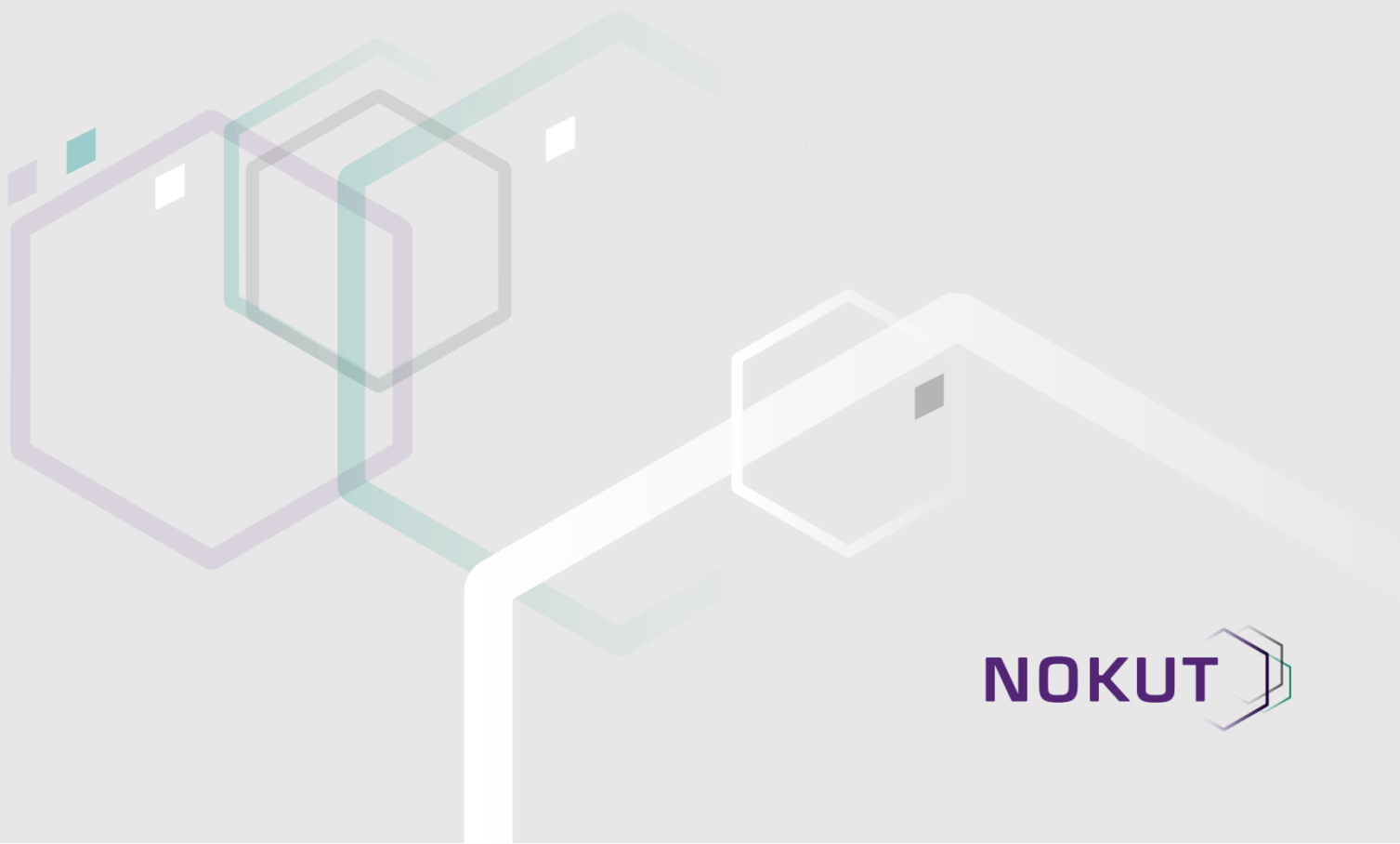


NOKUTs oppsummeringer

Nasjonal deleksamen i matematikk for grunnskolelærerutdanningen

Delrapport 2



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)
Dato:	12.07.2016

Sammendrag

11. mai 2016 arrangerte NOKUT i samarbeid med Nasjonalt råd for lærerutdanning (NRLU) og 13 utdanningsinstitusjoner den andre nasjonale deleksamenen i matematikk for grunnskolelærerutdanning 1–7 (GLU 1–7) og grunnskolelærerutdanning 5–10 (GLU 5–10). Totalt 997 studenter gjennomførte eksamen og fikk sensur. Eksamenen omfattet alle GLU-studenter som fulgte undervisning i vårsemesteret hvor eksamenstemaet undervisningskunnskap i brøk, desimaltall og prosentregning inngikk. I tillegg er det noen studenter som har tatt denne eksamenen som en kontinuasjonseksamen. Denne delrapporten må ses i sammenheng med den første delrapporten, som omhandlet den nasjonale deleksamenen som ble avholdt 1. desember 2015.

Resultatene for vårens nasjonale deleksamen er svake, både på nasjonalt nivå og på institusjonsnivå, og de er langt svakere enn resultatene på den nasjonale deleksamenen som ble avholdt i desember 2015. Våren 2016 strøk hele 37 prosent av kandidatene, og kun 6,6 prosent fikk karakteren A eller B. Andelen kandidater som fikk A eller B høsten 2015, var 28,3 prosent, og andelen som strøk, var 10,4 prosent. Årsakene til de svake resultatene er sammensatte, og våre analyser peker i retning av fire hovedfaktorer eller en kombinasjon av disse.

Én faktor er at et stort antall studenter med høy sannsynlighet har nedprioritert den nasjonale deleksamenen til fordel for ordinære eksamener i matematikk, og derfor kan vi ikke stole på at resultatene gir et korrekt bilde av studentenes kunnskapsnivå. Den nasjonale deleksamenen i grunnskolelærerutdanningene er obligatorisk, men ikke tellende for studentene. Dersom nasjonale deksamener skal kunne brukes til å gi nyttig informasjon om studentenes kunnskapsnivå og samtidig være et virkemiddel for institusjonene i deres egen kvalitetsutvikling, må man kunne stole på resultatene. Derfor bør nasjonale deksamener være tellende for studentene.

En annen faktor er at andelen studenter fra GLU 1–7 var langt høyere ved vårens eksamen enn ved høstens eksamen. Analysene viser at studenter ved GLU 5–10, som selv har valgt matematikk, gjør det bedre enn studenter ved GLU 1–7, som har matematikk som et obligatorisk emne.

Den tredje og siste faktoren er oppgavesettets vanskelighetsgrad og sensorveiledningen. Analysene fra de to eksamenene viser at det var noen flere enklere oppgaver i 2015 enn i 2016, og at det var enklere for studentene å oppnå ett poeng på enkeltoppgaver. Undersøkelsene blant sensorene støtter funnene fra Rasch-analysene. Det er med andre ord sannsynlig at oppgavesettet var noe vanskeligere, og at sensorveiledningen bidro til at det var vanskeligere å oppnå ett poeng på noen av enkeltoppgavene.

I tillegg til å forklare resultatforskjellene har vi også vist at studentenes startkompetanse (målt i karakterpoeng fra videregående skole) har stor innvirkning på resultatene. Det samme gjelder matematikkarakteren fra videregående skole. Studenter med 4 eller bedre i matematikk gjør det systematisk bedre enn studenter med dårligere enn 4 i matematikk. NOKUT har nå gjennomført nasjonal deleksamen i matematikk for grunnskolelærerutdanningene to ganger, og gjennomføringen av både eksamenen og sensuren har fungert svært godt. Nasjonale deksamener kan i prinsippet fungere som et godt virkemiddel i sektorens kvalitetsarbeid, men det må legges bedre til rette for at resultatene skal kunne gi et mer korrekt bilde av studentenes kunnskapsnivå.

Innhold

1	Innledning	1
1.1	Organisering av nasjonal deleksamen	1
1.1.1	Eksamensdagen	1
1.1.2	Sensur	1
1.2	Utvalget	2
2	Resultater	3
2.1	GLU 1–7 vs. GLU 5–10	4
2.2	Obligatorisk vs. tellende resultat	6
2.3	Karakterpoeng og matematikkarakter fra videregående skole	7
2.4	Eksamensoppgaver og sensorveiledning	11
3	Institusjonsresultater	12
3.1	Ordinær vs. nasjonal deleksamen våren 2016	14
4	Sensurreliabilitet	15
5	Konklusjon	17

1 Innledning

NOKUT fikk høsten 2014 i oppdrag av Kunnskapsdepartementet å gjennomføre en mulighetsstudie og et pilotprosjekt med nasjonale deleksamener i høyere utdanning. Tre rammeplanstyrte profesjonsutdanninger ble valgt ut til å delta i prosjektet, som finner sted fra 2015 til 2017. 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 deleksamen i matematikk i GLU 1–7 og i GLU 5–10 for studenter som har valgt matematikk, og som ble avholdt 11. mai 2016. Rapporten må ses i sammenheng med den første delrapporten, som kom i mars og omhandlet den nasjonale deleksamenen som ble avholdt i høstsemesteret 2015.¹ I den første delrapporten beskrev vi formålet med prosjektet og de viktigste innvendingene mot prosjektet, samt en del av utfordringene. Vi kommer derfor ikke videre inn på disse temaene i denne rapporten.

