

NOKUTs notater

Konsistente karakterer

Er en A en A?

Stephan Hamberg, Sverre Holm og Magnus Strand Hauge, 2017



NOKUT 

Innhold

1	Innledning	1
1.1	Karakterer som kvalitetsindikator	1
2	Nasjonal deleksamen	2
2.1	Bakgrunn	2
2.2	Karakter fra nasjonal deleksamen og studieprogrammet fordelt på studiesteder	3
2.3	Korrelasjoner mellom karakterer på nasjonal deleksamen og studieprogrammet	6
2.4	Hva sier forskningslitteraturen om årsaker til variasjon i karaktersetting?	11
3	Oppsummering og veien videre	13
4	Vedlegg	16
4.1	Supplement tabell 1. Resultater fra sykepleie.	16
4.2	Supplement tabell 2. Resultater fra bachelor i regnskap og revisjon.	17
4.3	Supplement tabell 3. Resultater fra grunnskolelærerutdanningen.	17

1 Innledning

1.1 Karakterer som kvalitetsindikator

Er det mulig å bruke karakterer som kvalitetsindikator i høyere utdanning? Hva skal karakterer være, og hvem skal bruke dem? Fra et kvalitetssikringsperspektiv vil konsistent karakterpraksis gjøre det mulig å bruke karakterer som en kvalitetsindikator. Mange mener kvalitet i seg selv er vanskelig å definere, men det er enda vanskeligere å måle. Ideelt sett vil en høyere utdanningsinstitusjon øke studentenes kunnskapsnivå betydelig med minst mulig ressurser. I en ideell situasjon kjenner vi studenters kunnskapsnivå før de går inn i høyere utdanning, og vi kjenner deres kunnskapsnivå når de fullfører graden sin. Økningen i kunnskapsnivå er da verdiskapingen gitt av institusjonen gjennom utdanningsløpet. Spørsmålet som reises er da om karakterene kan benyttes som en reliabel kvalitetsindikator for å måle dette.

Hovedformålet med dagens karaktersystem i høyere utdanning er at en gitt karakter skal reflektere det samme kunnskapsnivået uavhengig av hvilken institusjon studenten studerer ved. Da vil karakterene gi nøyaktig informasjon om studentens kunnskapsnivå til andre høyere utdanningsinstitusjoner og potensielle arbeidsgivere. Når en A er en A, uansett hvilken institusjon og studiefelt en student har valgt, muliggjør det også å vurdere verdiskapingen en bestemt institusjon bidrar med, så lenge vi kan kontrollere for kunnskapsnivået studentene hadde ved start på institusjonen. Dersom institusjonenes karaktersetting er systematisk forskjellig, vil for eksempel opptak av studenter til masterprogram bli feilaktig og urettferdig. De som har studert ved en institusjon med «snill» karaktersettingspraksis vil da lettere oppnå sine ønsker om studieplass ved at de fortrenger mulighetene til de som er utsatt for «streng» eller «korrekt» karaktersetting. Tilsvarende vil systematiske forskjeller i karaktersettingspraksis svekke karakterenes informasjonsverdi for potensielle arbeidsgivere.

En viktig forutsetning for at eksamenskarakterer skal kunne brukes til å skille mellom studentenes prestasjoner er at karaktersettingspraksisen er lik mellom institusjonene. Det er imidlertid mye som tyder på at bruken av karakterer varierer både mellom institusjoner og fag. Det kan også stilles spørsmål ved hvorvidt karaktersettingen på en enkelt eksamen er reliabel og rettferdig, noe som er viktig både for studentenes rettsikkerhet og institusjonenes troverdighet.

Dette reiser spørsmålet om en A er en A? Det er et uttalt mål at en A gitt ved en institusjon skal tilsvare en A gitt ved en annen institusjon. Det er tidligere dokumentert at dette ikke nødvendigvis er tilfelle. Forskningsrapporter fra Senter for økonomisk forskning (SØF), Møen og Tjelta (2010),

karakterundersøkelser fra UHRs karakterpaneler og tidligere NOKUT analyser viser at det er et stort element av tilfeldighet i karaktersettingen. SØF-rapporten viser blant annet at det gis bedre karakterer ved norske høyskoler og nye universiteter enn på de gamle universitetene. De konkluderer med at karaktersettingen varierer i en slik grad at karakterene sier lite om studentenes faktiske kunnskapsnivå. Dette er ikke bare et norsk fenomen. Studier fra andre Europeiske land (for eksempel De Paola, 2011) og USA (for eksempel Achen og Courant, 2009) viser de samme tendensene.

