NOKUTs oppsummeringer

Nasjonal deleksamen i matematikk for grunnskolelærerutdanningen høsten 2016

Mars 2017

NOKUT

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

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

Tittel:	Nasjonal deleksamen i matematikk for grunnskolelærerutdanningen
Forfattere:	Kristina C. Skåtun Kjersti Tokstad (prosjektleder)
Dato:	15. mars 2017

Sammendrag

1. desember 2016 arrangerte NOKUT i samarbeid med Nasjonalt råd for lærerutdanning (NRLU) og 12 utdanningsinstitusjoner den tredje nasjonale deleksamenen i matematikk for grunnskolelærerutdanning 1–7 (GLU 1–7) og grunnskolelærerutdanning 5–10 (GLU 5–10). Totalt 1059 studenter gjennomførte eksamen og fikk sensur. Eksamenen omfattet alle GLU-studenter som fulgte undervisning i høstsemesteret hvor eksamenstemaet undervisningskunnskap i brøk, desimaltall og prosentregning inngikk. I tillegg er det noen studenter som har tatt denne eksamenen som en kontinuasjonseksamen. Denne rapporten må ses i sammenheng med den første delrapporten, som omhandlet den nasjonale deleksamenen som ble avholdt 1. desember 2015, og den andre delrapporten som omhandlet deleksamenen som ble avholdt 11. mai 2016.

Resultatene for høstens nasjonale deleksamen var mye bedre enn for våren 2016, og tilnærmet like resultatene for høsten 2015. Høsten 2016 strøk kun 9,2 prosent av kandidatene, og 24,4 prosent fikk A eller B. Dette er tilnærmet likt resultatene for høsten 2015, hvor andelen kandidater som fikk A eller B var 28,3 prosent, og andelen som strøk var 10,4 prosent. Våren 2016 strøk hele 37 prosent av kandidatene, og kun 6,6 prosent fikk karakteren A eller B. Årsakene til at resultatene høsten 2016 var bedre enn våren 2016 er sammensatte, og våre analyser peker i retning av fire hovedfaktorer eller en kombinasjon av disse.

Én faktor er at andelen studenter fra GLU 1–7 var langt høyere ved vårens eksamen enn ved eksamenene på høsten. 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.

En annen faktor er at et stort antall studenter med høy sannsynlighet har nedprioritert den nasjonale deleksamenen våren 2016 til fordel for ordinære eksamener i matematikk. Nytt fra høsten 2016 er at nasjonal deleksamen nå er tellende på vitnemålet, og det får dermed større konsekvenser om studentene ikke består eksamen. Dette har trolig økt studentenes innsats og resultert i bedre karakterer enn for våren 2016.

Den tredje faktoren er studentenes startkompetanse, målt i karakterpoeng fra videregående skole. Startkompetanse viste seg å forklare noe av variasjonen i resultatene fra eksamen, der studenter som har gjort det bra på videregående også gjør det bedre på nasjonal deleksamen, men effektene var ikke veldig sterke og forklarer heller ikke endringen i resultatene over tid.

Den siste faktoren er oppgavesettets vanskelighetsgrad og sensorveiledningen. Analysene fra de tre eksamenene viser at det var noen flere enklere oppgaver i 2015 enn våren 2016, og at eksamenen høsten 2016 var en anelse lettere enn våren 2016. Undersøkelsene blant sensorene er i samsvar med funnene fra Rasch-analysene. Det er med andre ord sannsynlig at oppgavesettet var noe vanskeligere våren 2016.

NOKUT har nå gjennomført nasjonal deleksamen i matematikk for grunnskolelærerutdanningene tre ganger, og gjennomføringen av både eksamenen og sensuren har fungert svært godt. Nasjonale deleksamener kan i prinsippet fungere som et godt virkemiddel i sektorens kvalitetsarbeid, og etter at eksamenen nå er tellende, gir nok resultatene et mer korrekt bilde av studentenes kunnskapsnivå.

Innhold

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 1. desember 2016. Den første nasjonale deleksamenen i matematikk i grunnskolelærerutdanningen ble avholdt 1. desember 2015, og i den første rapporten 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, men viser heller til den første rapporten. Den andre nasjonale deleksamen ble avholdt i mai 2016. Resultatene på denne eksamenen var vesentlig dårligere enn høsten 2015, og resultatene og årsakene til denne nedgangen omtales i delrapport 2.

Det som er nytt fra og med høsten 2016, er at denne eksamenen nå teller på studentenes vitnemål. Det gjorde den ikke tidligere, og det fikk ingen konsekvenser om studentene ikke bestod eksamen. Dette har trolig ført til at studentene ikke har forberedt seg tilstrekkelig våren 2016 og forklarer i stor grad de dårligere resultatene. I denne rapporten beskriver vi resultatene for nasjonal deleksamen høsten 2016, sammenligner disse med de tidligere deleksamenene og peker på mulige årsaker til endringen i karaktersnitt over tid.

1.1 Organisering av nasjonal deleksamen

Det er den samme gruppen som har utviklet eksamensoppgavene og sensorveiledningene til alle 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 eksamensgjennomføringene. Dette inkluderer bestilling og fordeling av eksamenspapir, kandidatnumre til eksamen, fordeling av besvarelser til sensorkorpset samt formidling av eksamensresultatene til institusjonene.

1.1.1 Eksamensdagen

Den tredje nasjonale deleksamenen i GLU ble avholdt som en firetimers skriftlig skoleeksamen uten bruk av hjelpemidler 1. desember 2016. Det var 12 institusjoner som deltok, fordelt på 17 studiesteder. Studentene som deltok, var i sitt første og tredje semester og kom fra både GLU 1–7 og 5–10. Utvalget av GLU 1–7 og/eller GLU 5–10 på de ulike institusjonene vises i tabell 1. Totalt gjennomførte 1057 av 1202 oppmeldte kandidater eksamen, som tilsvarer ca. 88 prosent av de oppmeldte kandidatene.

1.1.2 Sensur

For å sikre likebehandling og full anonymitet ble det satt sammen et nasjonalt sensorkorps, og disse bestod av 50 personer. Alle lærerutdanningsinstitusjonene var representert i sensorkorpset. Sensorkorpset ble satt sammen av eksamensgruppen og formelt oppnevnt av alle institusjonene som gjennomførte den nasjonale deleksamenen.

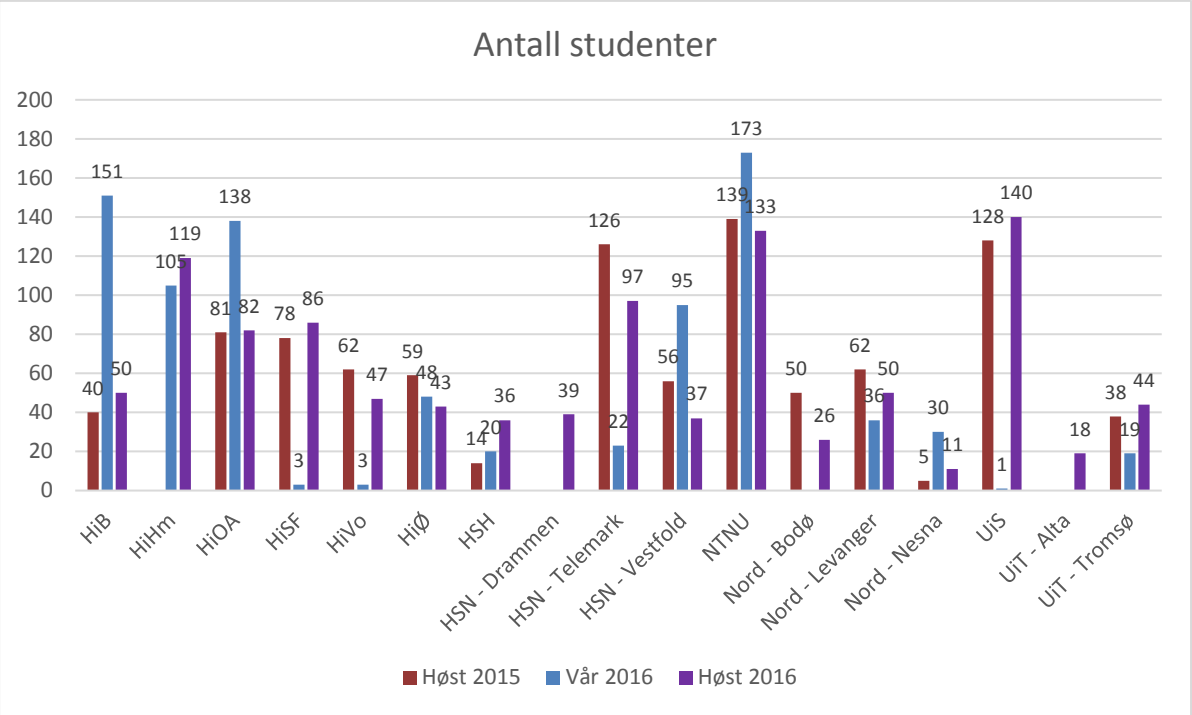
Hver besvarelse ble vurdert av to sensorer. For å få til en karakterkalibrering og en institusjonell spredning av besvarelsene delte NOKUT inn sensorene i til sammen 100 sensorpar. Hver sensor sensurerte sammen med tre eller fire andre. Hvert sensorpar sensurerte 10 eller 11 besvarelser, og hver enkelt sensor vurderte i gjennomsnitt 42 besvarelser.

