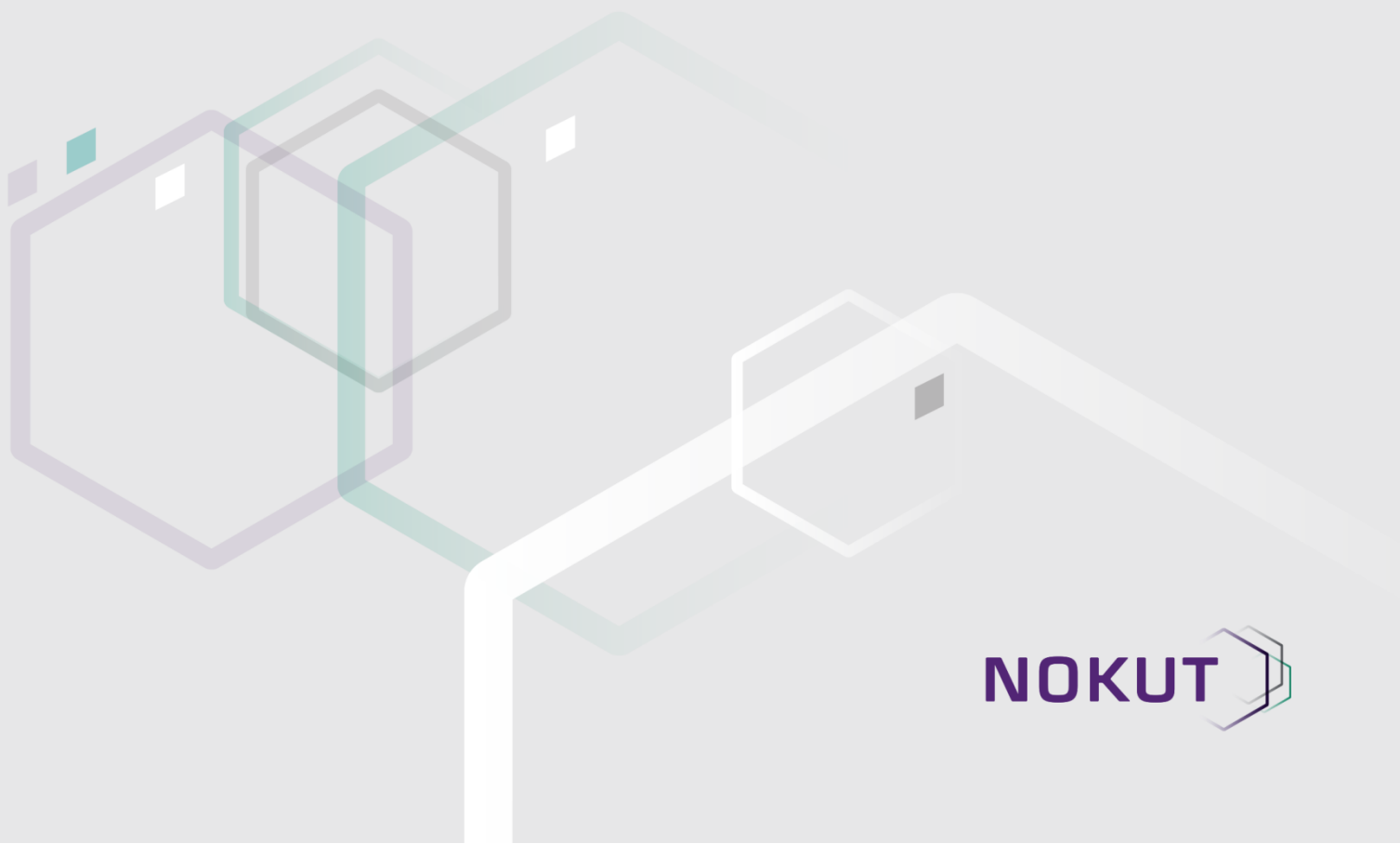


NOKUTs oppsummeringer

Nasjonale deleksamen i anatomi, fysiologi og biokjemi for sykepleierutdanningen – 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 anatomi, fysiologi og biokjemi for sykepleierutdanningen
Forfatter(e):	Kristina C. Skåtun Kjersti Tokstad (prosjektleder)
Dato:	15. mars 2017

Sammendrag

16. desember 2016 arrangerte NOKUT i samarbeid med Nasjonalt profesjonsråd for utdanning og forskning innen sykepleie (NPR-sykepleie) og 16 utdanningsinstitusjoner den andre nasjonale deleksamenen i anatomi, fysiologi og biokjemi for studenter i bachelorgradsstudiet i sykepleie. Totalt 4838 kandidater gjennomførte eksamen og fikk sensur.

Eksamenen 2015 var basert på frivillig deltagelse, og 12 institusjoner meldte seg frivillig til å delta i denne første piloten. Kriteriene for å delta var at institusjonene hadde tilnærmet likt pensum og læringsutbyttebeskrivelser, at de hadde mulighet til å avholde eksamenen på samme dag, og at de aksepterte å bruke de oppgavene som ble foreslått som ordinære eksamensoppgaver. Etter fusjonene i 2016 ble bachelorgradsstudiet i sykepleie tilbudt ved 16 institusjoner (fordelt på 30 studiesteder), og alle gjennomførte den nasjonale deleksamenen i anatomi, fysiologi og biokjemi som studentenes ordinære eksamen i faget. Denne rapporten må ses i sammenheng med den første rapporten, som omhandlet den nasjonale deleksamenen som ble avholdt 17. desember 2015.

Resultatene for høstens nasjonale deleksamen var noe bedre enn eksamenen høsten 2015. Andelen som fikk A eller B, var 26 prosent i 2016, sammenlignet med 19,2 prosent i 2015. Andelen stryk var også lavere i 2016, med 22 prosent i 2016 og 29 prosent i 2015. Den nasjonale gjennomsnittskarakteren for 2016 og 2015 var en D. Det er også en sammenheng med hvor bra studentene har gjort det på videregående skole (karakterpoeng) og hvor mange poeng de har fått på nasjonal deleksamen. Resultatene viser også en variasjon mellom studiestedene, der karaktergjennomsnittet rangerer fra en E til en svært god C.

NOKUT har nå gjennomført nasjonal deleksamen i anatomi, fysiologi og biokjemi 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.