1.1 Organisering av nasjonal deleksamen

Det er den samme gruppen som har utviklet eksamensoppgavene og sensorveiledningene til begge de nasjonale deleksamenene i grunnskolelærerutdanningene, mens NOKUT har hatt et overordnet ansvar overfor sektoren og departementet og et administrativt og logistisk ansvar i forbindelse med eksamensgjennomføringen, som bestilling og fordeling av eksamenspapir, kandidatnumre til eksamen, fordeling av besvarelser til sensorkorpset samt formidling av eksamensresultatene til institusjonene.

1.1.1 Eksamensdagen

Den andre nasjonale deleksamenen i GLU ble avholdt som en firetimers skriftlig skoleeksamen uten bruk av hjelpemidler 11. mai 2016. Det var 13 institusjoner som deltok, fordelt på 18 studiesteder, og dette gjelder studenter i andre og fjerde semester og GLU 1–7 og 5–10, med en klar hovedvekt (ca. 80 prosent) av studenter fra GLU 1–7. Totalt gjennomførte 997 av 1049 kandidater eksamen, noe som tilsvarer en gjennomføringsgrad på 95 prosent.

1.1.2 Sensur

For å sikre likebehandling og full anonymitet ble det satt sammen et sensorkorps på 48 personer, hvor alle lærerutdanningsinstitusjonene er representert i sensorkorpset. 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. Hver sensor

¹ «Nasjonal deleksamen i matematikk for grunnskolelærerutdanningen. Delrapport 1, mars 2016»: http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Kunnskapsbasen/Rapporter/H%c3%b8yere%20utdanning%20e2%80%93%20oppsummeringer/2016/Nasjonal_deleksamen_i_matematikk_for_grunnskolel%c3%a6rerutdanningen_2016-1.pdf

sensurerte sammen med tre eller fire andre. Hvert sensorpar sensurerte omtrent 11 besvarelser, og hver enkelt sensor vurderte i gjennomsnitt mellom 41 og 42 besvarelser.

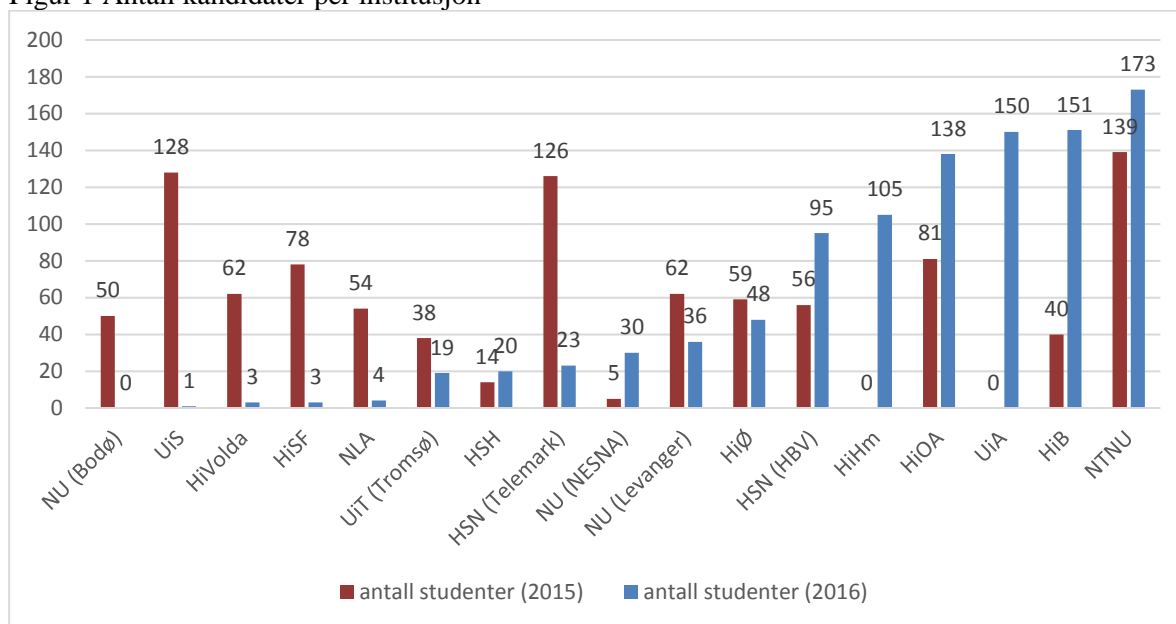
1.2 Utvalget

Som i desember 2015 deltok studenter fra både GLU 1–7 og GLU 5–10-programmer. Tabell 1 viser en oversikt over hvilke programmer som deltok fra hver institusjon, og figur 1 viser hvor mange studenter fra hver institusjon som tok eksamen i desember 2015 og i mai 2016.

Tabell 1 Programmer som deltok fra hver institusjon

Institusjon	Høst 2015	Vår 2016
HiOA	GLU 5–10	GLU 1–7
Nord Universitet (ikke HiNT/Nesna)	GLU 5–10 og 1–7	Ingen
UiS (kun 1 student i 2016)	GLU 5–10 og 1–7	Kun 1 student
HBV / HSN (2016)	GLU 5–10	GLU 5–10 og 1–7
NLA	GLU 5–10 og 1–7	Kun 9 studenter
HiST (2015) / NTNU (2016)	GLU 5–10	GLU 1–7
Nesna (kun 5 studenter i 2015) / NU (2016)	GLU 5–10	GLU 1–7
HiØ	GLU 5–10	GLU 1–7
HSH	GLU 5–10	GLU 5–10
UiT	GLU 5–10 og 1–7	GLU 1–7
HiT (2015) / HSN (2016)	GLU 5–10 og 1–7	GLU 1–7 og 5–10 (kun 15 stud.)
HiNT (2015) / NU (2016)	GLU 1–7	GLU 5–10
HiB	GLU 5–10	GLU 1–7
HiVo (kun 3 studenter i 2016)	GLU 5–10	GLU 1–7
HiHm	Ingen	GLU 5–10 og 1–7
UiA	Ingen	GLU 5–10 og 1–7
HiSF	GLU 5–10 og 1–7	Kun 3 studenter

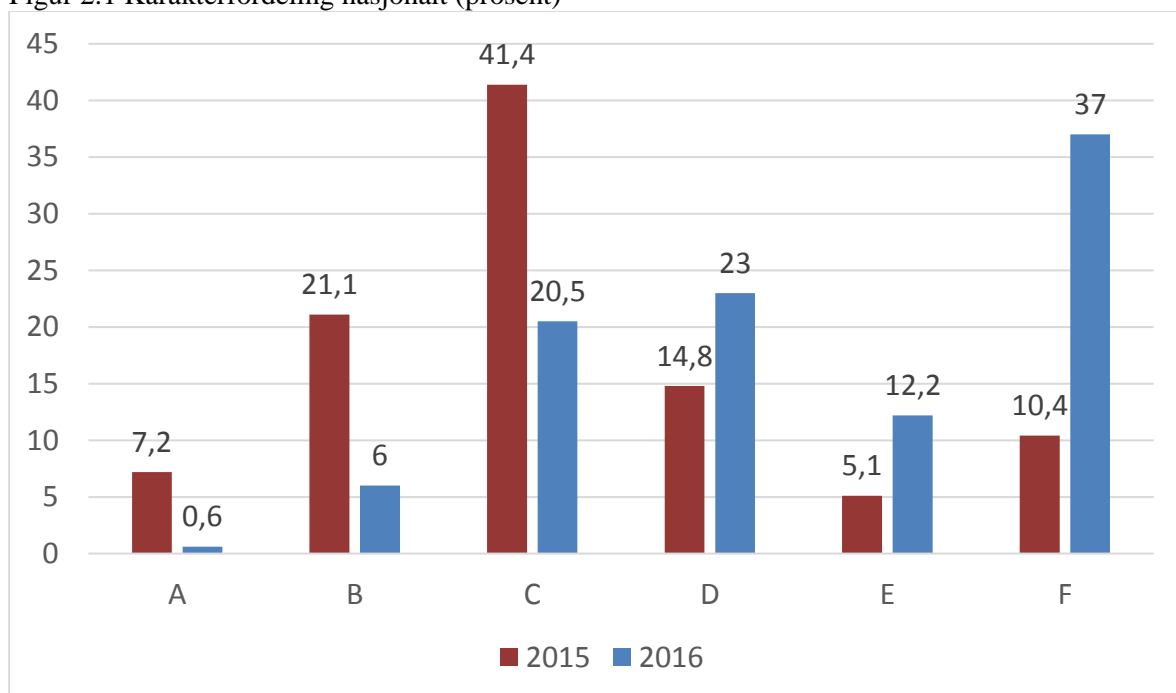
Figur 1 Antall kandidater per institusjon



2 Resultater

I den første delen av dette kapittelet fokuserer vi på nasjonale resultater og ikke på resultater på institusjonsnivå. Figur 2.1 viser karakterfordelingen fra høsten 2015 og våren 2016.

Figur 2.1 Karakterfordeling nasjonalt (prosent)



Som vi ser i figuren, var resultatene fra høsten 2015 markant bedre enn resultatene fra våren 2016. Våren 2016 strøk hele 37 prosent av studentene, og kun 6,6 prosent av studentene fikk karakteren A

eller B. Den nasjonale gjennomsnittskarakteren var D, men den lå meget nær E. Høsten 2015 strøk kun 10,4 prosent av kandidatene, og hele 28,3 prosent fikk A eller B. Gjennomsnittet høsten 2015 var C. Dersom vi gjør om karakterene til tall (A=5, B=4, C=3, D=2, E=1 og F=0), ser vi at gjennomsnittskarakteren var 2,8 høsten 2015 og 1,5 våren 2016. Den reelle forskjellen er altså 1,3 karakterpoeng. Hva kan forklare denne store forskjellen?