1.2 Utvalget

Som i 2015 og i mai 2016 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, i mai 2016 og i desember 2016.

Tabell 1 Programmer som deltok fra hver institusjon

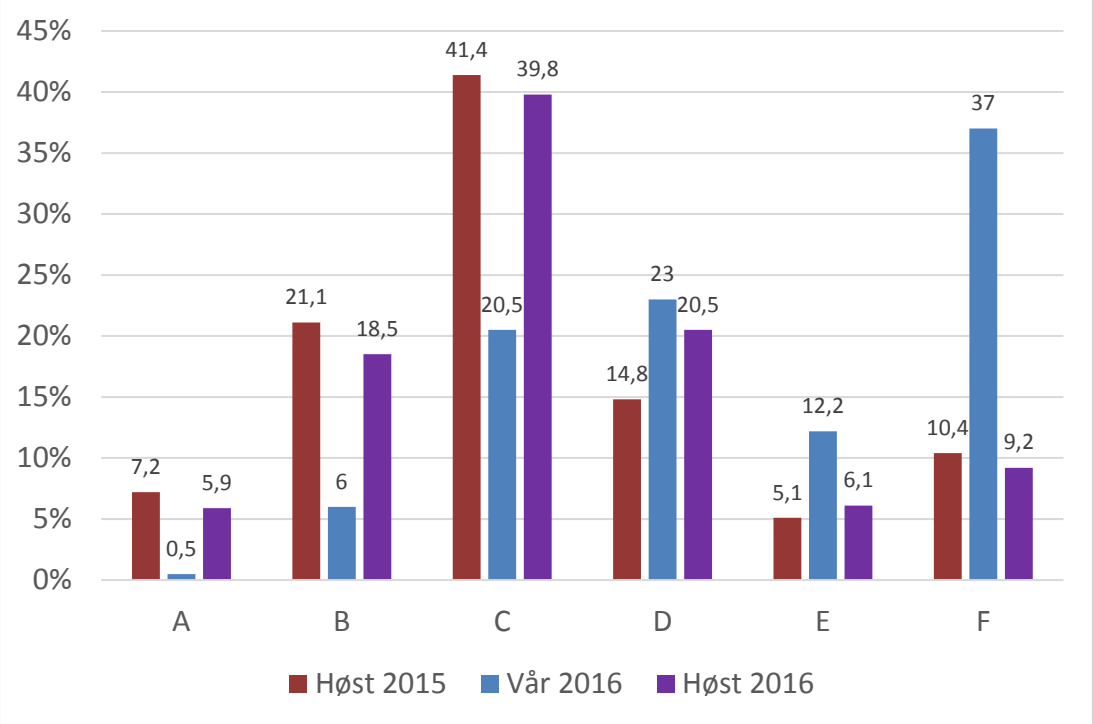
Institusjon	**Høst 2015**	**Vår 2016**	**Høst 2016**
HiB	GLU 5–10	GLU 1–7	GLU 5–10 og 1–7
HiHm	Ingen	GLU 5–10 og 1–7	GLU 1–7
HiOA	GLU 5–10	GLU 1–7	GLU 5–10 og 1–7
HiSF	GLU 5–10 og 1–7	Kun 3 studenter	GLU 5–10 og 1–7
HiVo	GLU 5–10	GLU 1–7 (kun 3 stud.)	GLU 5–10 og 1–7
HiØ	GLU 5–10	GLU 1–7	GLU 5–10
HSH	GLU 5–10	GLU 5–10	GLU 5–10 og 1–7
HSN (2016) / HBV	GLU 5–10	GLU 5–10 og 1–7	GLU 5–10
HSN (2016) / HiT (2015)	GLU 5–10 og 1–7	GLU 1–7 og 5–10 (kun 15 stud.)	GLU 5–10 og 1–7
NLA	GLU 5–10 og 1–7	Kun 9 studenter	Ingen
Nord (2016) / Bodø	GLU 5–10 og 1–7	Ingen	GLU 5–10
Nord (2016) / Levanger	GLU 1–7	GLU 5–10	GLU1–7
Nord (2016) / Nesna	GLU 5–10 (kun 5 stud.)	GLU 1–7	GLU1–7
NTNU (2016) / HiST (2015)	GLU 5–10	GLU 1–7	GLU 5–10
UiA	Ingen	GLU 5–10 og 1–7	Ingen
UiS	GLU 5–10 og 1–7	Kun 1 student	GLU 5–10 og 1–7
UiT	GLU 5–10 og 1–7	GLU1–7	GLU 5–10 og 1–7



Figur 1.1 Antall kandidater per institusjon for nasjonal deleksamen de tre siste semestrene.

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, våren 2016 og høsten 2016.



Figur 2.1 Karakterfordeling nasjonalt (prosent).

Som vi ser av figuren, var resultatene fra høsten 2016 markant bedre enn våren 2016, og kun noe svakere enn for høsten 2015. Kun 9,2 prosent av studentene strøk og 24,4 prosent fikk A eller B høsten 2016, sammenlignet med 37 prosent stryk og kun 6,5 prosent A eller B våren 2016. Den nasjonale gjennomsnittskarakteren høsten 2016 var C, som også var tilfellet høsten 2015. Våren var gjennomsnittskarakteren D, men den lå meget nær E. 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, 1,5 våren 2016 og 2,7 høsten 2016. Den reelle forskjellen er altså 1,2 karakterpoeng fra våren 2016 og tilnærmet lik for ett år siden.

Det er flere forskjellige forklaringer på de markant dårligere resultatene våren 2016. Disse har blitt beskrevet i delrapport 2 våren 2016, men hovedpoengene blir repetert her og satt i sammenheng med forbedringen i resultatene høsten 2016.