I dette notatet bruker vi data fra den seneste gjennomføringen av nasjonale deksamener for å belyse problemstillingen ytterligere.

2 Nasjonal deksamnen

2.1 Bakgrunn

Mange land, Norge inkludert, bruker nasjonale eksamener på grunnskole og videregående skole for å vurdere studentens kunnskaper og læringsutbytte. Dette har i liten grad blitt gjort i høyere utdanning. Høsten 2014 ga Kunnskapsdepartementet NOKUT i oppdrag å gjennomføre nasjonale deksamener i tre utdanningstyper: lærerutdanning, sykepleie og regnskap og revisjon. Formålet er å gi universiteter og høyskoler nyttig informasjon om studentenes kunnskapsnivå og la institusjoner få mulighet til å sammenligne seg med andre. Resultatene gir lærestedene mulighet til å utvikle og forbedre utdanningene. I tillegg definerte NOKUT kalibrering av sensur som et ytterligere mål.

Alle lærerutdanninger, sykepleieutdanninger og regnskapsstudier er pålagt å følge nasjonale rammeplaner. De nasjonale rammeplanene definerer læringsutbyttet og gir en liste over kurs studentene må ta. Innenfor rammeplanene gis institusjonene likevel frihet til å skreddersy de enkelte programmene. Dette gjelder blant annet når de skal tilby ulike kurs, når studentene gjennomfører praksis, hvilke lærings- og vurderingsmetoder som brukes og hvilken litteratur som benyttes. Til tross for disse individuelle tilpasningene gjør de nasjonale rammeplanene det mulig å teste alle studenter på et gitt studieprogram i bestemte emner. Det er derfor gjennomført landsdekkende eksamener i disse emnene der alle kandidatene har fått tilsvarende eksamen med målet om å se forskjeller i karakterene mellom institusjoner. Selv om dette er gjennomført som en nasjonal kartlegging, er disse eksamenene for en stor del gjort tellende for studentene.

I dette notatet viser vi analyser av data fra den seneste gjennomføringen av nasjonale deksamener for grunnskolelærerutdanning (mai 2017), regnskap og revisjon (mai 2017) og sykepleie (desember 2016).