Innhold

1	Innledning	1
1.1	Organisering av nasjonal deleksamen	1
1.1.1	Eksamensdagen	1
1.1.2	Sensur	2
1.2	Utvalget	2
2	Nasjonale resultater	4
2.1	Karakterpoeng fra videregående skole	5
2.2	Institusjonsbidrag.....	6
2.3	Eksamensoppgaver og sensorveiledning	9
3	Institusjonsresultater	9
4	Sensurreliabilitet	12
5	Konklusjon	13

1 Innledning

NOKUT fikk høsten 2014 i oppdrag av Kunnskapsdepartementet å gjennomføre en mulighetsstudie og et pilotprosjekt med nasjonale deksamener 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 deksamnen i anatomi, fysiologi og biokjemi i sykepleierutdanningen som ble avholdt 16. desember 2016. Den første nasjonale deksamnen i anatomi, fysiologi og biokjemi ble avholdt 17. 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.

Bachelorgradsstudiet i sykepleie ble tilbudt ved 28 utdanningsinstitusjoner før fusjonene trådte i kraft fra 1. januar 2016. Den nasjonale deksamnen som ble avholdt 17. desember 2015, omfattet kun 12 av de 28 institusjonene, og var en slags pilot i piloten. Forskjellene mellom institusjonenes undervisningsopplegg og vurderingsformer i anatomi, fysiologi og biokjemi var for store til at det var mulig å gjennomføre en fullskala eksamen allerede i 2015. Etter fusjonene i 2016 ble bachelorgradsstudiet i sykepleie tilbudt ved 16 institusjoner (fordelt på 30 studiesteder), og alle gjennomførte den nasjonale deksamnen i anatomi, fysiologi og biokjemi som studentenes ordinære eksamen i faget.

1.1 Organisering av nasjonal deksamnen

Det ble satt ned en arbeidsgruppe som utviklet den nasjonale deksamnen i anatomi, fysiologi og biokjemi i 2015. Med unntak av et par personer er det den samme arbeidsgruppen som har utviklet eksamenen i 2016. Ved siden av eksamensoppgaver og sensorveiledninger, har arbeidsgruppen også utviklet en felles læringsutbyttebeskrivelse og en detaljert fagbeskrivelse, som ble sendt ut til institusjonene i juni. Erfaringene fra den første nasjonale deksamnen tilsa at det var behov for et slikt dokument, som kunne fungere som et verktøy, siden studentene forholder seg til ulike læreverk. Dokumentet har på forhånd vært lagt frem til diskusjon i fagmiljøene og i Nasjonalt profesjonsråd for utdanning og forskning innen sykepleie.

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 sensorcorpset samt formidling av eksamensresultatene til institusjonene.

1.1.1 Eksamensdagen

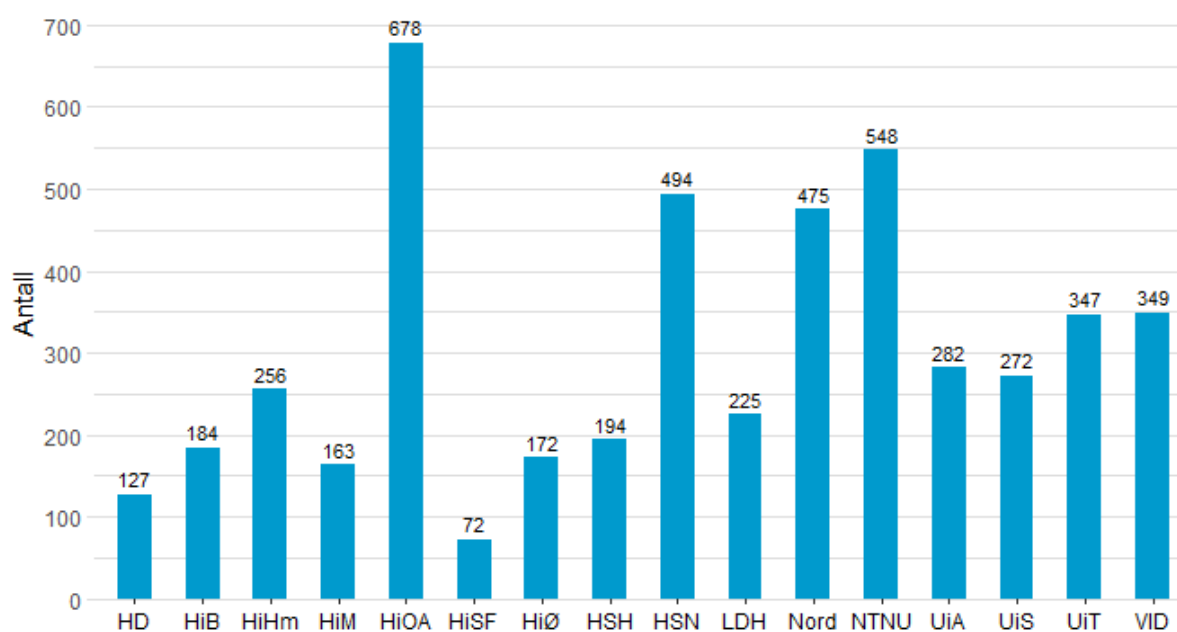
Den andre nasjonale deksamnen i anatomi, fysiologi og biokjemi ble avholdt som en firetimers skriftlig skoleeksamen uten bruk av hjelpemidler 16. desember 2016. Det var 16 institusjoner som deltok, fordelt på 30 studiesteder, og kandidatene var i all hovedsak i sitt første semester. Totalt gjennomførte 4838 av 5388 oppmeldte kandidater eksamen, som tilsvarer en gjennomføringsgrad på 89,8 prosent.