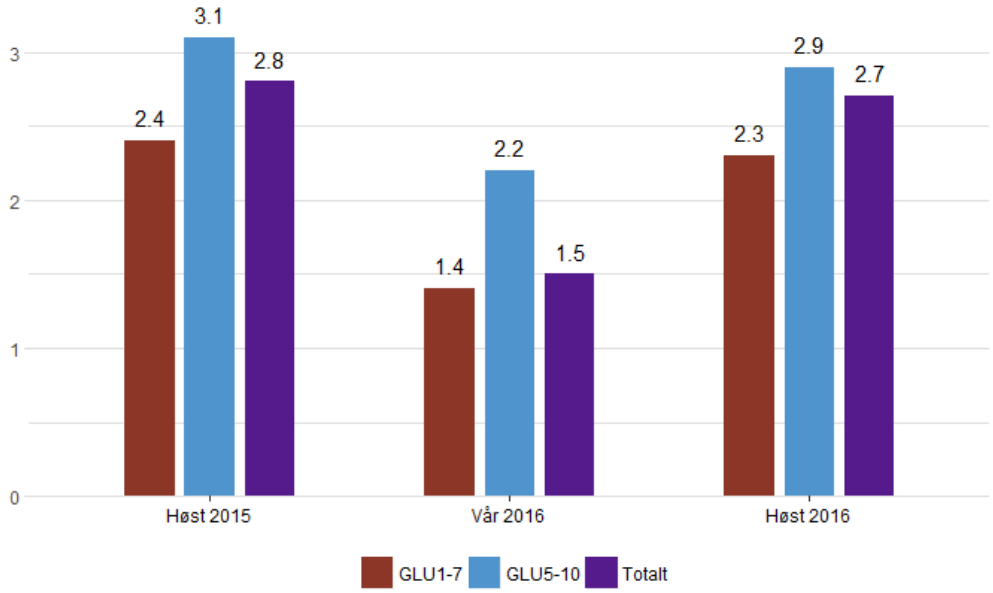
1. GLU 1–7-studenter har matematikk som obligatorisk fag, mens studentene fra GLU 5–10 som har tatt den nasjonale deleksamenen, har matematikk som valgfag. Det betyr at GLU 5–10-studenter sannsynligvis har en større interesse for matematikk og muligens bedre forkunnskaper (i gjennomsnitt) enn studenter fra GLU 1–7-programmer.
2. Nasjonal deleksamen er nå tellende, i motsetning til våren 2016. Dette har nok i stor grad økt studentenes innsats for å bestå eller få en god karakter på deleksamen.
3. Studentenes startkompetanse har noe effekt på studentenes karakterer på den nasjonale deleksamen.
4. Vanskelighetsgraden av eksamen. Denne vil naturlig nok variere litt mellom hver eksamen og vil kunne påvirke hvor godt studentene klarer å besvare oppgavene.

2.1 GLU 1–7 vs. GLU 5–10

En av årsakene til økningen i karaktersnitt høsten 2016 kan være at andelen GLU 1–7-studenter var mye høyere våren 2016 enn høsten 2016. Andelen av studentene som kom fra GLU 1–7 var 39 prosent for høsten 2016, 85 prosent for våren 2016 og 35 prosent for høsten 2015.[1]

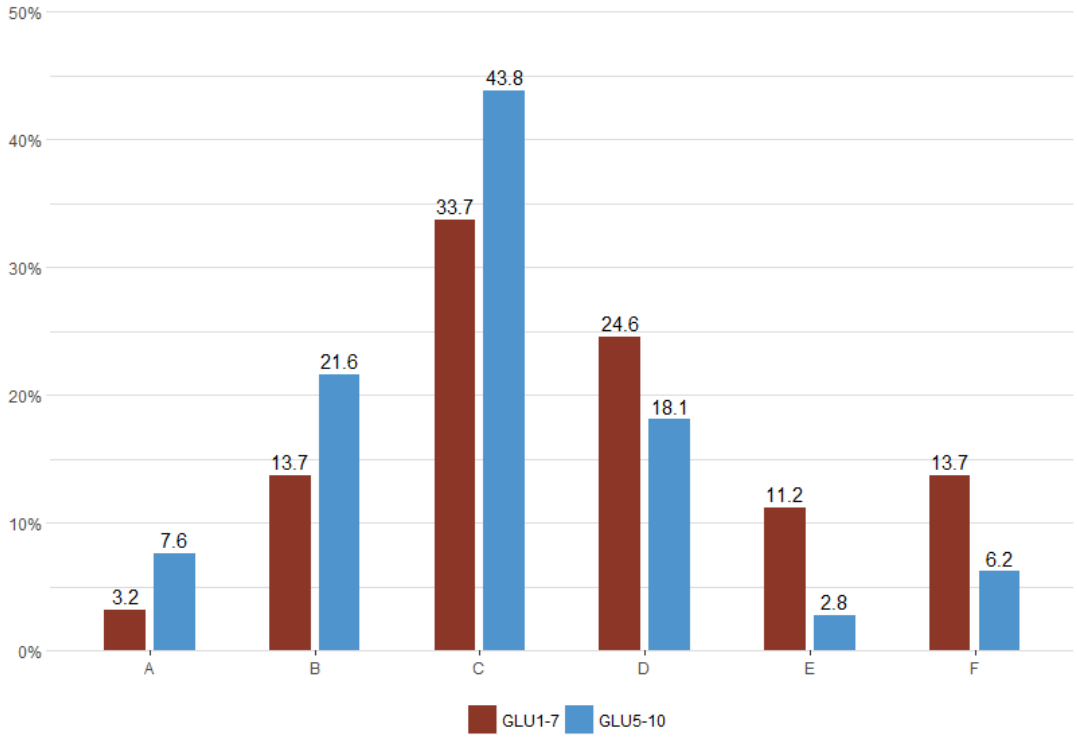
Resultatene fra siste deleksamen viser at GLU 5–10 gjør det signifikant bedre enn studentene fra GLU 1–7, i likhet med eksamen fra våren 2016. Figur 2.1.1 viser gjennomsnittskarakteren for GLU 1–7 og 5–10, samt totalen for alle studentene, for de tre siste semestrene.

[1] Vi har individdata for 1056 av 1057 av studentene høsten 2016, og for 982 av 997 studenter våren 2016. 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 2.2 Gjennomsnittskarakter for GLU 1–7, GLU 5–10, og totalt, fordelt på semester.

Av figuren ser vi at snittet for GLU 5–10 er høyere for alle semestrene, og at totalscoren ligger nærmere GLU 5–10 for eksamenene på høsten og nærmere GLU 1–7 på våren (gjennomsnittet trekkes mot GLU-typen med høyest antall studenter). For høsten 2016 er gjennomsnittskarakteren for GLU 5–10 2,9, som er 0,6 karakterpoeng bedre enn for GLU 1–7 (2,3), men dette utgjør i praksis en hel bokstavkarakter (GLU 1–7 har D i snitt, mens GLU 5–10 har C i snitt). Figur 2.1.2 viser karakterfordelingen for de to gruppene for høsten 2016.



Figur 2.3 Karakterfordeling GLU 1–7 vs. GLU 5–10 (høsten 2016).

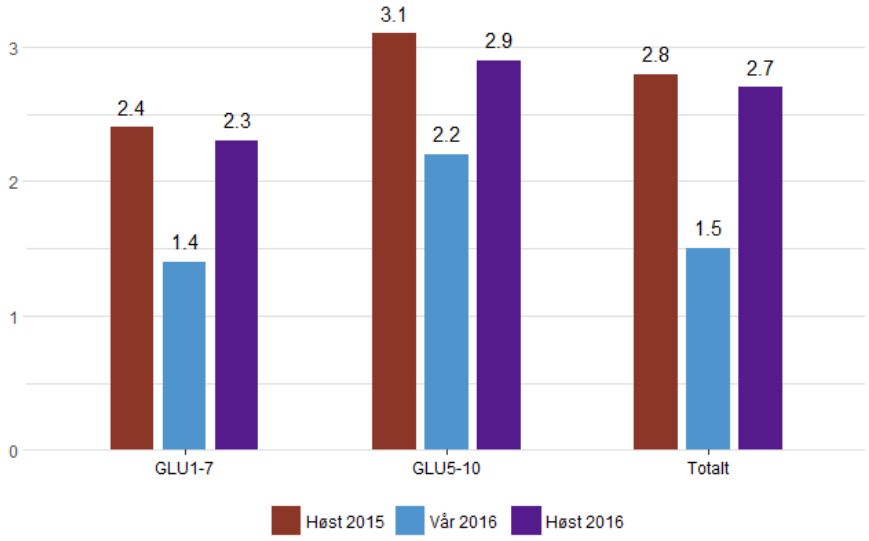
Vi ser også på karakterfordelingen at 73 prosent av GLU 5–10-studentene får C eller bedre, sammenlignet med 51 prosent hos GLU 1–7-studentene. Andelen stryk utgjør 6,2 prosent i GLU 5–10 og 13,7 prosent i GLU 1–7.