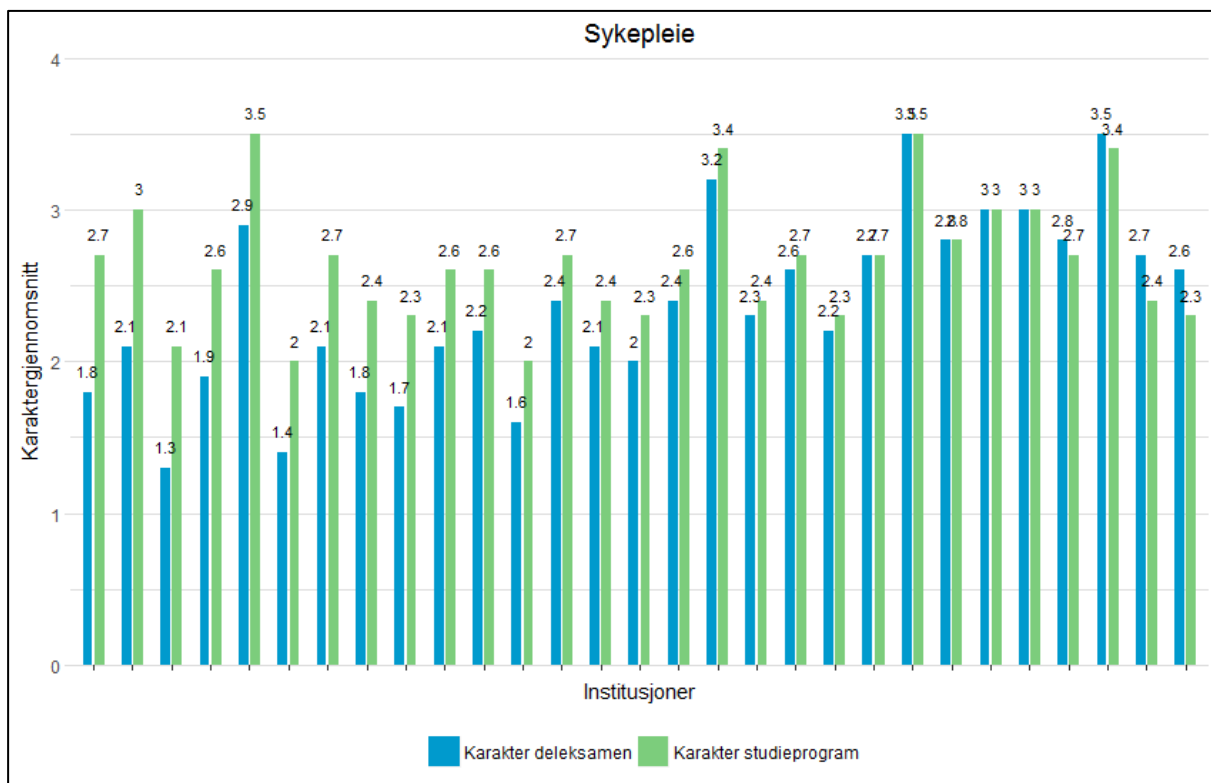
Vi vil belyse i hvilken grad karaktersettingen er konsistent eller varierer – både på tvers av institusjoner og fagområder. Dette gjør vi blant annet ved å sammenholde karakterer fra de nasjonale deleksamenene med karakterer gitt i ordinære eksamener ved studieprogrammene.

2.2 Karakter fra nasjonal deleksamen og studieprogrammet fordelt på studiesteder

Resultatene viser at det er forskjeller i karaktersetting på nasjonal deleksamen og studieprogrammet på de ulike institusjonene. Sammenlagt har om lag halvparten av institusjonene et karaktersnitt på studieprogrammet¹ som ligger en halv karakter eller mer over gjennomsnittet på nasjonal deleksamen. Differanser på mellom en halv og en karakter kan oppfattes som en veldig ulik karaktersetting på institusjonene. Generelt viser forskjellen mellom karakterene at det er karakteren på studieprogrammet som ligger høyere enn karakteren på nasjonal deleksamen. Dette tyder på at på de studiestedene der det er forskjell er det snillere karaktersetting på studieprogrammet sammenlignet med karaktersetting på nasjonal deleksamen. Noen studiesteder har lik karaktersetting, men det er kun 4 som har noe strengere karaktersetting på studieprogrammet og disse ligger innenfor 0,3 karakterpoeng i forskjell.