Det er i all hovedsak fire mulige forklaringer på den store forskjellen. Vi beskriver disse forklaringene her og kommer tilbake til hver enkelt senere i rapporten.

1. GLU 1–7-studenter har matematikk som obligatorisk fag, mens de studentene fra GLU 5–10 som har tatt den nasjonale deleksamenen, har matematikk som valgfag. Det betyr at GLU 5–10-studenter sannsynligvis har en større interesse for matematikk og muligens bedre forkunnskaper (i gjennomsnitt) enn studenter fra GLU 1–7-programmer. Høsten 2015 var ca. 65 prosent av studentene som tok den nasjonale deleksamenen, fra GLU 5–10-programmer, mens våren 2016 var det 20 prosent, mot henholdsvis 35 og 80 prosent studenter fra GLU 1–7-programmer. Som vi viser under, gjør GLU 5–10-studenter det markant bedre enn GLU 1–7-studenter på vårens eksamen.
2. Nasjonal deleksamen er obligatorisk, men ikke tellende. Alle studenter må ta den nasjonale deleksamenen, og resultatet blir oppført på vitnemålet, men selv om studentene stryker, får de vitnemål hvis de andre fagene er bestått. Ordningen var den samme under begge eksamenene, men høsten 2015 var sannsynligvis få studenter klar over dette. NOKUT beskrev denne problemstillingen i sin delrapport 1 i mars 2016, og flere presseoppslag fra juni 2016 indikerer at både studenter og faglærere var klar over dette før vårens eksamen. Dette har etter all sannsynlighet ført til at mange studenter har valgt å nedprioritere den nasjonale deleksamenen våren 2016.
3. Som vi viser under, er studentenes startkompetanse (karakterpoeng fra videregående skole) den variabelen med størst effekt på studentenes resultat på den nasjonale deleksamenen. En hypotese kan være at studentene som tok eksamen høsten 2015, hadde høyere startkompetanse enn studentene som tok eksamen våren 2016.
4. Til slutt kan det være at eksamensoppgavene våren 2016 var vanskeligere eller krevde mer av studentene enn oppgavene høsten 2015. Det er også mulig at sensorveiledningen bidro til at sensuren var strengere våren 2016 enn høsten 2015.

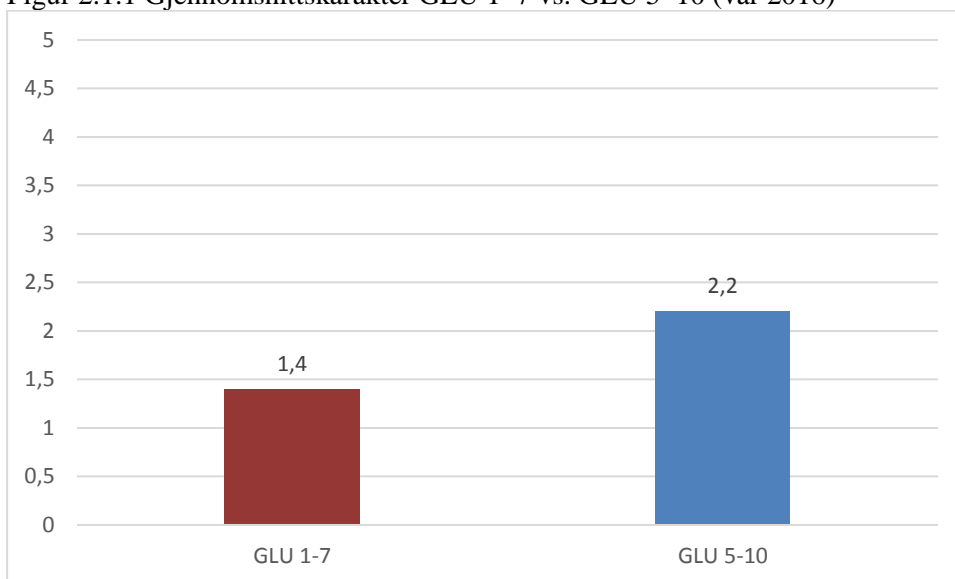
2.1 GLU 1–7 vs. GLU 5–10

Som nevnt over kan én av årsakene til den store forskjellen i resultatene fra høsten 2015 til våren 2016 være at andelen GLU 1–7-studenter var mye høyere våren 2016 enn høsten 2015. For våren 2016 vet vi at ca. 85 prosent av studentene kom fra GLU 1–7, mens for høsten 2015 antar vi at ca. 35 prosent av studentene gikk på GLU 1–7.²

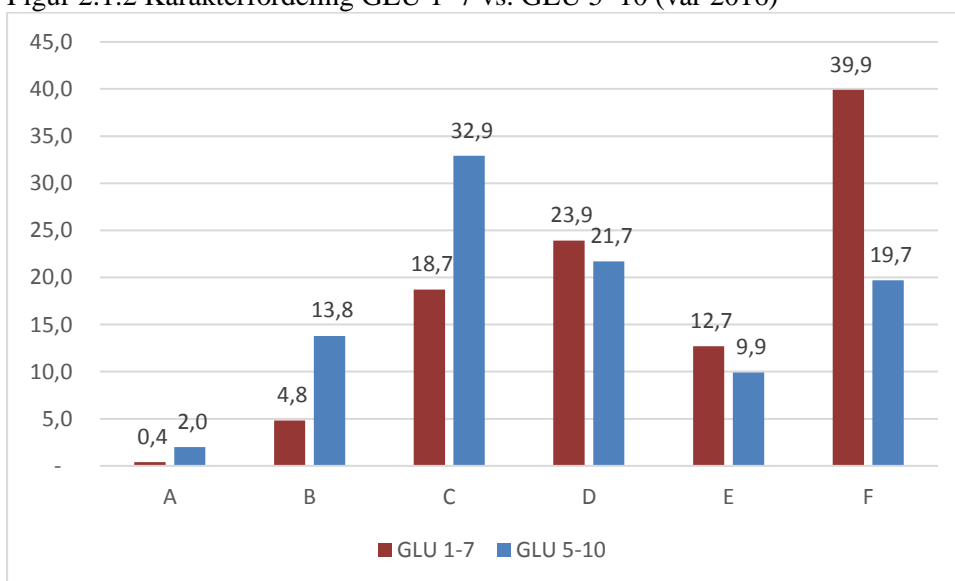
Basert på vårens resultater vet vi at studenter fra GLU 5–10 gjør det signifikant bedre enn studenter fra GLU 1–7. Figur 2.1.1 viser forskjellen i snittkarakter for studenter på de to ulike programmene, og figur 2.1.2 viser karakterfordelingen for de to gruppene.

² Vi har individdata for 982 av 997 studenter våren 2016. Vi mangler data for Høgskolen i Sørøst-Norge – campus Telemark, men vi kan uansett si at ca. 80 prosent kommer fra GLU 1–7. Vi mangler individdata for høsten 2015, men har gjennom kommunikasjon med institusjonene anslått at ca. 35 prosent av studentene som tok eksamen høsten 2015, gikk på GLU 1–7.

Figur 2.1.1 Gjennomsnittskarakter GLU 1–7 vs. GLU 5–10 (vår 2016)



Figur 2.1.2 Karakterfordeling GLU 1–7 vs. GLU 5–10 (vår 2016)



Som vi ser av figur 2.1.1 og 2.1.2, gjør studenter på GLU 5–10 det bedre enn studenter på GLU 1–7. Gjennomsnittskaracteren er 0,8 karakterpoeng bedre, men dette utgjør i praksis en hel bokstavkarakter (GLU 1–7 har E i snitt, mens GLU 5–10 har D i snitt). Vi ser også på karakterfordelingen at nesten 49 prosent av GLU 5–10-studentene får C eller bedre, mens kun 19 prosent av GLU 1–7-studentene får C eller bedre. Det er også stor forskjell i strykprosent (23,9 prosentpoeng). Disse forskjellene er både statistisk og substansielt signifikante.

GLU 5–10-studentene gjorde det altså bedre på vårens eksamen, og selv om vi på nåværende tidspunkt ikke kan teste om forskjellene var de samme høsten 2015, er det sannsynlig at det også da var en forskjell mellom 1–7 og 5–10. Det at majoriteten av studentene høsten 2015 var fra GLU 5–10, mens

majoriteten av studentene våren 2016 var fra GLU 1–7, forklarer sannsynligvis noe av den store forskjellen på resultatene. Det er likevel verdt å merke seg at begge gruppene hadde relativt svake resultater våren 2016. Forskjellen er såpass stor at andre forhold også må ha spilt inn.