Siden studentene i GLU 5–10 får bedre karakterer og utgjør en større andel høsten 2016 enn våren 2016, kan dette forklare noe av økningen i gjennomsnittskarakter observert ved siste deleksamen. Årsaken til at GLU 5–10 gjør det bedre, kan være at disse studentene selv har valgt matematikk, mens det for GLU 1–7-studenter er obligatorisk. Motivasjonsnivået og forutsetningene for å gjøre det bra på denne eksamenen er nok derfor forskjellig mellom disse gruppene.

2.2 Obligatorisk vs. tellende resultat

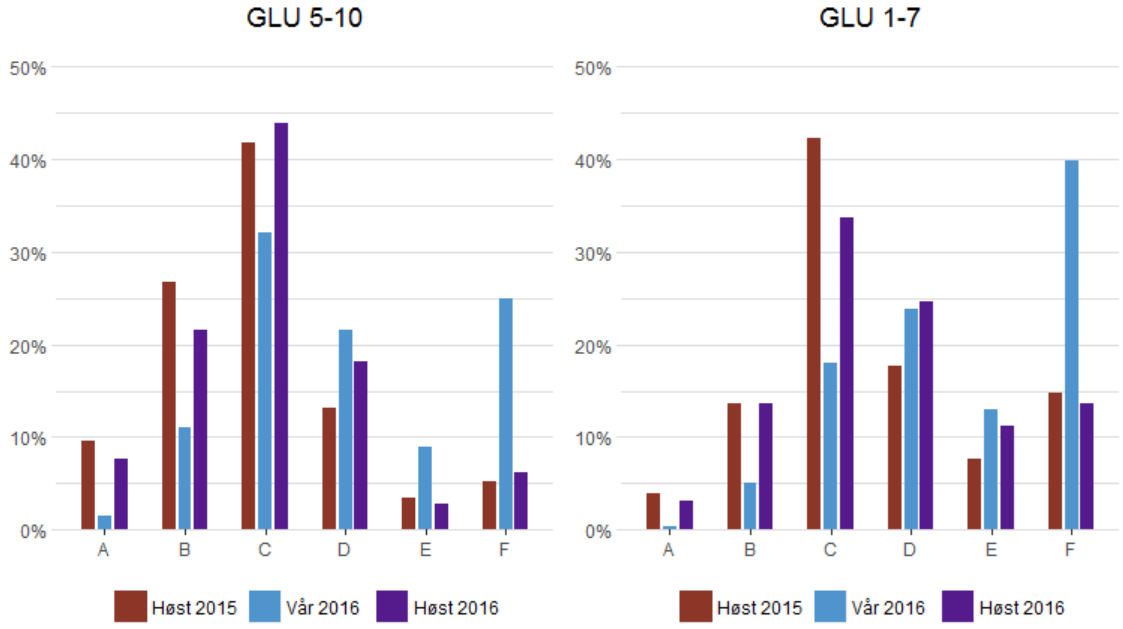
Den nasjonale deleksamenen i matematikk for grunnskolelærerutdanningene er en isolert eksamen som er obligatorisk. Hovedforskjellen fra høsten 2016 var at nasjonal deleksamen nå er tellende på vitnemålet, mens våren 2016 ville studentene få utstedt vitnemål selv om de strøk på eksamen. Resultatet var at studentene kunne nedprioritere den nasjonale deleksamenen uten at dette fikk særlige konsekvenser. Da den første deleksamenen ble avholdt i desember 2015, var dette lite kjent, noe som kan forklare hvorfor resultatene var mye bedre enn våren 2016.

Selv om vi ser at gjennomsnittskarakteren til GLU 5–10 er bedre enn GLU 1–7 på hver enkelt deleksamen, så ser vi også at både GLU 1–7 og 5–10 gjorde det vesentlig mye dårligere våren 2016 enn på høsteksamenene (figur 2.2.1). Fra figuren ser vi at GLU 1–7 scorer henholdsvis 2,4 og 2,3 høsten 2015 og høsten 2016 og kun 1,4 våren 2016, mens GLU 5–10 scorer henholdsvis 2,9 og 3,1 i gjennomsnitt på høsteksamenene og kun 2,2 på vårens eksamen.



Figur 2.4 Gjennomsnittskarakter for GLU 1–7, GLU 5–10 og totalt.

Hvis vi også ser på karakterfordelingen, har vårens eksamen en lavere andel A og B og en høyere andel stryk når vi sammenligner GLU 5–10-studentene over tid (figur 2.2.2). For eksempel får henholdsvis 36 og 29 prosent av GLU 5–10-studentene A og B på høsteksamenene, sammenlignet med kun 12,5 prosent våren 2016. Tilsvarende utgjør også strykprosenten kun henholdsvis 5 og 6 prosent på høsteksamenene, mens 25 prosent av studentene strøk våren 2016. Det samme gjelder om vi sammenligner kun GLU 1–7 studentene, der rundt 17 prosent får A og B på høsteksamenene, mens det kun er 5,4 prosent som får A og B våren 2016. Andelen stryk var også 40 prosent våren 2016, sammenlignet med 15 og 14 prosent på høsteksamenene.



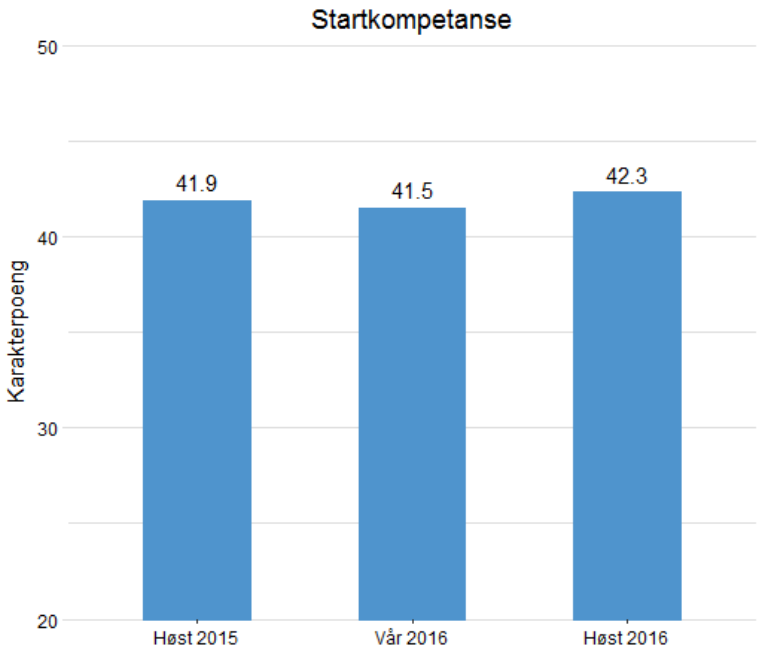
Figur 2.5 Karakterfordeling GLU 1–7 vs. GLU 5–10 de siste tre deleksamenene.

Hvis det kun var en høyere andel av GLU 5–10 som forklarer bedringen av resultatene på høstens eksamen, ville vi forventet å se like resultater for GLU 5–10 både på våren og høsten. Dette er ikke

tilfelle, noe som tyder på at forskjellen mellom eksamenene i stor grad skyldes at studentene har nedprioritert eksamenen i vårsemesteret. Resultatene illustrerer derfor at studentene får bedre resultater når de tror (høsten 2015) eller vet (høsten 2016) at eksamenen teller på vitnemålet, i motsetning til når eksamenen ikke teller (våren 2016). En tellende og obligatorisk eksamenen har trolig stor innvirkning på studentenes motivasjon og innsats, noe som igjen har resultert i en lavere strykprosent og et gjennomsnitt på en solid C.