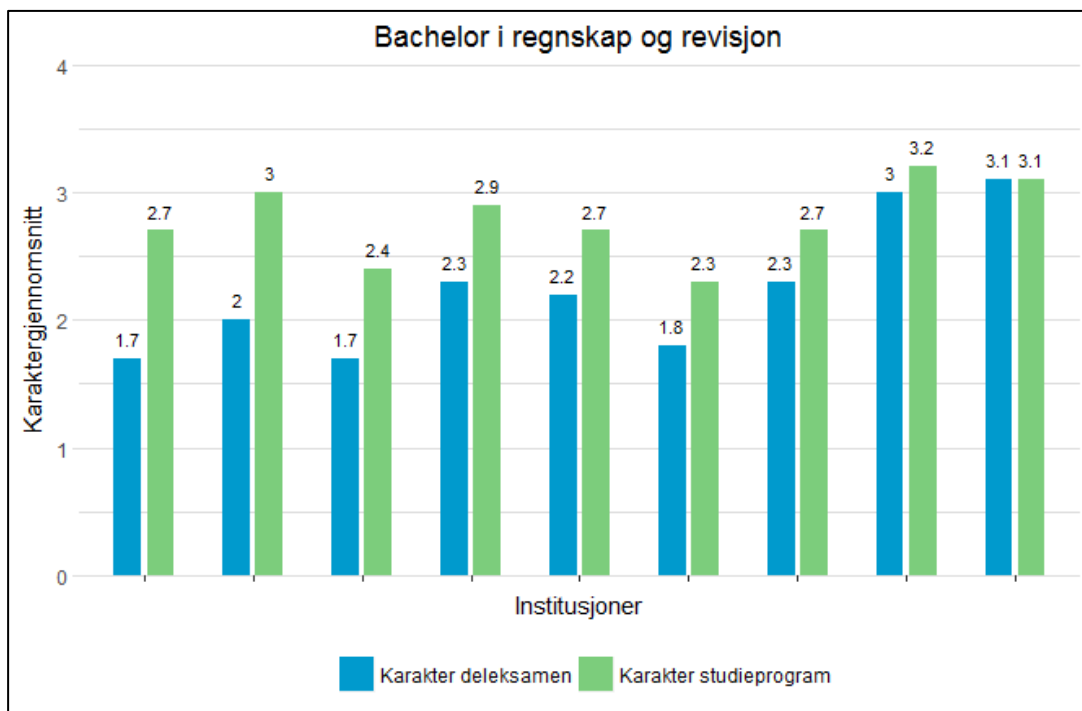
For eksamen for sykepleiestudentene ser vi at forskjellene mellom institusjonene rangerer fra en differanse på nesten en hel karakter (0,9 karakterpoeng), der studieprogramkarakteren er høyest, til lik karaktersetting, og en noe strengere karaktersetting på studieprogrammet (0,3 karakterpoeng differanse). Figur 1 viser karaktersnittene for hvert studiested.