1.1.2 Sensur

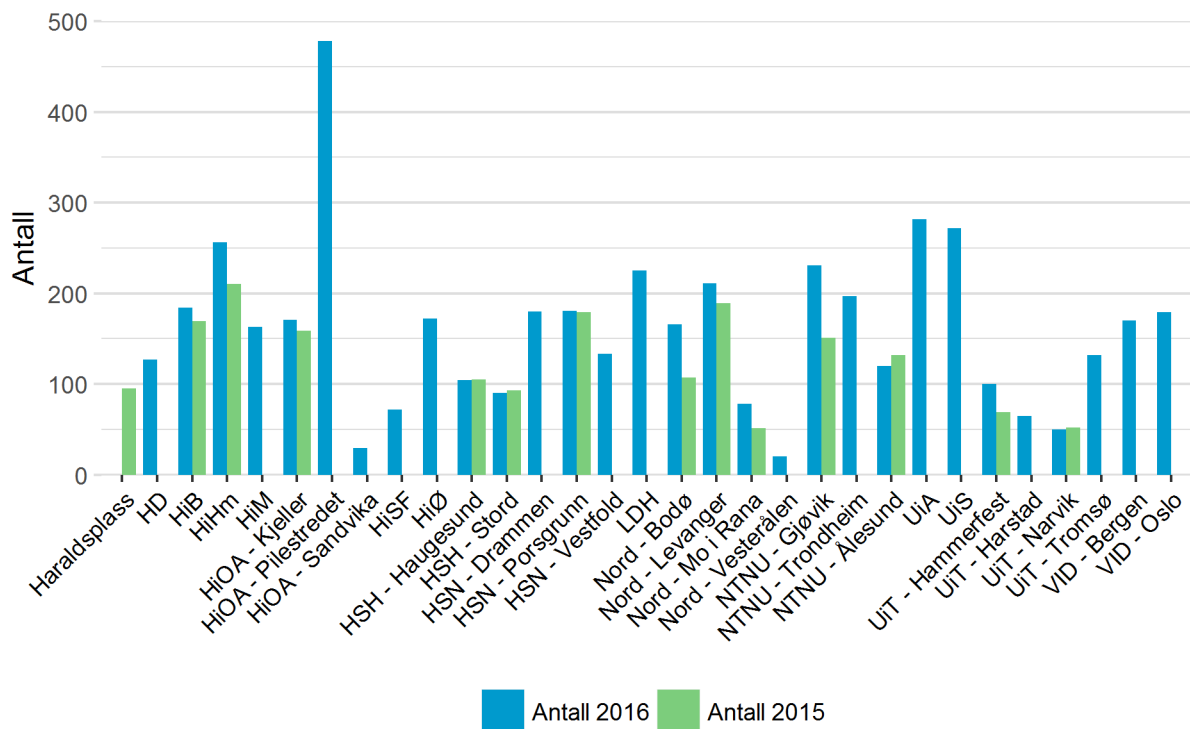
For å sikre likebehandling og full anonymitet ble det satt sammen et sensorcorps på 79 sensorer. Hver besvarelse ble vurdert av to sensorer. For å få til en karakterkalibrering og en institusjonell spredning av besvarelsene delte NOKUT inn sensorerne i til sammen 156 sensorpar, hvor hver sensor sensurerte sammen med 4 andre sensorer. Hvert sensorpar sensurerte i snitt 31 besvarelser, og hver sensor sensurerte i gjennomsnitt 122 besvarelser, som alle var jevnt og tilfeldig fordelt fra alle læresteder.

1.2 Utvalget

Sammenlignet med høsten 2015 var det flere institusjoner som deltok høsten 2016, og antall kandidater var dermed høyere (4838 i 2016 versus 1761 i 2015). Figur 1.1 viser antall kandidater i 2016 fordelt på institusjoner, og figur 1.2 og tabell 1.1 viser antall kandidater fordelt på institusjoner og studiesteder for de to siste årene. Gjennomsnittsantallet var på 161 kandidater, og høyest antall kandidater var på Høgskolen i Oslo og Akershus, NTNU, Høgskolen i Sørøst-Norge og Nord universitet.



Figur 1.1 Antall kandidater per institusjon for eksamen 2016.



Figur 1.2 Antall kandidater per studiested for eksamen 2016 og 2015.

Institusjon	Antall 2016	Antall 2015
Haraldsplass	–	95
HD	127	–
HiB	184	169
HiHm	256	210
HiM	163	–
HiOA – Kjeller	171	159
HiOA – Pilestredet	478	–
HiOA – Sandvika	29	–
HiSF	72	–
HiØ	172	–
HSH – Haugesund	104	105
HSH – Stord	90	93
HSN – Drammen	180	–
HSN – Porsgrunn	181	179
HSN – Vestfold	133	–
LDH	225	–
Nord – Bodø	166	107
Nord – Levanger	211	189
Nord – Mo i Rana	78	51
Nord – Vesterålen	20	–
NTNU – Gjøvik	231	151
NTNU – Trondheim	197	–
NTNU – Ålesund	120	132
UiA	282	–
UiS	272	–
UiT – Hammerfest	100	69

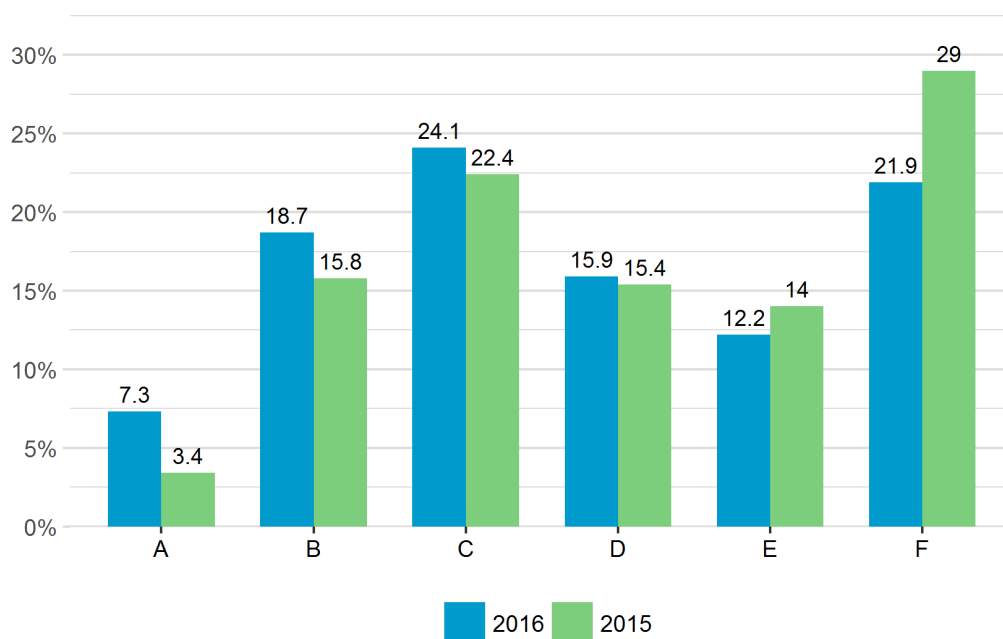
UiT – Harstad	65	–
UiT – Narvik	50	52
UiT – Tromsø	132	–
VID – Bergen	170	–
VID – Oslo	179	–
Totalt	4838	1761

Tabell 1.1 Antall kandidater for 2015 og 2016, fordelt på studiested.