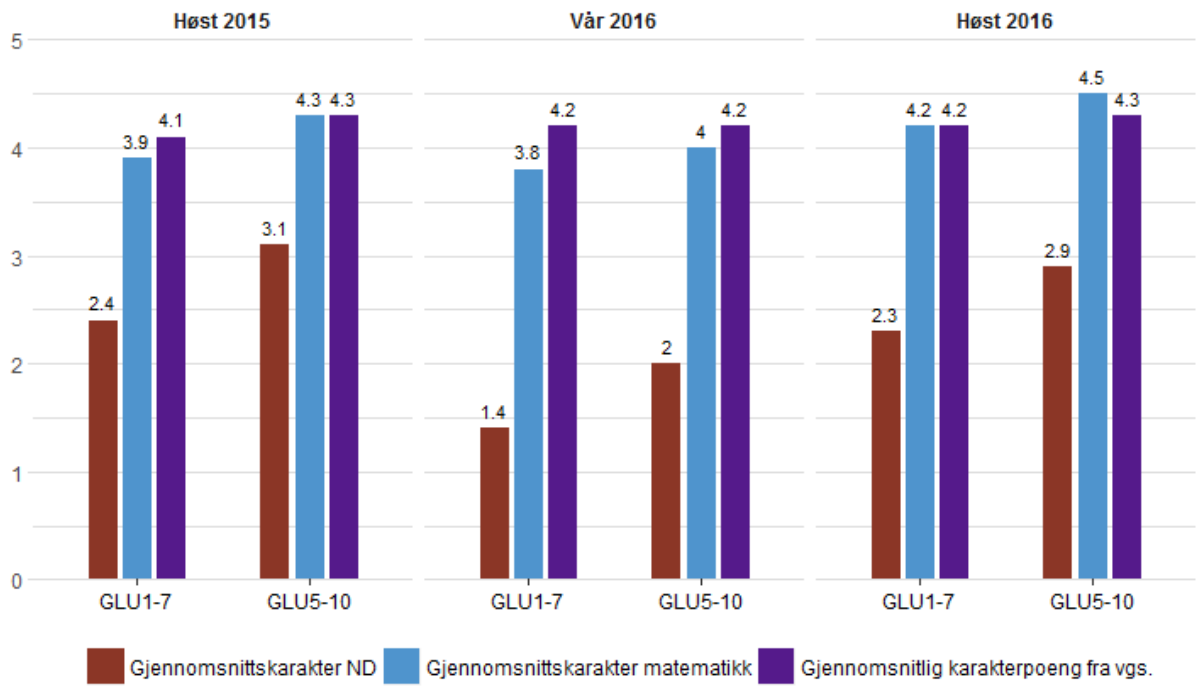
2.3 Karakterpoeng og matematikkarakter fra videregående skole

Studentenes startkompetanse, i form av karakterpoeng og matematikkarakter fra videregående skole, har også noe innvirkning på studentenes resultater. Hvis vi ser på startkompetansen til studentene for de tre siste deleksamenene, var karakterpoengene fra videregående nærmest uendret. I gjennomsnitt hadde studentene 42,4 karakterpoeng fra videregåendeskole høsten 2016, som er nesten identisk med gjennomsnittene fra våren 2016 (41,5) og høsten 2015 (41,9) (figur 2.3.1). Studentenes startkompetanse har dermed vært den samme for alle de nasjonale deleksamenene, noe som betyr at de dårligere resultatene våren 2016 ikke kan skyldes forskjeller i karakterpoeng.



Figur 2.6 Karakterpoeng for de tre siste nasjonale deleksamenene.

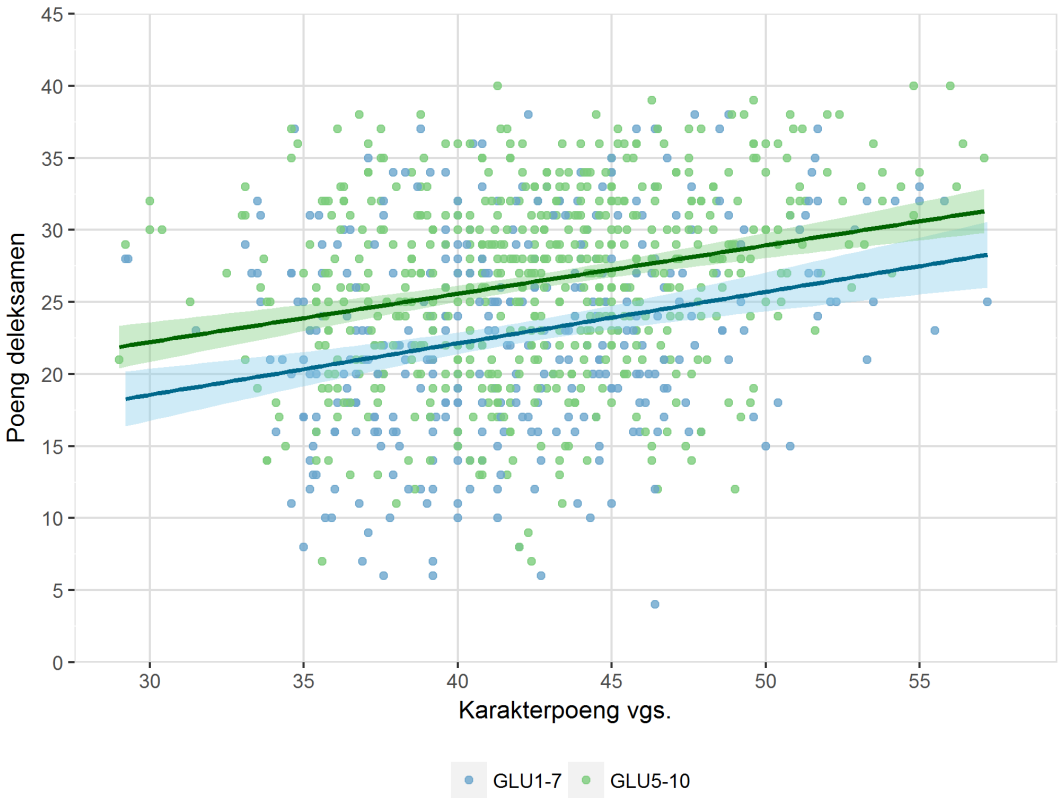
Hvis vi ser på sammenhengen mellom karakterer fra videregående skole og antall poeng på nasjonal deleksamen fordelt på GLU-type, ser vi at matematikkarakteren til studentene i GLU 5–10 er signifikant høyere enn for GLU 1–7 ($t=6,7$, $p<0,001$) (figur 2.3.2). Det er også en signifikant forskjell i karakterpoeng mellom gruppene ($t=2,7$, $p=0,006$), men denne er svakere enn for matematikkarakterene.



Figur 2.7 Karakterer fra nasjonal deleksamen og videregående skole fordelt på GLU-type over tid.