¹ Karaktersnitt på studieprogrammer er gjennomsnittet av studentenes gjennomsnittskarakter på alle emnene de har tatt som inngår i studieprogrammet.



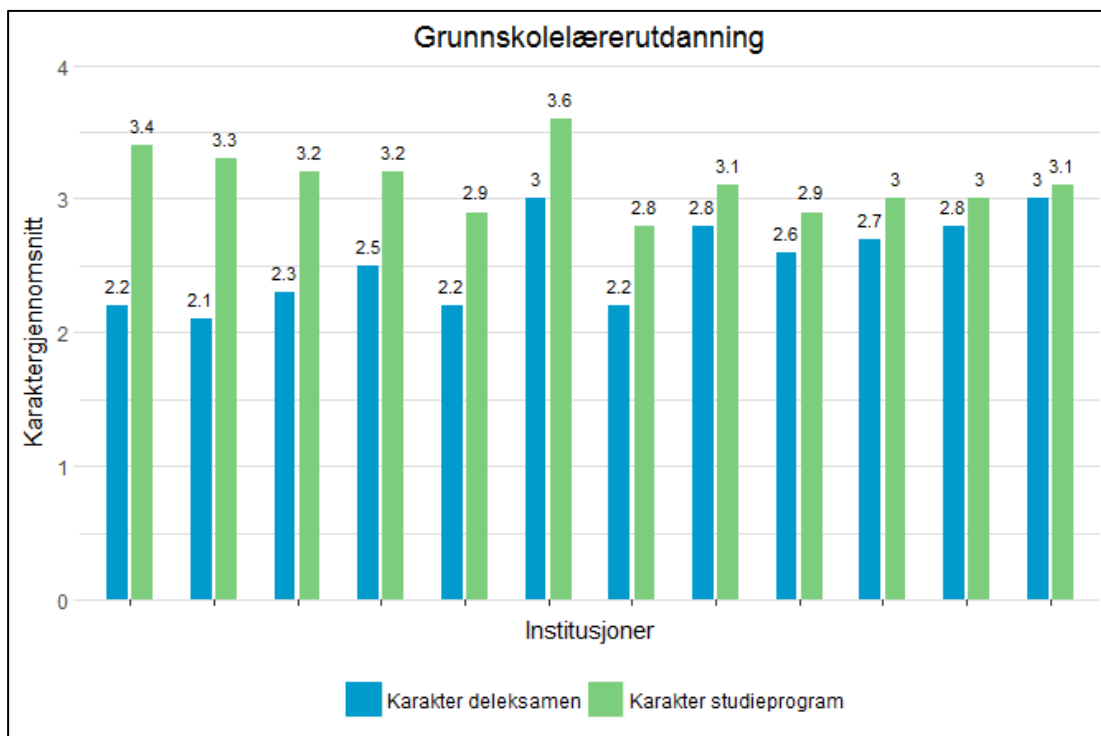
Figur 1. Karaktersnitt for studieprogram og nasjonal deleksamen i sykepleie, der studiestedene med størst differanse er rangert lengst til venstre i figuren.