2 Nasjonale resultater

I dette kapittelet presenter vi resultatene fra den nasjonale deleksamenen i anatomi, fysiologi og biokjemi, som ble gjennomført 16. desember 2016. Først presenterer vi tall for hele populasjonen, og deretter resultatene på institusjonsnivå. Vi avslutter med en analyse av sensuren.

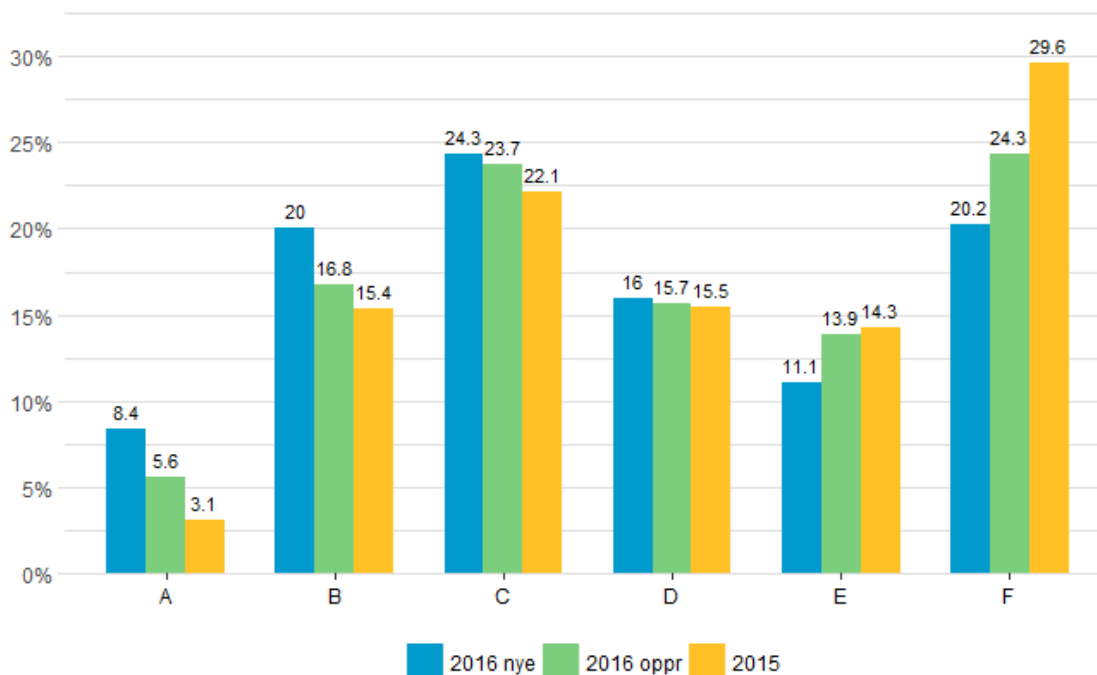
Resultatene for 2016 var generelt bedre enn for 2015. Andelen som fikk A eller B, var 26 prosent i 2016, sammenlignet med 19,2 prosent i 2015. Andelen stryk var også lavere i 2016, med 22 prosent i 2016 og 29 prosent i 2015. Den nasjonale gjennomsnittskarakteren for 2016 og 2015 var en D. Dersom vi gjør om karakterene til tall (A=5, B=4, C=3, D=2, E=1 og F=0), ser vi at gjennomsnittskarakteren var 1,9 i 2015 og 2,3 i 2016. Den reelle forskjellen er altså 0,4 karakterpoeng, som utgjør nesten en halv karakter. Forskjellen i karakterpoeng mellom årene er signifikant ($t=8$, $p<0,001$). Figur 2.1 viser karakterfordelingen for 2015 og 2016.



Figur 2.1 Karakterfordeling nasjonalt for eksamen 2015 og 2016.

Det er viktig å merke seg at antallet kandidater er vesentlig mye høyere i 2016, og flere institusjoner har deltatt. Studentutvalget er dermed veldig ulikt i 2015 og 2016, og kan delvis forklare forskjellen i gjennomsnittskarakter. En annen grunn til en forbedring av resultatene kan være at studentene nå vet bedre hvordan eksamenen er lagt opp og har hatt muligheten til å øve på den eksamenen som allerede har blitt gjennomført (oppgavene ligger tilgjengelig på NOKUTs hjemmesider).

Vi ønsket derfor også å sammenligne kun de studiestedene som var med begge årene for å undersøke om karakterene også har blitt bedre her, eller om de i hovedsak skyldes bedre resultater fra studentene på de nye institusjonene fra 2016. Resultatene for studiestedene som var med begge årene, viser en gjennomsnittskarakter på 2,1 i 2016, sammenlignet med 1,9 i 2015. Denne forskjellen er mindre enn for hele utvalget fra 2016, men den er fortsatt signifikant ($t=4,4$, $p<0,001$). Gjennomsnittskarakteren for de nye studiestedene i 2016 er på 2,4, som er signifikant høyere enn studiestedene som var med i både 2016 og 2015 (henholdsvis $t=5,6$ og $t=10,2$, med $p<,001$). Figur 2.2 viser karakterfordelingen for studiestedene som var med begge årene og de nye studiestedene i 2016.



Figur 2.2 Karakterfordeling nasjonalt for de nye studiestedene som har deltatt i 2016 (2016 nye), og de studiestedene som har deltatt begge årene (2016 oppr og 2015).