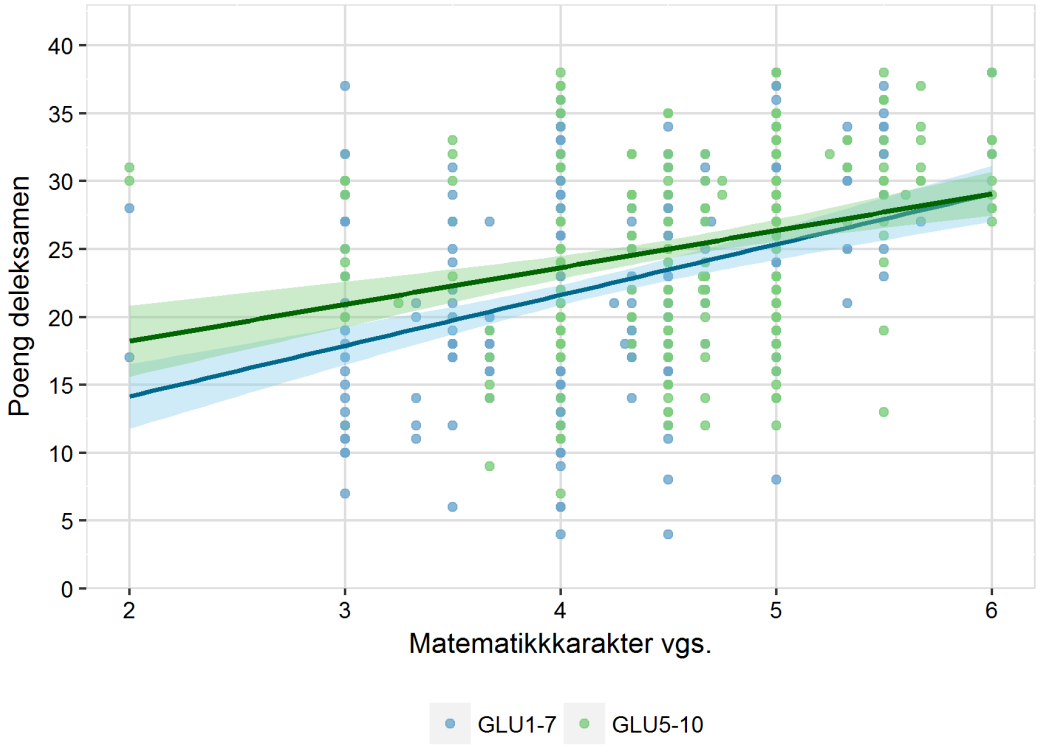
Hvis vi ser på korrelasjonen mellom karakterpoeng fra videregående skole[2] og antall poeng på nasjonal deleksamen høsten 2016, får vi en signifikant svak positiv korrelasjon ($r=0,27$, $p<0,001$). En regresjonsanalyse, der antall poeng på den nasjonale deleksamenen er avhengig variabel og karakterpoeng er uavhengig variabel, viser at karakterpoeng forklarer 7 prosent av variasjonen i resultatene. Analysen viser at for hvert karakterpoeng høyere en student har fra videregående skole, gir det 0,37 poeng mer i uttelling på den nasjonale deleksamenen. Dette betyr at de som har gjort det bra på videregående skole, også får noe høyere karakter på nasjonal deleksamen, men her er det nok mange andre faktorer som også forklarer variasjonen. Figur 2.3.3 viser effekten av karakterpoeng på studentenes totalscore på nasjonal deleksamen høsten 2016. Av figuren ser vi også at GLU 5–10 ligger jevnt høyere enn GLU 1–7.

[2] Vi har karakterpoeng fra videregående skole for 980 av 1057 studenter. Vi har mottatt bakgrunnsdata på individnivå fra FSAT.



Figur 2.8 Effekten av karakterpoeng fra videregående skole på totalscore på nasjonal deleksamen. Blå og grønne sirkler viser individuelle studenter fra henholdsvis GLU 1–7 og GLU 5–10, linjene er regresjonslinjer for hver GLU-type, og skygget område viser 95 prosent konfidensintervall.

Vi ønsket også å undersøke effekten av matematikkarakter fra videregående skole på antall poeng på deleksamen. Siden vi kun fikk oppgitt matematikkarakter for 638 av studentene, ble denne analysen kun gjort på dette utvalget. Matematikkarakter hadde en noe høyere korrelasjon ($r=0,37$, $p<0,001$) med deleksamenspoeng enn det karakterpoeng hadde, og forklarte 13 prosent av variasjonen. Analysen viser at for hver økning i matematikkarakter, så øker antall poeng på deleksamen med 3,5. Figur 2.3.4 viser effekten av matematikkarakter på studentenes totalscore på den nasjonale deleksamenen. Figuren viser også at GLU 5–10 ligger høyere enn GLU 1–7, men at antall poeng på deleksamen jevner seg ut med bedre matematikkarakterer.



Figur 2.9 Effekten av matematikkarakter fra videregående skole på totalscore på nasjonal deleksamen. Blå og grønne sirkler viser individuelle studenter fra henholdsvis GLU 1–7 og GLU 5–10, linjene er regresjonslinjer for hver GLU-type, og skygget område viser 95 prosent konfidensintervall.

2.4 Eksamensoppgaver og sensorveiledning

Det finnes en mulighet for at eksamensoppgavene høsten 2016 var vanskeligere eller lettere enn eksamensoppgavene høsten 2015 og våren 2016. For å vurdere dette har vi gjennomført en såkalt Rasch-analyse.[3] 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, og at kandidater fikk 1 poeng uavhengig av deres dyktighet. Analysen for oppgavesettet våren 2016 viste at også dette oppgavesettet holdt høy kvalitet, men at noen få oppgaver var så vanskelige at nesten ingen studenter greide å løse disse korrekt. Analysen fant derimot ingen indikasjoner på at kandidatene som fikk 1 poeng på en oppgave, fikk det uavhengig av deres dyktighet. Oppgavene på eksamenen høsten 2016 var overordnet sett tilfredsstillende, men med noe lav spredning i vanskelighetsgrad. En svakhet var også at 1 poeng ble relativt lite brukt i flere av oppgavene. Det vil si at studentene enten fikk galt (0 poeng) eller riktig (2 poeng), og disse oppgavene skiller dermed dårligere mellom studenter på ulike nivåer. Sammenlignet med eksamenen våren 2016 fremstår oppgavene på høstens eksamen som noe enklere, men dette kan også skyldes dyktigere studenter.

Vi har også bedt sensorene om å vurdere høstens eksamen, der 33 av 50 sensorer besvarte en kort undersøkelse[4]. Sensorene kunne svare på ulike påstander på en skala fra 1 til 5, samt komme med kommentarer. Vedrørende vanskelighetsgraden på eksamen svarte 79 prosent av sensorene at eksamenssettet var «passe vanskelig», mens 18 prosent mente det var for lett og 3 prosent at det var

[3] Analysen er utført av Rolf Vegar Olsen ved Centre for Educational Measurement (CEMO), UiO.
[4] Høsten 2015 svarte 39 av 48 sensorer på undersøkelsen, og våren 2016 svarte 27 av 48 sensorer.

noe for vanskelig. Dette tyder på at sensorene syntes at vanskelighetsgraden var riktig tilpasset emnet og studentene.

Vi spurte også sensorene om de syntes vårens oppgavesett var lettere, hadde lik vanskelighetsgrad eller var vanskeligere enn vårens oppgavesett, og her var gjennomsnittsscoren 2,5 (der 1 var «eksamenssettet fra desember 2016 var lettere», 3 var «de hadde samme vanskelighetsgrad» og 5 var «eksamenssettet fra desember 2016 var vanskeligere»)[5]. Dette tyder på at sensorene mener at høstens oppgaver har tilnærmet lik vanskelighetsgrad som våren 2016, men likevel kanskje noe lettere. Til sammenligning fra vårens eksamen scoret sensorene 3,4 når de sammenlignet våren 2016 med høsten 2015 (der 1 var «mye lettere i 2016» og 5 var «mye vanskeligere i 2016»). Til sammen indikerer dette at oppgavene kan ha vært noe vanskeligere våren 2016 enn høsten 2015, og en anelse lettere igjen høsten 2016.

Fra sensorveiledningen i 2015 var det relativt lett å oppnå 1 poeng på enkeltoppgaver, mens det ble enklere å skille mellom 0 og 1 poeng på våren 2016 (67 prosent av sensorene mente det var «mye enklere å skille mellom 0 og 1 poeng»). For høsten 2016 scoret sensorene i gjennomsnitt 3,7 på om det var enkelt å skille mellom 0 og 1 poeng (der 1 var vanskelig og 5 var lett). Dette tyder på at sensorene mener det var vanskeligere for studentene å få 1 poeng på enkeltoppgaver på eksamenene i 2016 enn på høsten 2015.