For bachelor i regnskap og revisjon ser vi også at flere studiesteder har stor forskjell mellom karakterer på nasjonal deleksamen og studieprogrammet (Figur 2). To av studiestedene har en hel karakter forskjell, mens kun ett studiested har helt lik karakter (3,1) og kun tre har en differanse på mindre enn en halv karakter. Igjen er det snillere karaktersetting på studieprogrammet.



Figur 2. Karaktersnitt for studieprogram og nasjonal deleksamen i bachelor i regnskap og revisjon, der studiestedene med størst differanse er rangert lengst til venstre i figuren.

For grunnskolelærerutdanningen har institusjonene med størst differanse hele 1,2 karakterpoeng mer på studieprogrammet enn på nasjonal deleksamen (Figur 3). Fem av institusjonene har en differanse innenfor en halv karakter, mens 7 har en forskjell på over en halv karakter.



Figur 3. Karaktersnitt for studieprogram og nasjonal deleksamen i grunnskolelæreutdanningen, der studiestedene med størst differanse er rangert lengst til venstre i figuren.

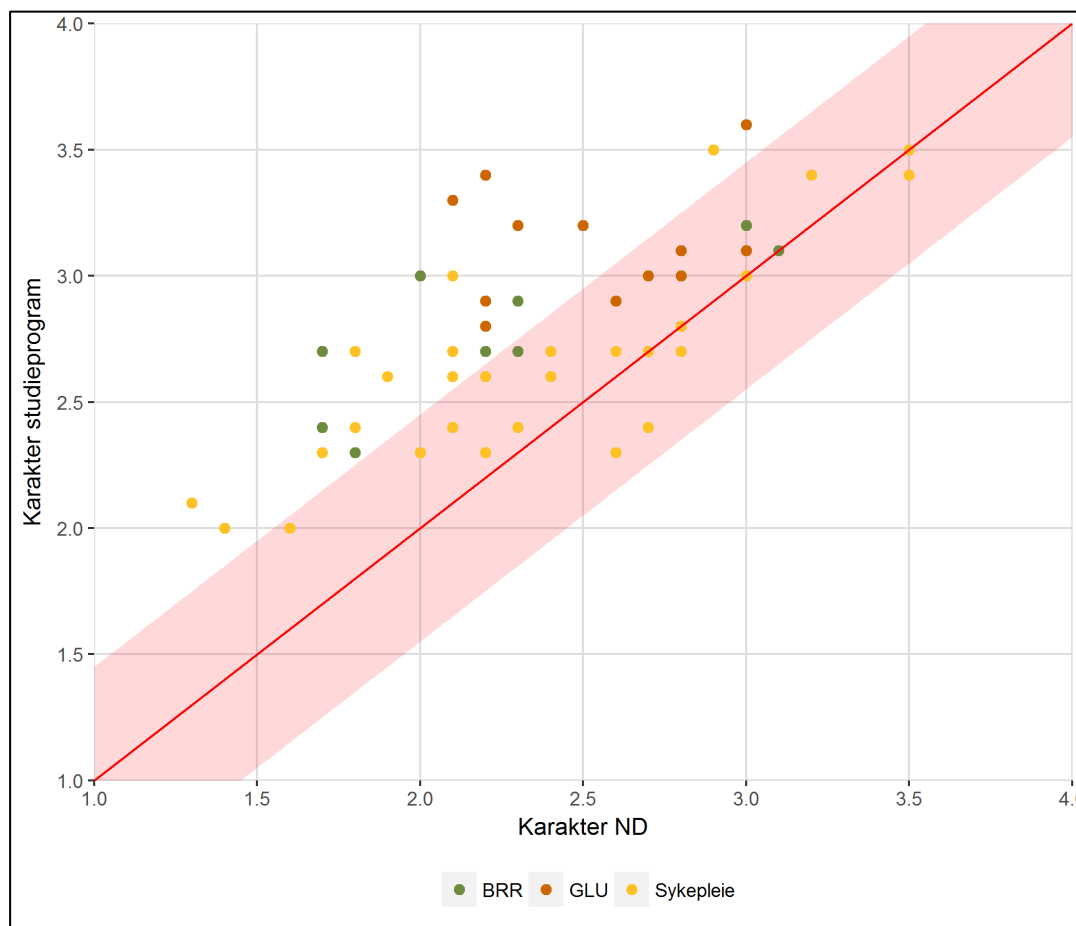
Hvis vi sammenligner de tre nasjonale deleksamenene, observerer vi også forskjeller mellom fagfelt. I nasjonal deleksamen for sykepleierstudenter ser vi at karakterene på studieprogrammet og deleksamen generelt ligger mye nærmere hverandre enn for grunnskolelærerutdanningene og regnskap og revisjon. For sykepleie er det kun 10 av 29 studiesteder som har en differanse på en halv karakter eller mer. For regnskap og revisjon er det 6 av 10 studiesteder og for grunnskolelærerutdanningene 7 av 12. I dette materialet er karaktersetningen jevnere mellom nasjonal deleksamen og ordinær eksamen for sykepleieutdanningene enn for grunnskolelærerutdanningene og regnskap og revisjon.