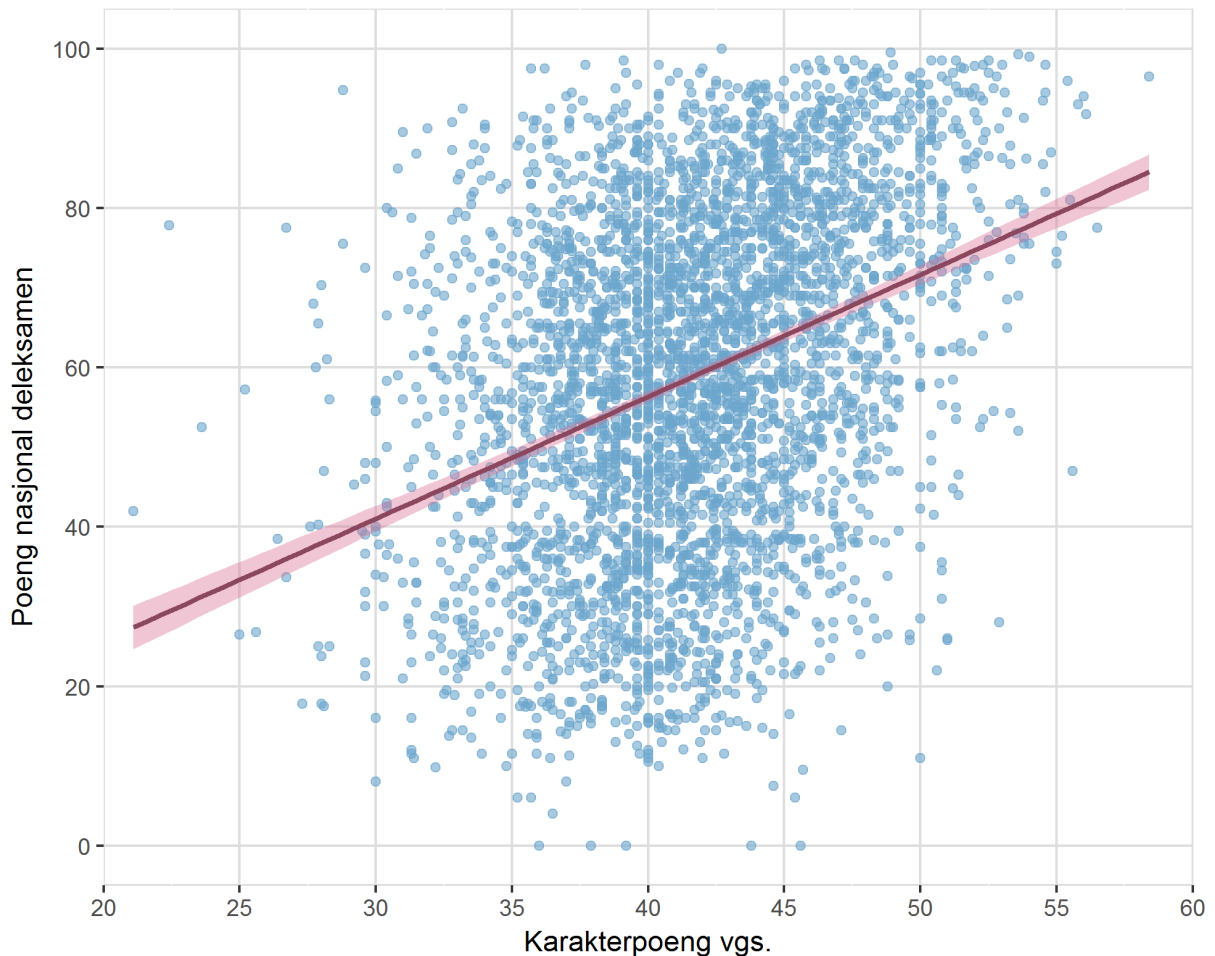
Resultatene indikerer at de studiestedene som var med begge årene, i snitt gjør det noe bedre på eksamenen i 2016, men at de nye studiestedene igjen gjør det bedre. Hvor bra studentene gjør det på eksamenen kan skyldes flere faktorer, slik som kvaliteten på undervisningen og studentenes motivasjon for å gjøre det bra. Forskjellene kan også forklares med studentenes startkompetanse, målt i form av karakterpoeng fra videregående skole, og vanskelighetsgraden på eksamensoppgavene. For å undersøke underliggende årsaker til resultatene på eksamen har vi sett nærmere på studentenes startkompetanse, foretatt en analyse av vanskelighetsgraden på eksamensoppgavene, og sett på forskjeller mellom institusjonene. Disse funnene vil bli diskutert i de følgende kapitlene.

2.1 Karakterpoeng fra videregående skole

Hvis vi ser på korrelasjonen mellom karakterpoeng fra videregående skole¹ og antall poeng på nasjonal deleksamen høsten 2016, får vi en signifikant svak positiv korrelasjon ($r=0,35$, $p<0,001$). En regresjonsanalyse, der antall poeng på den nasjonale deleksamenen er avhengig variabel og

¹ Vi har karakterpoeng fra videregående skole for 4193 studenter. Vi har mottatt bakgrunnsdata på individnivå fra CERES (tidligere FSAT).

karakterpoeng er uavhengig variabel, viser at karakterpoeng forklarer 12 prosent av variasjonen i resultatene. Analysen viser at for hvert karakterpoeng høyere en student har fra videregående skole, gir det 1,5 poeng mer i uttelling på totalscoren på nasjonal deleksamen. Dette betyr at de som har gjort det bra på videregående skole, får høyere karakter på nasjonal deleksamen høsten 2016 (figur 2.3). Vi har ikke bakgrunnsdata for eksamenen i 2015 og kan dermed ikke sammenligne denne effekten med tidligere resultater.



Figur 2.3 Effekten av karakterpoeng fra videregående skole på totalscore på nasjonal deleksamen. Blå sirkler viser individuelle studenter fra nasjonal deleksamen, den røde linjen er regresjonslinjen og rosaskyget område viser 95 prosent konfidensintervall.

2.2 Institusjonsbidrag

Som nevnt over hadde vi ikke tilgang til individdata etter eksamenen høsten 2015. Dette har vi nå. I tillegg til å se på sammenhengen mellom karakterpoeng fra videregående skole og totalscoren på eksamen, kan vi nå også kontrollere for andre individspesifikke variabler, som for eksempel kjønn og alder. I tabell 2.1 viser vi resultatene fra en slik enkel regresjonsmodell.

Avhengig variabel: Totalt antall poeng (0–100)	
	Modell 1
Karakterpoeng fra VGS	1,7*** (0,07)
Alder	0,65*** (0,07)
Kvinne	-4,9*** (1,00)
Konstant	-2,27*** (3,6)
N	3912
***p > .01, **p > .05, *p > .1	

Tabell 2.1 Regresjonsanalyse

Analysen viser at hvert karakterpoeng høyere fra videregående skole gir 1,7 poeng mer i uttelling på totalscoren på den nasjonale deleksamenen², hvert år eldre studentene er gir 0,65 poeng mer i uttelling på totalscoren og kvinner scorer 4,9 poeng færre enn menn.³ Det er viktig å påpeke at analysen kun forklarer 15 prosent av variasjonen i resultatene.

Regresjonsmodellene kan også brukes til å vurdere institusjonenes bidrag til studentenes resultater på de nasjonale deleksamenene. Institusjonsbidragsmodellen som presenteres her, er en enkel modell som gir indikasjoner på institusjonenes bidrag, men den gir ikke noe fasitsvar.