2.2 Obligatorisk vs. tellende resultat

Den nasjonale deleksamenen i matematikk for grunnskolelærerutdanningen er en isolert eksamen som er obligatorisk, men ikke tellende for studentene. Det betyr at resultatet kommer på studentenes vitnemål, men selv om en student stryker på eksamen, vil vedkommende få utstedt vitnemål. Dette er ingen optimal løsning, og NOKUT poengterte dette i delrapport 1 i mars 2016. Hovedgrunnen til at det ble slik, var at dette var et pilotprosjekt, og at det var nødvendig for å få satt i gang prosjektet med eksamener i 2015. Da den første deleksamenen ble avholdt i desember 2015, var dette lite kjent, men både studenter og faglærere ble gjort oppmerksomme på dette i evalueringsmøter og i NOKUTs første rapport i mars 2016. Konsekvensen av denne løsningen er at studentene kan nedprioritere den nasjonale deleksamenen uten at dette får særlige konsekvenser. Flere oppslag i pressen etter at NOKUT publiserte de første resultatene mandag 20. juni, indikerer at mange studenter har gjort nettopp dette før vårens nasjonale deleksamen. For eksempel sa rektor ved Høgskolen i Sørøst-Norge (HSN) følgende til Khrono 21. juni 2016:

«På denne prøven er som kjent stryk lik bestått. Det vet studentene. Resultatet av prøven gir med andre ord ikke et bilde av studentenes kunnskap eller mangelfulle kunnskap i matematikk, men synes heller å vise at studenter tenker strategisk. De prioriterer arbeidsinnsats der det gir uttelling.»

Til VG 22. juni 2016 sa leder for Pedagogstudentene i Utdanningsforbundet Silje Marie Bentzen at «... studenter ga tilbakemelding på at nasjonal deleksamen ble avholdt nært i tid med ordinære eksamener, og at flere kan ha nedprioritert den på grunn av det.»

I samme VG-artikkel sier en lærerstudent: «Når det nærmet seg eksamenstid fikk vi beskjed om at denne eksamen ikke ga oss noen studiepoeng. Jeg fikk inntrykk av det hadde ikke noe å si om vi strøk, eller hvilken karakter vi fikk på deleksamen. ... Det sier seg selv at de fleste ikke leste eller tenkte på denne deleksamenen i det hele tatt.»

NOKUT har også sendt ut en undersøkelse til alle sensorer som deltok på vårens eksamen. Flere av disse har også poengtert at de kjenner til at studentene nok har nedprioritert den nasjonale deleksamenen.

Det at en del studenter har valgt å nedprioritere den nasjonale deleksamenen, fører til at NOKUT, Kunnskapsdepartementet og sektoren for øvrig ikke kan stole på at resultatene gir et korrekt bilde av studentenes kunnskapsnivå. Dette er uheldig. Som vi beskriver under, mener NOKUT at nasjonale deksamener kan være et godt virkemiddel for institusjonene til å videreutvikle sine utdanninger, men det krever at resultatene er robuste og reliable. Dersom et høyt antall studenter opplever at de kan nedprioritere denne eksamenen, vil sektoren ikke ha mulighet til å utnytte eksamenens fulle potensial.

NOKUT ser for seg en relativt enkel løsning på dette problemet. Vi foreslår at den nasjonale deleksamenen vil utgjøre en viss andel av studentenes totale matematikkarakter. Studentene vil med andre ord ta to eksamener i matematikk. Den ene eksamenen er den nasjonale deleksamenen i brøk, prosent og desimaltall, mens den andre eksamenen dekker resten av pensumet. Man kan se for seg at

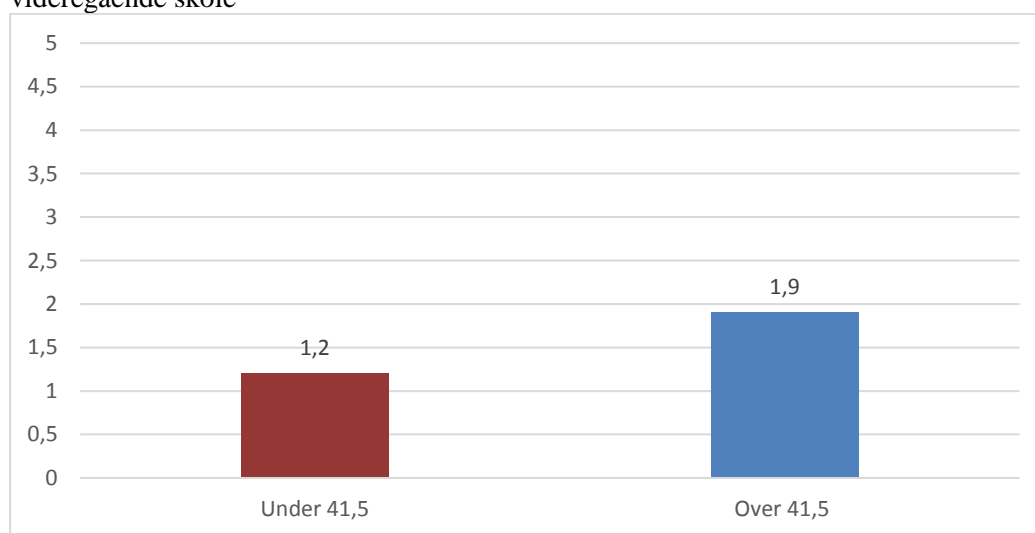
den nasjonale deleksamenen for eksempel teller 30 prosent av studentenes totale matematikkarakter. På denne måten vil eksamenen både være obligatorisk og tellende.

2.3 Karakterpoeng og matematikkarakter fra videregående skole

Resultatene fra vårens eksamen viser at selv om en del studenter kan ha nedprioritert den nasjonale deleksamenen, har studentenes generelle startkompetanse og matematikkarakter fra videregående skole likevel en stor innvirkning på studentenes resultater.³

Gjennomsnittlig karakterpoeng fra videregående skole for studentene som tok vårens eksamen, var 41,5. I figur 2.3.1 og 2.3.2 viser vi gjennomsnittskarakteren og karakterfordelingen for studenter som lå over og under dette gjennomsnittet.

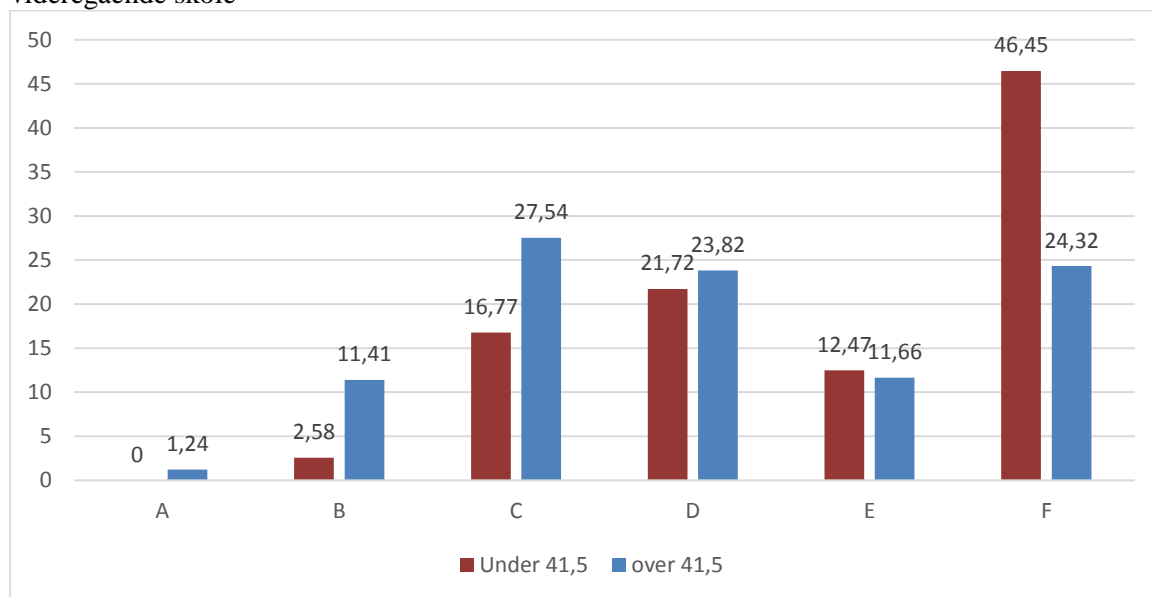
Figur 2.3.1 Gjennomsnittskarakter på nasjonal deleksamen våren 2016 basert på karakterpoeng fra videregående skole⁴



³ I denne delen av analysen forutsetter vi at det ikke er noen systematisk skjevhet i hvilke studenter som har nedprioritert den nasjonale deleksamenen. Vi tror ikke at studenter med svakere forkunnskaper har vært mer tilbøyelige til å nedprioritere vårens eksamen.

⁴ Vi har kun karakterpoeng fra videregående skole for 868 av 997 studenter. 465 studenter ligger under 41,5 karakterpoeng, og 403 studenter ligger over. Vi har mottatt bakgrunnsdata på individnivå fra FSAT. Analysene i denne rapporten er basert på det første uttrekket. Etter hvert som vi oppdager feil i datagrunnlaget, vil FSAT kjøre uttrekket igjen, og vi vil motta data for flere individer.