2.5 Oppsummering resultater

Sammenlagt ser vi at det er flere faktorer som forklarer karaktervariasjonen på nasjonal deleksamen høsten 2016, inkludert GLU-type, om eksamen er tellende, startkompetanse og vanskelighetsgraden på eksamen. Basert på de analysene vi har hatt mulighet til å gjøre, er det sannsynligvis to hovedgrunner til at resultatene høsten 2016 var bedre enn våren 2016. Den første er at andelen av studentene som går på GLU 5–10, er høyere for høsten 2015 og høsten 2016 enn for våren 2016. GLU 5–10 gjør det signifikant bedre enn GLU 1–7, og det kan delvis forklare hvorfor resultatene er bedre på høst-eksamenene. Den andre hovedårsaken er at eksamen nå er tellende, og det får dermed større konsekvenser om studentene ikke består eksamen. Dette har trolig økt studieinnsatsen og ført til at strykprosenten er vesentlig lavere enn for våren 2016.

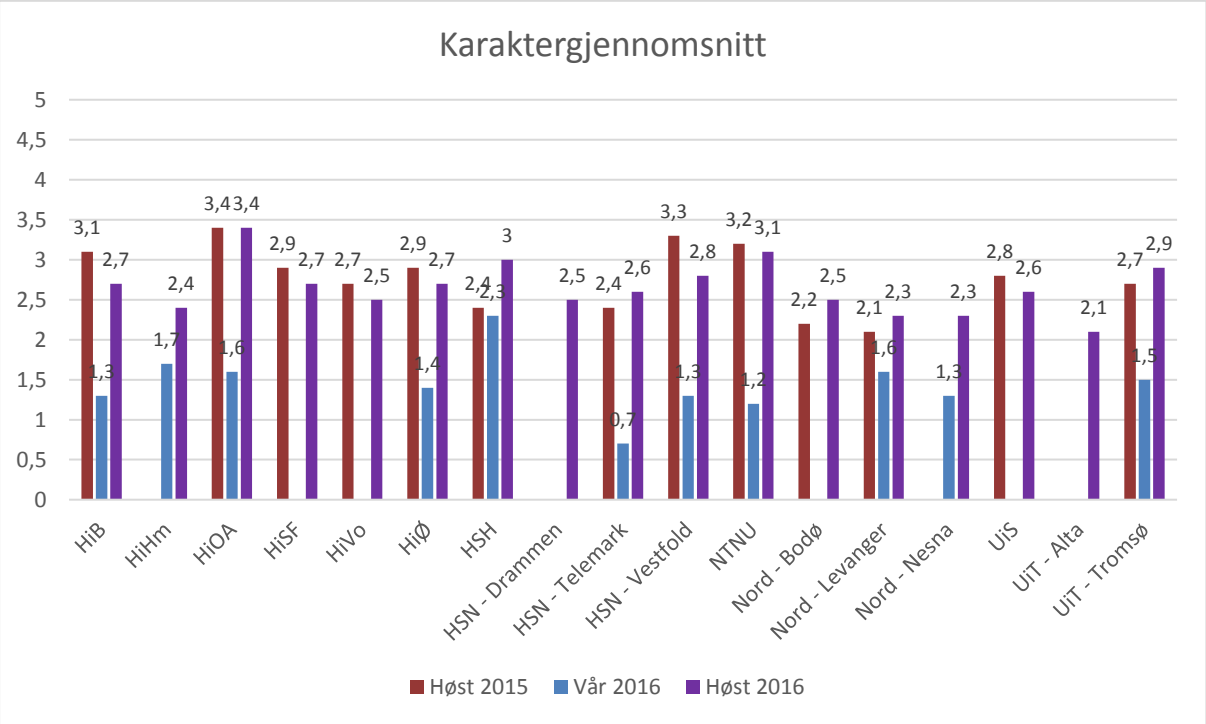
Av andre faktorer som spiller inn, ser vi at også starkompetanse forklarer noe av variasjonen i hvor mange poeng studentene får på deleksamen, der matematikkarakteren forklarer mer enn karakterpoengene. Eksamenssettet ble også oppfattet som en anelse lettere enn for våren 2016, og eksamenssettets noe lavere vanskelighetsgrad kan også ha påvirket resultatene.

3 Institusjonsresultater

Antall studenter og karaktergjennomsnitt varierer noe for hver institusjon. Figur 3.1 viser gjennomsnittskarakteren på den nasjonale deleksamenen for høsten 2015, våren 2016 og høsten 2016. Tabellen under viser antallet studenter og 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

[5] 30 av de svarende var også sensor våren 2016.

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. Det er også viktig å påpeke at man må vise forsiktighet i tolkningen av resultatene, siden noen av gruppene er veldig små, noe som fører til at individuelle forskjeller i større grad påvirker gjennomsnittet.



Figur 3.1 Gjennomsnittskarakter per institusjon på nasjonal deleksamen.

Høsten 2016 deltok studenter fra 12 forskjellige institusjoner, med et gjennomsnittsantall på 62 studenter. Flest antall studenter finner vi på Universitetet i Stavanger (140), NTNU (133), og Høgskolen i Hedmark (119). Karaktergjennomsnittet rangerer fra 2,1 til 3,4, hvor institusjonene med høyest snitt var Høgskolen i Oslo og Akershus (3,4), NTNU (3,1) og Høgskolen Stord/Haugesund (3,0).

Tabell 3 Karakterfordeling per institusjon høsten 2016 (prosent).

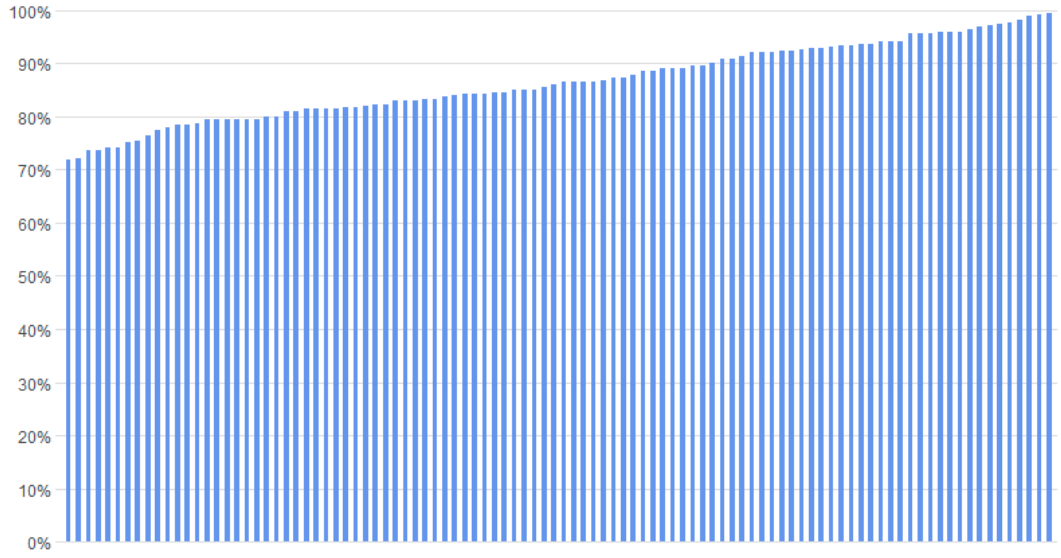
Institusjon	**Antall**	**A**	**B**	**C**	**D**	**E**	**F**
HiB	50	6	16	48	10	8	12
HiH	119	2,5	21	27,7	21,8	13,4	13,4
HiOA	82	8,5	35,4	45,1	8,5	0	2,4
HSN – Drammen	39	2,6	15,4	35,9	30,8	5,1	10,3
HSN – Telemark	97	7,8	13,3	41,8	21,9	5,0	10,3
HSN – Vestfold	37	5,4	27	32,4	21,6	2,7	10,8
HiVo	47	4,3	14,9	36,2	25,5	10,6	8,5
HiØ	43	7	18,6	34,9	23,3	9,3	7
HSH	36	13,9	19,4	36,1	22,2	0	8,3
HiSF	86	4,7	17,4	44,2	17,4	9,3	7
Nord – Bodø	26	0	15,4	46,2	26,9	0	11,5