For å «måle» institusjonsbidraget bruker man regresjonsanalysen på individnivå til å predikere hva gjennomsnittspoengscoren per institusjon skulle ha vært basert på studentenes startkompetanse, kjønn og alder. Man ser så på forskjellen mellom den predikerte gjennomsnittsscoren og den faktiske gjennomsnittsscoren. Institusjoner med en faktisk gjennomsnittsscore som er høyere enn den predikerte gjennomsnittsscoren, har et positivt institusjonsbidrag, mens institusjoner med en faktisk gjennomsnittsscore som er lavere enn den predikerte gjennomsnittsscoren, har et negativt institusjonsbidrag. I tabellen nedenfor illustrerer vi metoden basert på resultater og data fra høstens eksamen i anatomi, fysiologi og biokjemi. Vi har valgt å vise dette basert på studiested, ettersom flere institusjoner tilbyr bachelorgradsstudiet i sykepleie på flere studiesteder. Dette gir også de forskjellige studiestedene på samme institusjon en mulighet til å sammenligne seg med hverandre internt i organisasjonen, noe som vi håper bidrar til at fagmiljøene kommuniserer med hverandre om hvordan de kan utvikle seg videre.

² Dersom vi ikke kontrollerer for alder og kjønn, gir hvert karakterpoeng 1,5 poeng mer på den nasjonale deleksamenen.

³ Av de 3912 studentene vi har data med karakterpoeng, kjønn og alder for, er kun 412 studenter menn.

Studiested	Antall studenter	Faktisk gj.snittscore	Predikert gj.snittscore	Differanse
HiSF	69	67,32	58,08	9,23
HiB	172	73,19	66,63	6,56
VID – Bergen	155	68,58	62,53	6,04
NTNU – Trondheim	193	73,71	68,38	5,33
HiHm	235	60,02	54,83	5,19
UiA	268	65,03	60,03	5,00
VID – Oslo	167	64,65	60,26	4,39
UiT – Hammerfest	94	51,00	47,20	3,80
HD	114	62,82	59,39	3,43
HiOA – Kjeller	159	62,47	59,34	3,14
LDH	212	68,01	64,95	3,06
HiØ	161	57,85	55,66	2,19
NTNU – Ålesund	116	61,23	59,42	1,81
UiT – Tromsø	129	62,99	62,15	0,84
Nord – Levanger	199	57,68	56,85	0,83
NTNU – Gjøvik	228	56,19	56,28	-0,09
UiT – Harstad	63	51,11	51,60	-0,49
HiM	151	56,06	56,89	-0,82
UiT – Narvik	50	50,17	51,69	-1,52
HSN – Vestfold	124	55,99	57,73	-1,74
HiOA – Pilestredet	441	55,83	59,04	-3,21
HSH	176	51,99	56,65	-4,66
Nord – Mo i Rana	77	46,27	51,66	-5,39
HSN – Porsgrunn	160	48,36	54,01	-5,65
UiS	250	52,80	59,03	-6,23
Nord – Bodø	153	48,66	55,07	-6,40
HSN – Drammen	163	52,05	59,26	-7,20

Tabell 2.2 Institusjonsbidrag

Tabellen over viser institusjonsbidraget til 11 institusjoner hvor mer enn 50 kandidater avla eksamen i desember 2016. Vi bruker Høgskulen i Sogn og Fjordane som eksempel. 69 studenter tok eksamen, og gjennomsnittsscoren var 67,3 poeng (man kan oppnå maksimalt 100 poeng). Basert på startkompetanse, kjønn og alder predikerer modellen at gjennomsnittsscoren skulle ha vært 58 poeng. Studentene fikk altså i overkant av 9 poeng mer i gjennomsnitt enn det man kunne forvente basert på modellen. Ser man på Høgskolen i Sørøst-Norge, studiested Drammen, fikk studentene der 7,2 poeng færre i gjennomsnitt enn modellen predikerer. Som nevnt over forklarer regresjonsmodellen lite av variasjonen i resultatene, og utregningen av institusjonsbidraget er kun en indikasjon. Vi tror likevel at en slik beregning kan hjelpe fagmiljøene når de ønsker å finne andre fagmiljøer å samarbeide med for å videreutvikle studiekvaliteten.

2.3 Eksamensoppgaver og sensorveiledning

Dersom en nasjonal eksamen skal ha en verdi for institusjonene og for studentene, er det viktig at eksamenssettet er av høy kvalitet, og at oppgavene skiller godt mellom sterke og svake studenter.

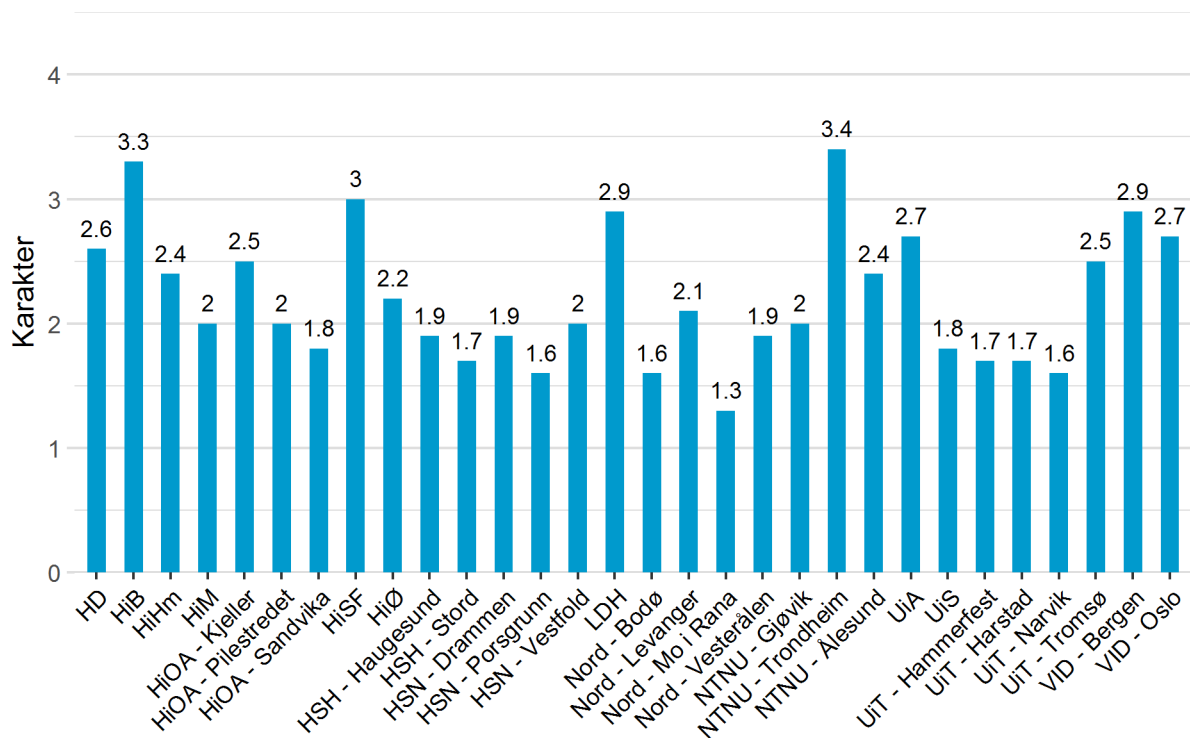
For å vurdere kvaliteten på eksamenssettet har vi gjennomført en såkalt Rasch-analyse.⁴ Resultatene fra høsten 2016 viser at eksamenssettet har en høy reliabilitet og totalt sett separerer godt mellom studenter med ulik dyktighet. Det er imidlertid en del oppgaver som ikke evner å skille mellom dyktighetsnivåer, og studentene får «for mange» poeng på disse. Noen av oppgavene er også litt for like hverandre og spør etter samme type informasjon. Flervalgsoppgavene har også i gjennomsnitt for lav vanskelighetsgrad, mens de åpne oppgavene er bedre tilpasset dyktighetsnivået til studentene. Dersom man skal fortsette med nasjonale deksamener, er det viktig at man videreutvikler eksamenssettet, slik at det blir best mulig. Rasch-analysen som er gjort i forbindelse med høstens eksamen, er en god start i så måte.