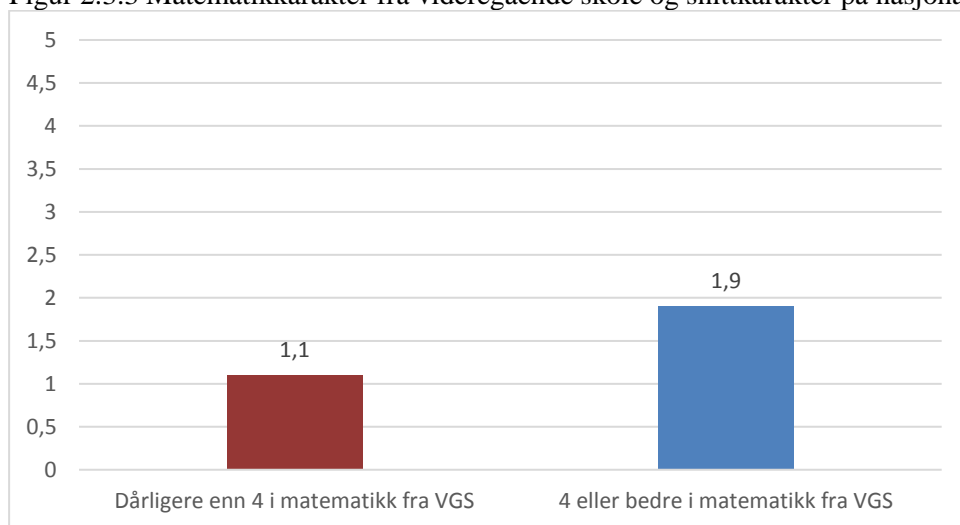
Figur 2.3.2 Karakterfordeling på nasjonal deleksamen våren 2016 basert på karakterpoeng fra videregående skole



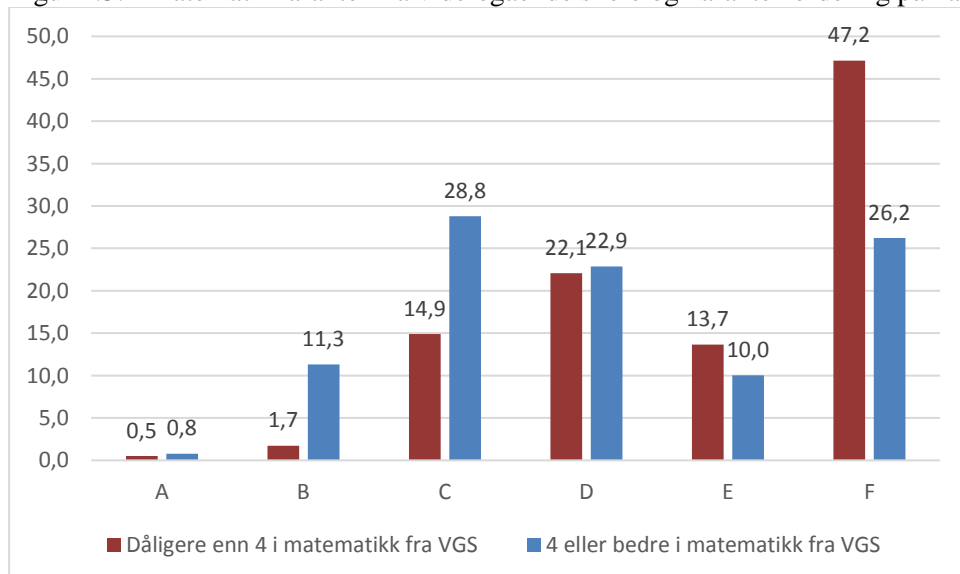
Som vi ser av figur 3.3.1 og 3.3.2, presterer de som har over 41,5 karakterpoeng fra videregående skole, bedre enn de som har færre karakterpoeng fra videregående skole. For studenter med mer enn 41,5 karakterpoeng er gjennomsnittskarakteren på nasjonal deleksamen 1,9 (D), mens for studenter med mindre enn 41,5 karakterpoeng er gjennomsnittskarakteren på nasjonal deleksamen 1,2 (E). Denne forskjellen er også tydelig på karakterfordelingen. Av studentene med mer enn 41,5 karakterpoeng får 40 prosent C eller bedre, mens 24,3 prosent strøk. Av studentene med mindre enn 41,5 karakterpoeng fikk kun 19,4 prosent C eller bedre, og hele 46,5 prosent strøk.

Dersom vi ser på studentenes matematikkarakter fra videregående skole, ser vi de samme effektene. NOKUTs datamateriale inkluderer studentenes gjennomsnittskarakter i alle matematikkfag de har tatt på videregående skole. I analysen under splitter vi studentene i de som har 4 eller bedre i snitt, og de som har dårligere enn 4 i snitt i matematikk fra videregående skole. Figur 2.3.3 viser gjennomsnittskarakteren på nasjonal deleksamen, og figur 2.3.4 viser karakterfordelingen på de to gruppene.

Figur 2.3.3 Matematikkarakter fra videregående skole og snittkarakter på nasjonal deleksamen⁵



Figur 2.3.4 Matematikkarakter fra videregående skole og karakterfordeling på nasjonal deleksamen



Som med karakterpoeng ser vi at studentene med 4 eller bedre i snitt i matematikk fra videregående skole gjorde det signifikant bedre enn studentene med dårligere enn 4 i snitt i matematikk. Snittkarakteren på den nasjonale deleksamenen var én karakter høyere, og karakterfordelingen viser også store forskjeller.

Effektene av karakterpoeng og matematikkarakter er statistisk og substansielt signifikante. I tabell 2 viser vi resultatene av en regresjonsanalyse. Den avhengige variabelen er antall poeng kandidatene fikk på den nasjonale deleksamenen (studentene kunne oppnå totalt 40 poeng), og de uavhengige variablene er kjønn, alder, GLU 1–7 eller GLU 5–10, karakterpoeng og hvilket studieår studentene var i da de tok nasjonal deleksamen.

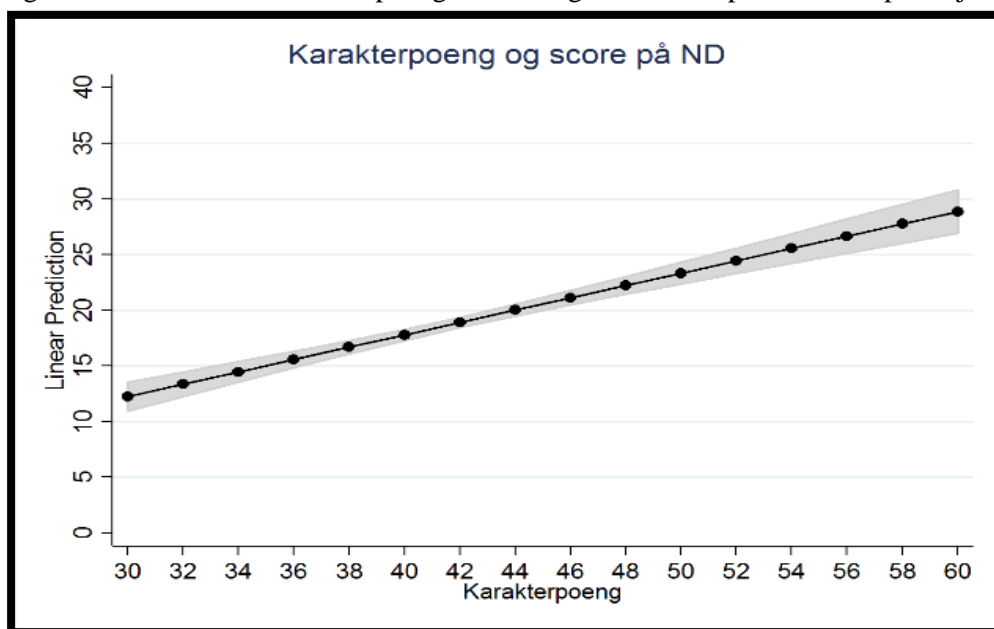
⁵ Vi har kun matematikkarakter for 792 av 997 kandidater. Av de 792 kandidatene hadde 403 under 4 i snitt i matematikk fra videregående skole, og 389 hadde over 4 i snitt i matematikk.

Tabell 2 Regresjonsanalyse nasjonal deleksamen. Resultater våren 2016

Avhengig variabel: Total score ND	
Karakterpoeng	0,55*** (0,05)
Kvinne	-2,67*** (0,64)
Alder	0,09 (0,05)
Kull	-0,77 (0,48)
GLU 5-10	3,83*** (0,75)
Constant	1547,82 (986,7)
N	773
***p > .01, **p > .05, *p > .1	

Regresjonsanalysen viser at karakterpoeng, kjønn og hvilken grunnskolelærerutdanning (1–7 eller 5–10) man tar, alle er signifikante variabler som kan forklare variasjonen i studentenes resultater på den nasjonale deleksamenen. For hvert karakterpoeng høyere en student har fra videregående skole, gir det 0,55 poeng mer i uttelling på den nasjonale deleksamenen. Kvinner scorer 2,67 poeng lavere enn menn, og GLU 5–10-studenter scorer 3,83 poeng høyere enn GLU 1–7-studenter. Figur 2.3.5 viser effekten av karakterpoeng på studentenes totalscore på den nasjonale deleksamenen.

Figur 2.3.5 Effekten av karakterpoeng fra videregående skole på totalscore på nasjonal deleksamen



Denne delen av analysen viser at studentenes startkompetanse forklarer mye av variasjonen på vårens nasjonale deleksamen, men ikke på resultatforskjellen mellom høsten og vårens eksamen. Studentene som tok eksamen våren 2016, hadde 41,5 karakterpoeng fra videregående skole, mens studentene som tok eksamen høsten 2015, hadde 41,9 karakterpoeng fra videregående skole. Forskjellen er liten og

ikke statistisk signifikant. Det er dermed ikke studentenes startkompetanse som forklarer resultatforskjellene på de to eksamenene.