2.3 Korrelasjoner mellom karakterer på nasjonal deleksamen og studieprogrammet

I Figur 4 plotter vi karakteren på nasjonal deleksamen mot karakteren på studieprogrammet for hver institusjon. Hvert studiested er representert med en farget sirkel i figuren, hvor fargekoden skiller mellom de tre ulike deleksamenene. Her ser vi at karakterene er positivt korrelerte, det vil si at en høyere karakter på studieprogrammet også tilsvarer en høyere karakter på nasjonal deleksamen.

Hvis det hadde vært slikt at en A på nasjonal deleksamen tilsvarte en A på studieprogrammet, ville vi forventet å se en 1 til 1 stigning i karakter. Dette forholdet er representert med en rød linje i Figur 4.

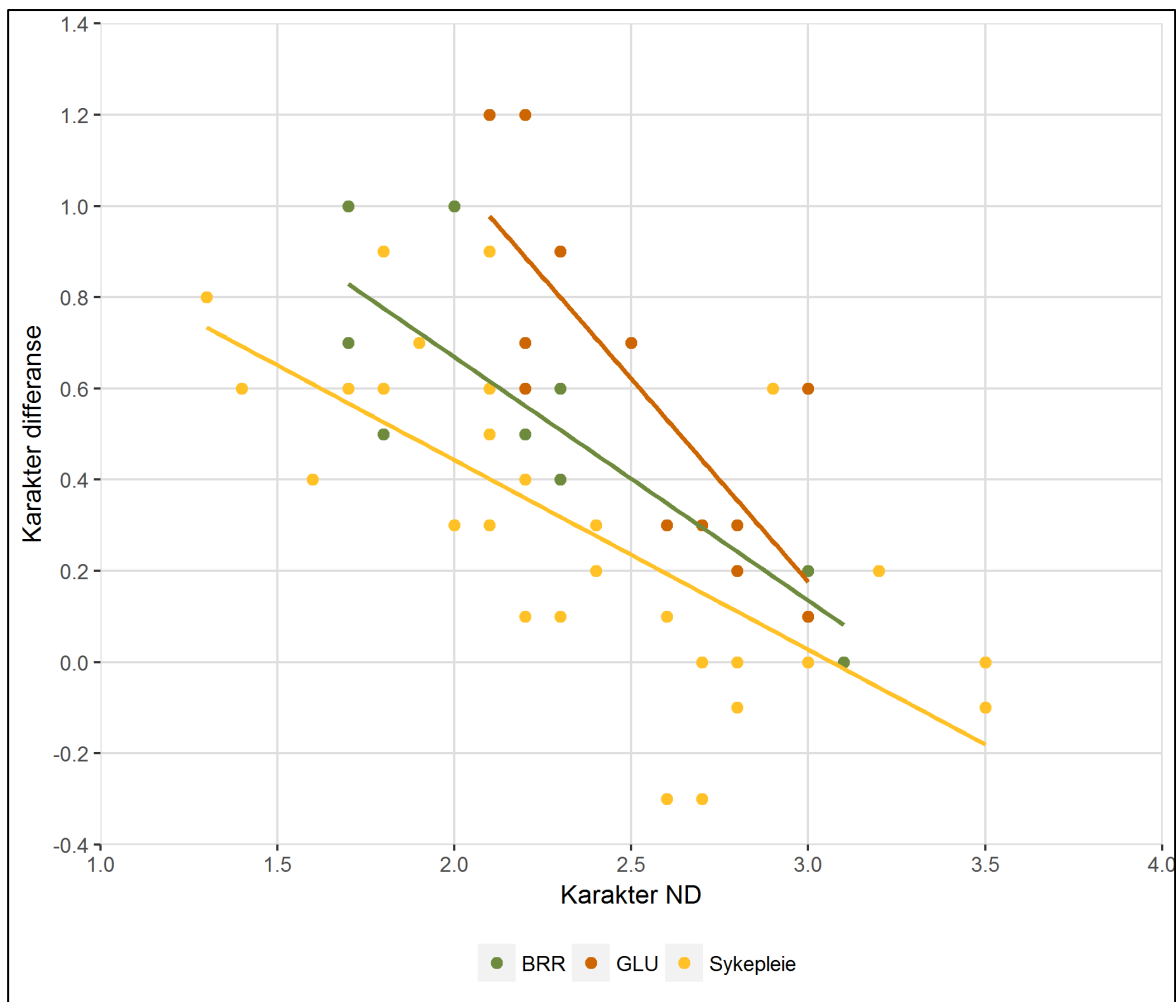
Som vi ser fra figuren, er om lag halvparten (27 av 50) studiesteder på eller rundt denne linjen (skygget område representerer institusjonene med mindre enn en halv karakter i forskjell). Dette vil si at en C i snitt på studieprogrammet også tilsvarer en C i snitt på deleksamen. På den annen side ligger nesten halvparten av institusjonene over denne røde linjen. Det vil si at karakteren på studieprogrammet er høyere enn tilsvarende karakter på nasjonal deleksamen for disse institusjonene.



Figur 4. Spredningsplot av karakterer på nasjonal deleksamen og studieprogrammet for hvert studiested fordelt på tre fagfelt. Den røde linjen viser diagonalen der karaktersetningen er helt lik, og det rosaområdet viser området som har mindre enn en halv karakter i differanse i karaktersetningen.