Vi har også bedt sensorene om å vurdere høstens eksamen, der 69 av 79 sensorer besvarte en kort undersøkelse. Sensorene kunne svare på ulike påstander på en skala fra 1 til 6, samt komme med kommentarer. Vedrørende vanskelighetsgraden på eksamenen, svarte sensorene i snitt 3,2 (der 1 er «lett» og 6 er «vanskelig»). Dette tyder på at sensorene synes at vanskelighetsgraden er riktig tilpasset emnet og studentene.

3 Institusjonsresultater

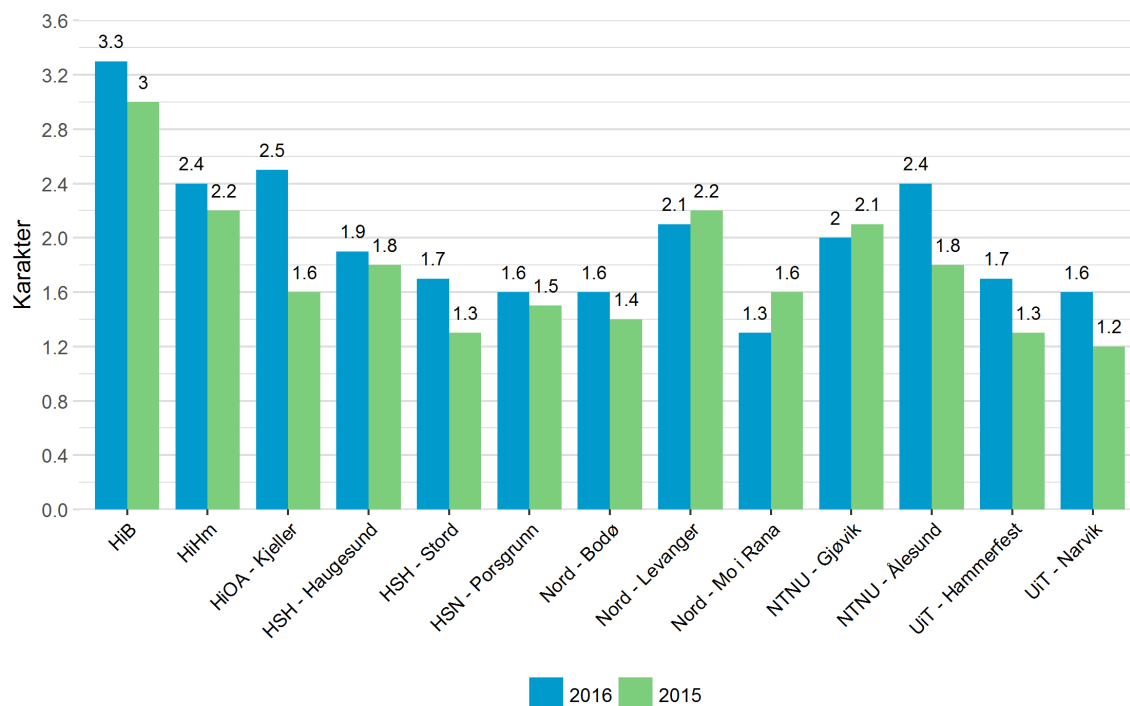
Antall kandidater og karaktergjennomsnitt varierer noe for hver institusjon. Figur 3.1 viser gjennomsnittskarakteren på den nasjonale deksameneren for 2016, og tabell 3.1 viser antall kandidater og karakterfordelingen per institusjon.

⁴ Analysen er utført av Øystein Guttersrud ved Naturfagsenteret, UiO.



Figur 3.1 Gjennomsnittskarakterer per institusjon på nasjonal deleksamen.

Fra figuren ser vi at det er stor variasjon i gjennomsnittskarakterer for de ulike studiestedene. Karaktergjennomsnittet rangerer fra 1,3 til 3,4, der institusjonene som scorer høyest er NTNU Trondheim, Høgskolen i Bergen, Høgskulen i Sogn og Fjordane, Lovisenberg diakonale høgskole og VID Bergen. Disse institusjonene har også lavest strykprosent, som også er et godt mål på om studentene har lyktes i faget. Høgskolen i Bergen og Høgskulen i Sogn og Fjordane har spesielt lav strykprosent, med rundt 5 prosent. Hvis vi sammenligner resultatene for studiestedene som har deltatt begge årene, ser vi at for de fleste stedene er karakterene blitt noe bedre i 2016 (figur 3.2). Høgskolen i Oslo og Akershus Kjeller og NTNU Ålesund viser størst forbedring i karakterer.



Figur 3.2 Gjennomsnittskarakter på nasjonal deleksamen for institusjonene som har deltatt de to siste årene.