Basert på de analysene vi har hatt mulighet til å gjøre, er det sannsynligvis to hovedgrunner til at resultatene våren 2016 var signifikant dårligere enn høsten 2015. Den første er at majoriteten av studentene som tok eksamen høsten 2015, går på GLU 5–10, mens majoriteten som tok eksamen våren 2016, går på GLU 1–7. Analysene over viser at GLU 5–10-studenter gjorde det signifikant bedre på vårens eksamen.

Den andre grunnen er at studentene ble gjort oppmerksomme på at den nasjonale deleksamenen, selv om den er obligatorisk, ikke er tellende. Vi tror derfor at en relativt stor andel studenter nedprioriterte den nasjonale deleksamenen. Dette gjør resultatene mindre reliable, og som nevnt tidligere kan man ikke stole på at resultatene gir et korrekt bilde av studentenes kunnskapsnivå.

Analysene over er likevel interessante. Dersom vi forutsetter at det ikke er en systematisk skjevhet i hvilke studenter som nedprioriterte vårens eksamen, ser vi at det å ha 4 i matematikk fra videregående skole har en signifikant effekt på studentenes resultater på den nasjonale deleksamenen. Det samme gjelder karakterpoeng fra videregående skole. Studentenes startkompetanse har med andre ord stor betydning for hvor godt de presterer på nasjonal deleksamen. Dette så vi også ved høstens eksamen, og det er ikke overraskende når vi vet at eksamenen avholdes i studentenes første studieår.

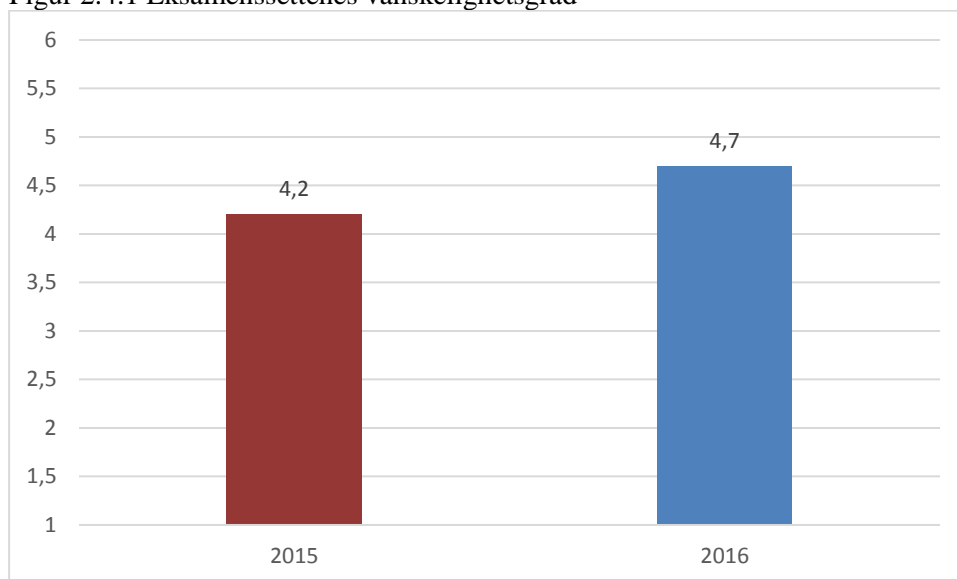
2.4 Eksamensoppgaver og sensorveiledning

Det finnes en mulighet for at eksamensoppgavene våren 2016 var vanskeligere enn eksamensoppgavene høsten 2015. For å vurdere dette har vi gjennomført en såkalt Rasch-analyse.⁶ Vi utførte samme analyse på eksamensoppgavene for høsten 2015, og den analysen konkluderte med at oppgavesettet holdt høy kvalitet, men at det var litt for mange enkle oppgaver. Analysen fra 2015 viste også at det på noen oppgaver var slik at kandidater fikk ett poeng uavhengig av deres dyktighet. Analysen for oppgavesettet våren 2016 viser at også dette oppgavesettet holdt høy kvalitet, men at noen få oppgaver var så vanskelige at nesten ingen studenter greide å løse disse korrekt. En årsak til dette kan være at oppgavene var kvalitativt vanskeligere, men det kan også være fordi studentene nedprioriterte eksamen. Analysen fant derimot ingen indikasjoner på at kandidatene som fikk ett poeng på en oppgave, fikk det uavhengig av deres dyktighet.

I tillegg til Rasch-analysen har vi også bedt sensorene om å vurdere vanskelighetsgraden på vårens eksamen, samt å sammenligne vanskelighetsgraden på de to eksamenene som har vært holdt. Av 48 sensorer har 44 av dem sensurert oppgaver både i 2015 og 2016. Høsten 2015 svarte 39 av 48 sensorer på undersøkelsen, mens 27 av 48 sensorer har svart på vårens undersøkelse.

⁶ Analysen er utført av Rolf Vegar Olsen ved Centre for Educational Measurement (CEMO), UiO.

Figur 2.4.1 Eksamenssettenes vanskelighetsgrad



Vi ser altså at sensorene mener at oppgavene var noe vanskeligere i 2016 når vi sammenligner svar fra de to undersøkelsene. Vi spurte også sensorene om de syntes vårens oppgavesett var lettere, like vanskelig eller vanskeligere enn høstens oppgavesett, og her var gjennomsnittsscoren 3,4 på en skala fra 1–5 (der 1 var «mye lettere i 2016» og 5 var «mye vanskeligere enn i 2015»). Begge de to spørsmålene indikerer altså at oppgavene kan ha vært noe vanskeligere i 2016 enn i 2015.

En annen mulighet er at sensorveiledningen i 2015 gjorde det relativt enkelt å oppnå ett poeng på enkeltoppgaver (se diskusjon over). Vi spurte derfor vårens sensorer om det var enklere eller vanskeligere å skille mellom null og ett poeng på vårens eksamen enn på høstens eksamen. Her svarte 67 prosent av sensorene at det var «mye enklere å skille mellom null og ett poeng på vårens eksamen». Gjennomsnittsscoren på spørsmålet var 1,4 (der 1 var «mye enklere»). Det er altså klart at sensorene mener det var vanskeligere for studentene å få ett poeng på enkeltoppgaver på vårens eksamen enn på høstens eksamen.

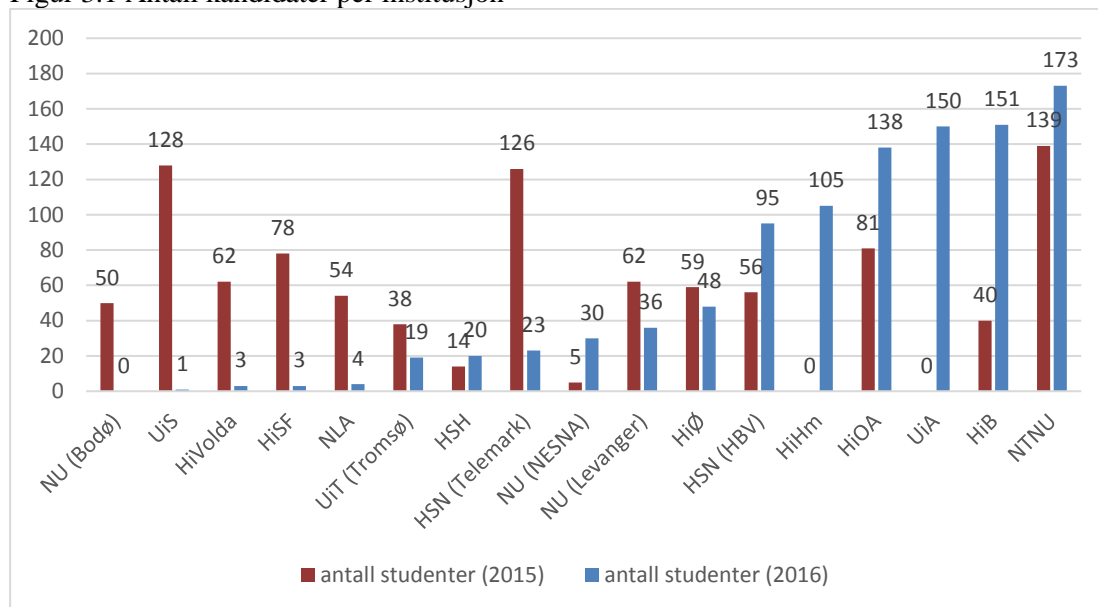
Analysen av oppgavesettet og sensorveiledningen viser altså at vårens oppgaver var noe vanskeligere, og at sensorveiledningen gjorde det noe vanskeligere å oppnå ett poeng på enkeltoppgaver. Når vi legger dette til de andre forklaringene, begynner vi å få en helhetlig forklaring på hvorfor årets eksamensresultater var såpass mye dårligere enn i 2015.