Nord – Levanger	50	2	6	42	34	4	12
Nord – Nesna	11	0	0	45,5	45,5	0	9,1
NTNU	133	11,3	22,6	42,9	15	2,3	6
UiT – Alta	19	0	5,3	42,1	26,3	10,5	15,8
UiT – Tromsø	44	2,3	22,7	50	15,9	4,5	4,5
UiS	140	5	14,3	40,7	22,9	6,4	10,7

4 Sensurreliabilitet

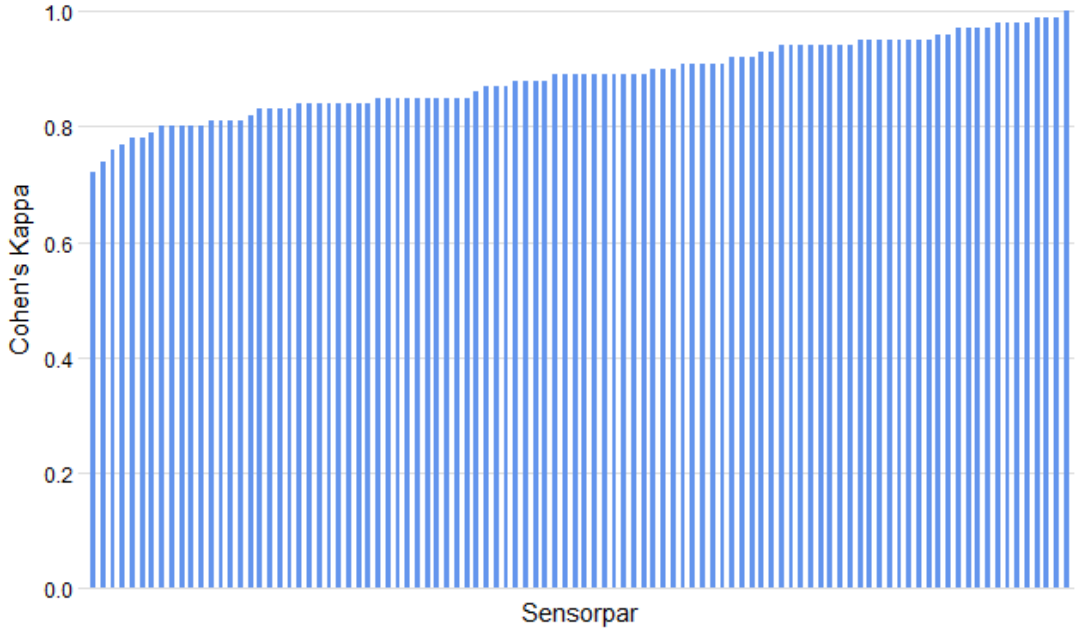
Før vi konkluderer i denne delrapporten, vil vi si noe om sensurreliabiliteten. I samarbeid med eksamensgruppen ba NOKUT alle sensorer om å rapportere detaljert sensur. Vi ba hver sensor om å oppgi hvor mange poeng hver kandidat fikk på hver enkelt eksamensoppgave. Sensorene fikk et poengskjema hvor 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. På denne måten har NOKUT mulighet til å vurdere sensurreliabiliteten.

Som beskrevet over var det totalt 50 sensorer som sensurerte den nasjonale deleksamenen. Disse ble delt inn i grupper på fem, og i hver gruppe sensurerte alle sensorene sammen med hverandre. Til sammen var det 100 sensorpar, og hvert sensorpar samsensurerte 10 eller 11 eksamensoppgaver. Hvert eksamenssett inneholdt 20 spørsmål, så i gjennomsnitt sensurerte hvert sensorpar mellom 200 og 220 enkeltoppgaver sammen. Figur 4.1 viser hvor mange prosent av disse oppgavene hvert sensorpar hadde vurdert likt, der hver stolpe representerer ett sensorpar, og høyden på stolpen viser hvor mange prosent av oppgavene sensorene har vurdert likt. Hver enkelt sensor vurderte i gjennomsnitt 42 besvarelser.



Figur 4.1 Sensurreliabilitet: Prosentvis enighet.

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.[6] 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.



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 100 sensorparene har alle en kappa-koeffisient på over 0,70. Hvor godt er dette? I forskningslitteraturen opererer man med følgende skala:[7]

- 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 og våren 2016, hvor henholdsvis 83 prosent og 96 prosent av sensorparene hadde en kappa-koeffisient på over 0,60. Det at sensurreliabiliteten nå er enda høyere, 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 tre 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.

[6] Se for eksempel Krippendorf, K. (2004). «Reliability in content analysis: Some common misconceptions and recommendations» Human Communications Research 300: 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.
[7] Se for eksempel Landis, J. & Koch, G. (1977). «The Measurement of Observer Agreement for Categorical Data» Biometrics 33: 159–174.

Resultatene etter klagesensuren bekrefter den høye sensurreliabiliteten. Etter matematikkeksamenen i grunnskolelærerutdanningene høsten 2016 var det totalt 17 av 1059 kandidater (1,6 prosent) som klagde og fikk ny sensur. Av disse gikk fem kandidater opp én karakter, mens to kandidater gikk opp mer enn én karakter.

Tabell 4 Klagesensur GLU høsten 2016

Opprinnelig karakter	Antall	Antall forbedret karakter	Antall dårligere karakter	Opp 1 karakter	Ned 1 karakter	Opp mer enn 1 karakter
F	11	2	0	1	0	1
E	3	2	0	1	0	1
D	2	1	0	1	0	0
C	1	0	0	0	0	0
B	0	0	0	0	0	0
SUM	16	5	0	3	0	2

5 Konklusjon

Resultatene på den nasjonale deleksamenen i matematikk for grunnskolelærerutdanningen høsten 2016 var bedre enn våren 2016 og tilnærmet like resultatene for høsten 2015. Hovedfokuset i denne delrapporten har vært å beskrive resultatene for høstens eksamen, samt å belyse hvilke faktorer som kan forklare hvorfor vårens resultater var så mye svakere enn resultatene for høsten 2015 og høsten 2016.

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.

Først og fremst viser analysen at et stort antall studenter med høy sannsynlighet nedprioriterte den nasjonale deleksamenen våren 2016, ettersom denne ikke var tellende på vitnemålet. Etter at eksamenen ble gjort tellende høsten 2016, har vi sett en forbedring i resultatene. Gjennomsnittskarakteren og -fordelingen er lik den for høsten 2015, hvor studentene ikke visste at eksamenen ikke var tellende. Det er dermed helt tydelig at når studentene tror (høsten 2015) eller vet (høsten 2016) at eksamenen teller, øker de innsatsen og får bedre karakterer enn når de vet at den ikke teller (våren 2016). Det gir sektoren som helhet mer robuste resultater, som de kan bruke i sitt utviklingsarbeid.

Den andre faktoren er at det var flere studenter fra GLU 5–10 som avla eksamen på høsten enn på våren 2016, 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. Når vi kun sammenlignet innad i hver GLU-type, så vi imidlertid at resultatene fortsatt var svakere på våren, og endringen i resultater over tid kan derfor ikke bare skyldes forskjeller i andelen studenter i hver GLU-type.

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 eksamenen høsten 2016 var en anelse lettere enn våren 2016. Undersøkelsene blant sensorene bekrefter funnene fra Rasch-analysene. Det er med andre ord sannsynlig at oppgavesettet var noe vanskeligere våren 2016 enn høsten 2015 og høsten 2016, og det kan forklare noe av forskjellene i resultatene.

I tillegg til å forklare resultatforskjellene har vi også vist at studentenes startkompetanse (målt i karakterpoeng fra videregående skole) har en innvirkning på resultatene, der høyere karakterer fra videregående er positivt korrelert med antall poeng på deleksamen. Det samme gjelder matematikkarakteren fra videregående skole.

NOKUT har nå gjennomført nasjonal deleksamen i matematikk for grunnskolelærerutdanningene tre ganger. Gjennomføringen av eksamenene og sensuren har fungert meget godt, og i prinsippet tror NOKUT at nasjonale deleksamener kan fungere som et godt virkemiddel i sektorens kvalitetsarbeid.