Institusjon	Antall	A	B	C	D	E	F
HD	127	4.7	29.1	26	15	7.1	18.1
HiB	184	15.2	36.4	28.3	9.2	5.4	5.4
HiHm	256	5.9	17.6	30.5	18	11.3	16.8
HiM	163	5.5	14.1	22.1	21.5	10.4	26.4
HiOA – Kjeller	171	6.4	22.8	24.6	19.3	11.1	15.8
HiOA – Pilestredet	478	5.6	16.3	21.3	16.1	14	26.6
HiOA – Sandvika	29	3.4	3.4	27.6	27.6	6.9	31
HiSF	72	15.3	22.2	30.6	16.7	9.7	5.6
HiØ	172	3.5	19.8	25	18	9.9	23.8
HSH – Haugesund	104	3.8	14.4	25	14.4	12.5	29.8
HSH – Stord	90	2.2	6.7	26.7	16.7	14.4	33.3
HSN – Drammen	180	5.6	14.4	18.9	17.2	10	33.9
HSN – Porsgrunn	181	2.2	11.6	18.8	14.9	13.3	39.2
HSN – Vestfold	133	4.5	13.5	27.1	15.8	13.5	25.6
LDH	225	13.3	28.4	24	16	8.4	9.8
Nord – Bodø	166	3	12	20.5	10.2	16.9	37.3
Nord – Levanger	211	5.2	16.6	22.3	18	19.4	18.5
Nord – Mo i Rana	78	3.8	6.4	15.4	14.1	15.4	44.9
Nord – Vesterålen	20	0	15	30	15	10	30
NTNU – Gjøvik	231	4.3	15.2	25.1	14.7	14.7	26
NTNU – Trondheim	197	26.4	29.4	20.3	8.6	7.1	8.1
NTNU – Ålesund	120	7.5	18.3	25.8	19.2	15	14.2
UiA	282	8.5	24.5	26.2	18.8	11	11
UiS	272	1.8	11.8	22.4	19.9	17.3	26.8
UiT – Hammerfest	100	4	12	17	18	19	30
UiT – Harstad	65	1.5	10.8	27.7	7.7	16.9	35.4
UiT – Narvik	50	4	10	12	22	18	34

UiT – Tromsø	132	9.1	21.2	27.3	15.9	9.1	17.4
VID – Bergen	170	14.1	27.6	27.6	8.8	8.2	13.5
VID – Oslo	179	10.6	21.8	29.6	14	8.9	15.1

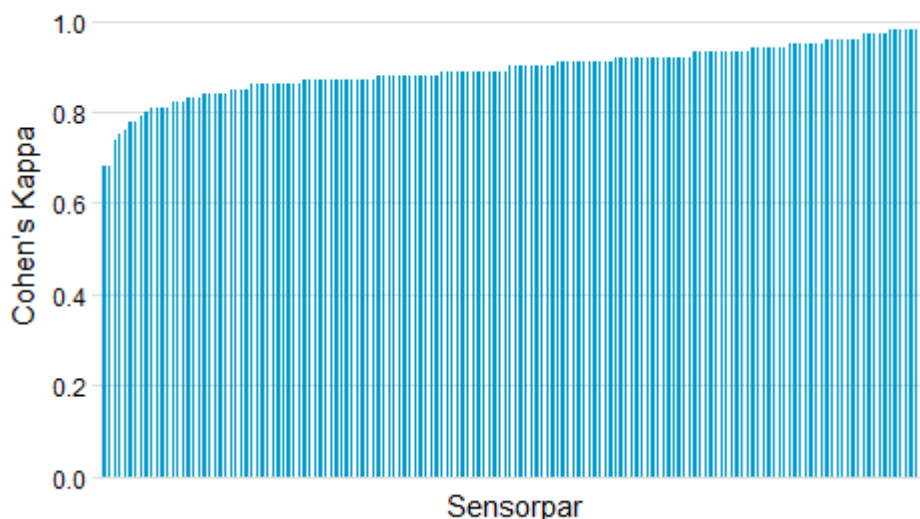
Tabell 3.1 Karakterfordeling per institusjon høsten 2016 (prosent).

4 Sensurreliabilitet

Før vi konkluderer i denne rapporten, vil vi si noe om sensurreliabiliteten. I samarbeid med eksamensgruppen ba NOKUT alle sensorer om å rapportere detaljert sensor. 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 79 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 156 sensorpar, og hvert sensorpar samsensurerte i snitt 31 eksamensoppgaver.

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.⁵ 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 4.1 viser vi kappa-koeffisienten per sensorpar.



Figur 4.1 Sensurreliabilitet: Cohens kappa.

⁵ 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.

Kappa-koeffisienten beveger seg fra 0 (der man ikke er enig på noen oppgaver) til 1 (hvor man er enig på alle oppgaver). Av sensorparene har 95 prosent en kappa-koeffisient på over 0,80. Hvor godt er dette? I forskningslitteraturen opererer man med følgende skala:⁶

- 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

Analysene viser dermed at sensorene i stor grad er betydelig enige om den totale poengsummen. Dette er ikke spesielt overraskende, ettersom eksamensgruppen lagde en meget detaljert sensorveiledning.

5 Konklusjon

Etter pilottesting av nasjonal deleksamen i 2015 ble deleksamenen høsten 2016 gjennomført i fullskala-format som sykepleierstudentenes ordinære eksamen i faget og inkluderte hele 16 institusjoner fordelt på 30 studiesteder.

Resultatene høsten 2016 var noe bedre enn resultatene fra eksamenen fra høsten 2015. Andelen som fikk A eller B, var 26 prosent i 2016, sammenlignet med 19,2 prosent i 2015. Andelen stryk var også lavere i 2016, med 22 prosent i 2016 og 29 prosent i 2015. Den nasjonale gjennomsnittskarakteren for 2016 var en D, men resultatene fra de ulike studiestedene varierte fra en E til en svært god C. Det viste seg også å være en sammenheng mellom karakterpoeng fra videregående skole og antall poeng på nasjonal deleksamen, der de med høyere startkompetanse gjorde det bedre på deleksamenen.

Sensurreliabiliteten viste seg også å være god, noe som tyder på at sensorene som evaluerer den samme eksamenen, i stor grad gir like poeng på oppgavene. Eksamenssettet var også overordnet tilfredsstillende, men noen oppgaver var for enkle (flervalgsoppgavene), eller skilte dårlig mellom studenter med ulike ferdighetsnivåer.

NOKUT har nå gjennomført nasjonal deleksamen i anatomi, fysiologi og biokjemi 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.

⁶ Se for eksempel Landis, J. & Koch, G. (1977). «The Measurement of Observer Agreement for Categorical Data» *Biometrics* 33: 159–174.