3 Institusjonsresultater

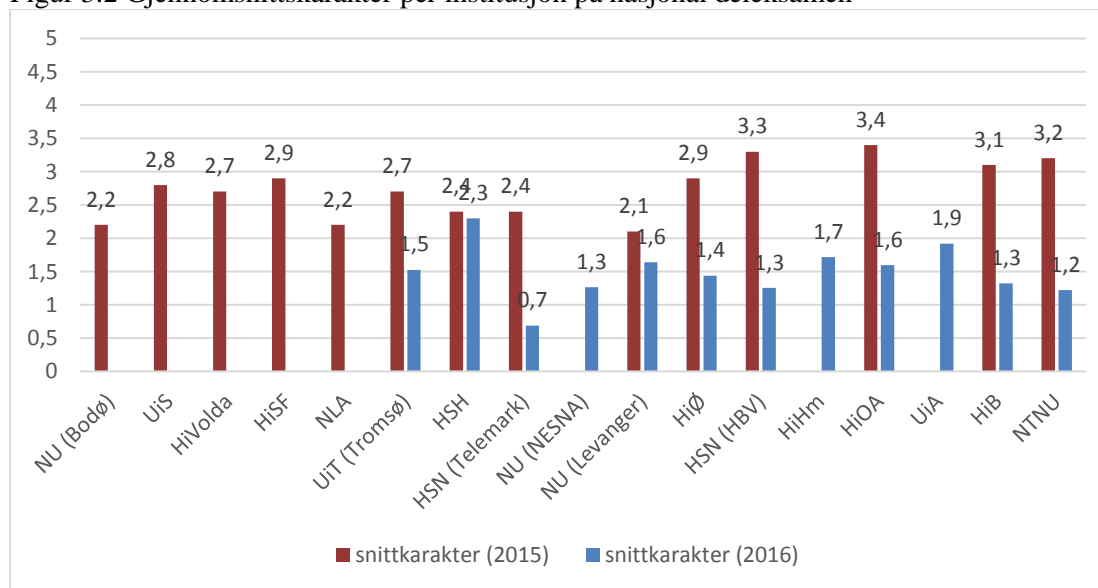
Når vi nå viser resultatene for hver institusjon, vil vi minne om at tallene må leses med forsiktighet så lenge vi ikke kan kontrollere for hvilke studenter som har nedprioritert denne eksamenen. Vi velger likevel å vise resultatene, slik at sektoren selv kan se de store forskjellene. Den første figuren viser antallet studenter på hver institusjon, og den andre viser gjennomsnittskaracteren på den nasjonale deleksamenen for høsten 2015 og våren 2016. Tabellen under viser karakterfordelingen per institusjon. Vi har slått sammen alle kandidatene for hver institusjon, selv om noen går på GLU 1–7 og andre på GLU 5–10. Grunnen til dette er at det ved noen institusjoner er så få studenter ved det ene

studieprogrammet at vi ikke kan sikre studentenes personvern hvis vi skiller de to gruppene på institusjonsnivå.

Figur 3.1 Antall kandidater per institusjon



Figur 3.2 Gjennomsnittskaracter per institusjon på nasjonal deleksamen



Tabell 3 Karakterfordeling per institusjon 2016 (prosent)

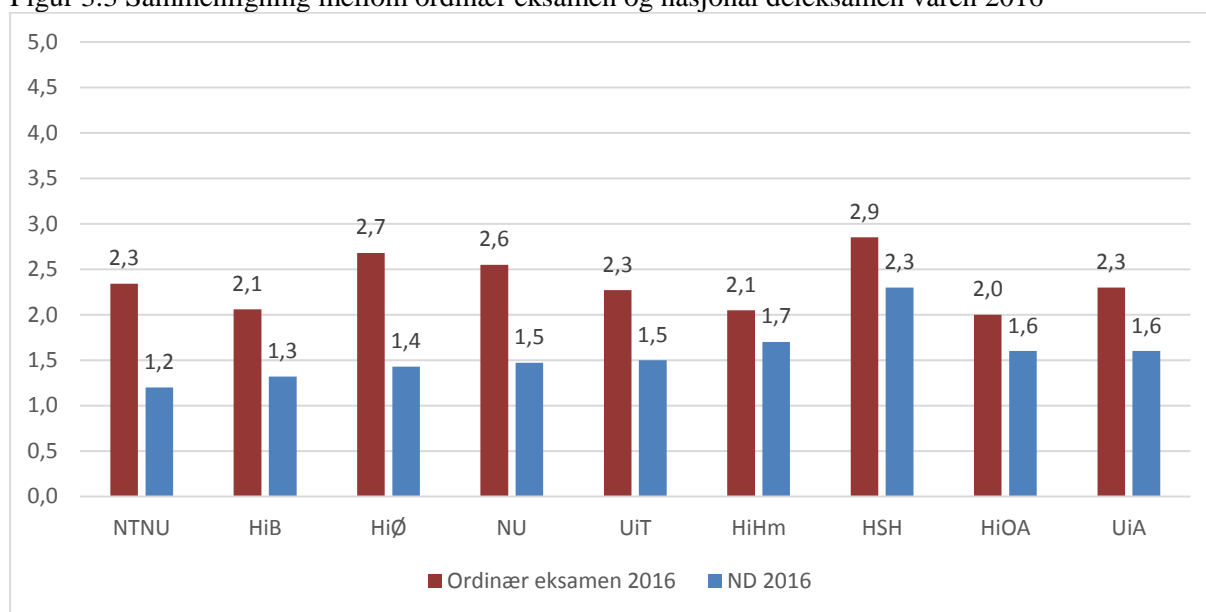
Karakterfordeling 2016						
Institusjon	A	B	C	D	E	F
HiB	0	6,0	13,9	25,2	16,6	38,4
HiHm	0	9,5	22,9	28,6	7,6	31,4
HiOA	0	7,2	26,1	20,3	11,6	34,8
HiSF	For få kandidater					

HiVolda	For få kandidater					
HiØ	0	2,1	29,2	20,8	6,3	41,7
HSH	0	10,0	45,0	25,0	5,0	15,0
HSN (Drammen)	0	0	3,2	12,9	16,1	67,7
HSN (Sandvika)	0	4,2	16,7	29,2	20,8	29,2
HSN (Telemark)	0	0	6,3	12,5	25,0	56,3
HSN (Vestfold)	0	4,9	26,8	24,4	19,5	24,4
NLA	For få kandidater					
NTNU	0	4,0	15,0	24,3	12,1	44,5
NU (Levanger)	0	5,6	27,8	22,2	13,9	30,6
NU (NESNA)	3,3	0	16,7	16,7	26,7	36,7
UiA	3,3	11,3	24,7	24,0	8,0	28,7
UiS	For få kandidater					
UiT (Tromsø)	0	0	31,6	26,3	5,3	36,8

3.1 Ordinær vs. nasjonal deleksamen våren 2016

De aller fleste studenter som avla den nasjonale deleksamenen våren 2016, avla også en ordinær matematikkeksamen denne våren. I delrapport 1 fra mars 2016 sammenlignet vi resultatene for de studentene som hadde tatt ordinær eksamen, med resultatene fra nasjonal deleksamen høsten 2015. Den analysen viste at lokale eksamener og sensur holdt et meget godt nivå, og at det ikke fantes noen grunn til å tro at institusjonene var «snille» med eksamensoppgaver eller sensur. Dersom våre antagelser stemmer (se over) om at studentene har nedprioritert den nasjonale deleksamenen fordi den ikke er tellende, bør vi i vårens sammenligning finne at studentene presterer bedre på lokale ordinære og tellende eksamener enn på den nasjonale deleksamenen. I figur 3.3 viser vi gjennomsnittresultatet for de to eksamenene fra våren 2016. Som man ser, er det relativt store forskjeller i gjennomsnittskaracteren for de fleste institusjonene. Ved NTNU, Høgskolen i Bergen, Høgskolen Stord/Haugesund og Nord universitet er gjennomsnittskaracteren én bokstavkaracter høyere på ordinær eksamen, mens ved Høgskolen i Østfold er forskjellen to bokstavkaracterer. Ved Høgskolen i Hedmark, Universitetet i Agder og Høgskolen i Oslo og Akershus presterte også studentene bedre på den ordinære eksamenen, men forskjellen utgjorde mindre enn én bokstavkaracter.

Figur 3.3 Sammenligning mellom ordinær eksamen og nasjonal deleksamen våren 2016⁷



Sammenligningen styrker vår antagelse om at studentene har nedprioritert den nasjonale deleksamenen til fordel for den ordinære eksamenen, og den styrker også NOKUTs anbefaling om at den nasjonale deleksamenen i matematikk for grunnskolelærerutdanningen må bli tellende for studentene.

4 Sensurreliabilitet

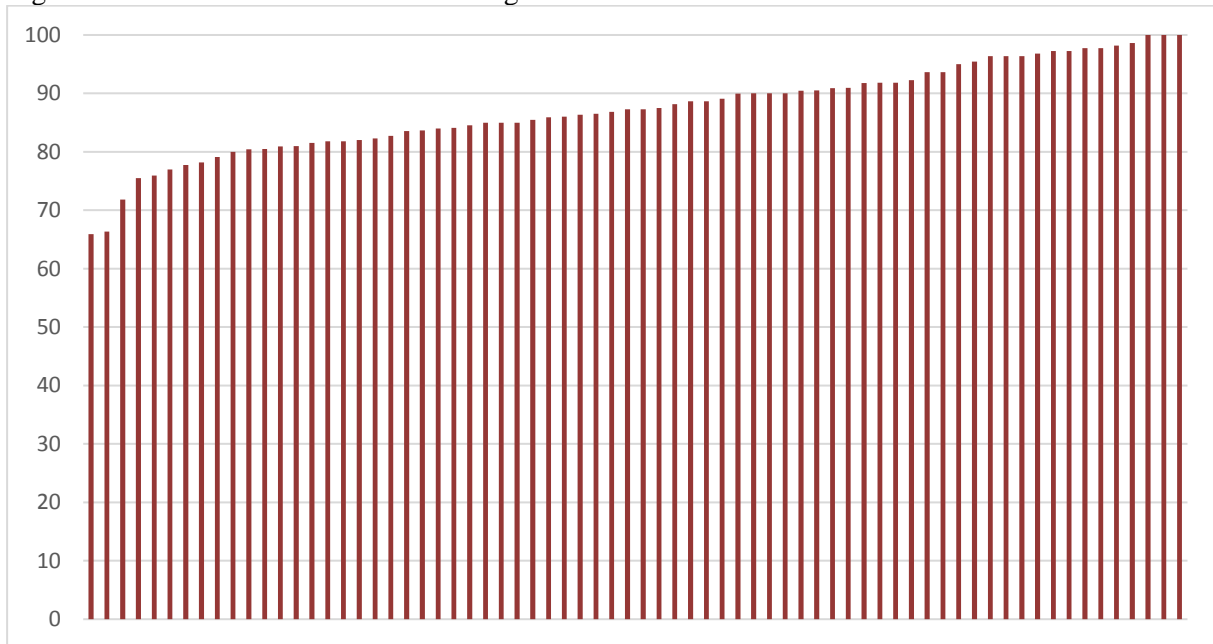
Før vi konkluderer i denne delrapporten, vil vi si noe om sensurreliabiliteten. I samarbeid med eksamensgruppen ba NOKUT alle sensorer om å rapportere detaljert sensur. Vi ba hver sensor om å oppgi hvor mange poeng hver kandidat fikk på hver enkelt eksamensoppgave. Sensorene fikk et poengskjema hvor de skulle fylle inn kandidatnummer og 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 totalscoren til hver enkelt kandidat. Sensorene ble så enige seg imellom om hvor mange poeng kandidaten skulle få på hver oppgave og på totalscoren.⁸ På denne måten har NOKUT mulighet til å vurdere sensurreliabiliteten.

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 eksamensoppgaver. Til sammen var det 92 sensorpar. Hvert eksamenssett inneholdt 20 spørsmål, så i gjennomsnitt sensurerte hvert sensorpar 220 enkeltoppgaver sammen. Figur 4.1 viser hvor mange prosent av de 220 oppgavene hvert sensorpar hadde vurdert likt.

⁷ Ved Universitetet i Stavanger, NLA Høgskolen, Høgskulen i Volda og Høgskulen i Sogn og Fjordane tok alle kandidatene den nasjonale deleksamenen som utsatt prøve. Ved Høgskolen i Oslo og Akershus er matematikkfaget delt opp over to år. Kandidatene som tok den nasjonale deleksamenen våren 2016, avla sin ordinære eksamen våren 2015. Ved Høgskolen i Sørøst-Norge skal studentene ha skriftlig eksamen i matematikk høsten 2016, og vi har dermed ikke noe sammenligningsgrunnlag.

⁸ Alle sensorene rapporterte ikke detaljert sensur til NOKUT, vi viser her resultatene fra de sensorene som rapporterte til NOKUT.

Figur 4.1 Sensurreliabilitet: Prosentvis enighet

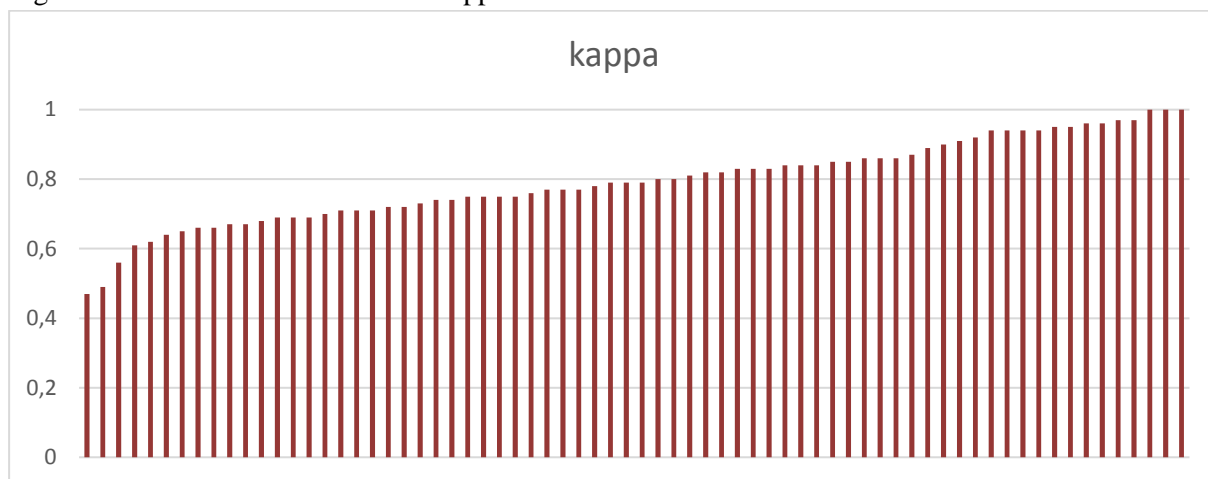


Hver stolpe i figuren representerer ett sensorpar, og høyden på stolpen viser hvor mange prosent av oppgavene sensorene hadde vurdert likt.

Ifølge forskningslitteraturen innenfor statistikk og reliabilitetsmålinger er «prosentvis 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.⁹ 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 viser vi kappa-koeffisienten per sensorpar.

⁹ Se for eksempel, Krippendorff, 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.

Figur 4.2 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 70 sensorparene vi har data for, har 67 sensorpar (96 prosent) en kappa-koeffisient på over 0,60. Hvor godt er dette? I forskningslitteraturen opererer man med følgende skala:¹⁰

- 0,0–0,20 svak enighet
- 0,21–0,40 rimelig enighet
- 0,41–0,60 moderat enighet
- 0,61–0,90 betydelig enighet
- 0,91–1,0 nesten perfekt enighet

Resultatene er med andre ord meget gode, og noe bedre enn resultatene fra høsten 2015, hvor 83 prosent av sensorparene hadde en kappa-koeffisient på over 0,60. At sensurreliabiliteten er enda bedre, er ikke overraskende med tanke på at eksamensgruppen har hatt mer tid til å lage gode oppgaver og sensorveiledning, samt at sensorene har samsensurert med hverandre to ganger. 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.

5 Konklusjon

Resultatene på den nasjonale deleksamenen i matematikk for grunnskolelærerutdanningen våren 2016 var svake uansett om man ser på nasjonalt nivå eller på institusjonsnivå. Resultatene var også mye svakere enn resultatene på den nasjonale deleksamenen som ble avholdt i desember 2015. Hovedfokuset i denne delrapporten har vært å belyse hvilke faktorer som kan forklare hvorfor årets resultater var så mye svakere enn fjorårets resultater.

Analysene viser at det ikke finnes én enkeltfaktor som kan forklare de svake resultatene våren 2016. Årsaken er etter alt å dømme en kombinasjon av tre faktorer, og vi kan på nåværende tidspunkt ikke vurdere effekten av enkeltfaktorene.

¹⁰ Se for eksempel Landis, J. & Koch, G. (1977). «The Measurement of Observer Agreement for Categorical Data» Biometrics 33: 159–174.

Først og fremst viser analysen at et stort antall studenter med høy sannsynlighet har nedprioritert den nasjonale deleksamenen til fordel for ordinære lokale eksamener i matematikk. Dette er helt sentralt, og dersom man skal fortsette med nasjonale deleksamener, viser årets eksamen i matematikk at disse eksamenene bør være tellende og ikke kun obligatoriske. En slik endring vil gjøre det meningsfullt for studentene både å forberede seg til eksamen og å yte sitt beste, og det vil gi sektoren som helhet robuste resultater som de kan bruke i sitt utviklingsarbeid.

Den andre faktoren er at det var flere studenter fra GLU 1–7 som avla vårens eksamen enn høstens eksamen, og at GLU 5–10-studenter, som selv har valgt matematikk, gjør det bedre enn studenter på GLU 1–7, som har matematikk som et obligatorisk emne og dermed ikke kan velge det bort.

Den tredje og siste faktoren er oppgavesettets vanskelighetsgrad og sensorveiledningen. Rasch-analysene fra de to eksamenene viser at det var noen flere enklere oppgaver i 2015 enn i 2016, og at det var enklere for studentene å oppnå ett poeng på enkeltoppgaver. Undersøkelsene blant sensorene støtter funnene fra Rasch-analysene. Det er med andre ord sannsynlig at oppgavesettet var noe vanskeligere, og at sensorveiledningen bidro til at det var vanskeligere å oppnå ett poeng på noen av enkeltoppgavene.

I tillegg til å forklare resultatforskjellene har vi også vist at studentenes startkompetanse (målt i karakterpoeng fra videregående skole) har stor innvirkning på resultatene. Det samme gjelder matematikkarakteren fra videregående skole. Studenter med 4 eller bedre i matematikk gjør det systematisk bedre enn studenter med dårligere enn 4 i matematikk.

NOKUT har nå gjennomført nasjonal deleksamen i matematikk for grunnskolelærerutdanningene to ganger. Gjennomføringen av eksamenen og sensuren har fungert meget godt, og i prinsippet tror NOKUT at nasjonale deleksamener kan fungere som et godt virkemiddel i sektorens kvalitetsarbeid. Men for at det skal kunne fungere, må NOKUT, Kunnskapsdepartementet og institusjonene ha tillit til kandidatenes resultater, og som denne rapporten tydelig viser, kan vi ikke ha det uten at eksamensresultatene blir tellende.