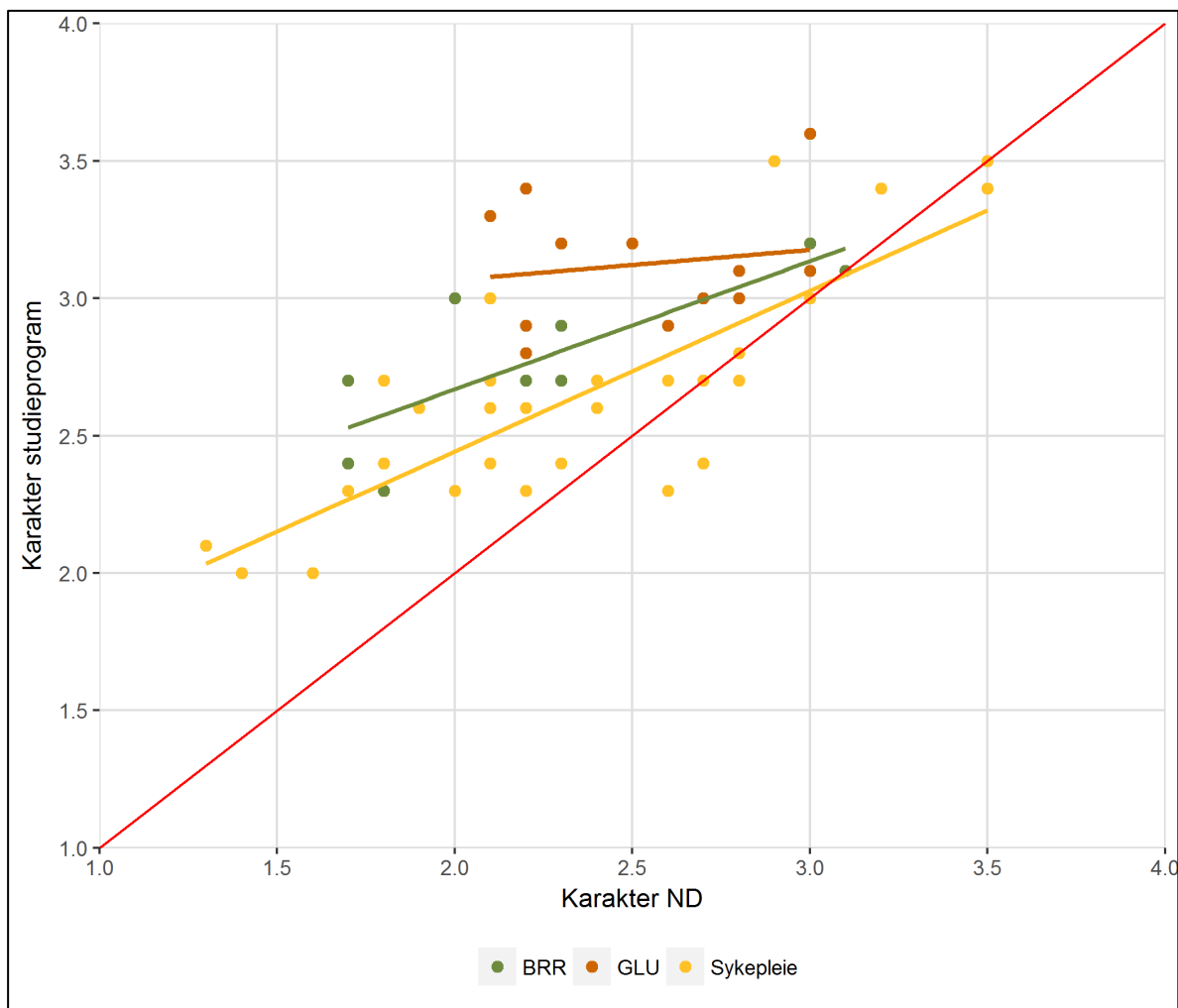
Figuren viser også at studiestedene med høyere karakterer på nasjonal deleksamen (høyere enn 2,5 i snitt, tilsvarende en lav C) har bedre samsvar mellom karakterer på studieprogrammet og nasjonal deleksamen. De studieprogrammene som har en gjennomsnittskarakter under 2,5 på den nasjonale deleksamen har generelt snillere karaktersetning. Dette viser at det er de studieprogrammene som har et lavt snitt på deleksamen som i størst grad har tilsvarende høyere karaktergjennomsnitt på studieprogrammet. Dette er også illustrert i Figur 5, der karakteren på nasjonal deleksamen er plottet mot differansen i karakterene mellom studieprogrammet og deleksamen. Her er det tydelig at

studiestedene med lav score på nasjonal deleksamen også har en stor forskjell i karaktersetting på studieprogrammet. Dette viser at det er ulik karaktersetting mellom studiestedene, spesielt de som scorer lavere på deleksamen, og at en A ikke alltid tilsvarer en A når vi sammenligner på tvers av institusjoner og studieprogrammer.



Figur 5. Spredningsplot av karakterer på nasjonal deleksamen og differansen i karaktersetting med studieprogrammet fordelt på tre fagfelt. Forholdet viser at en lavere karakter på deleksamen har en større forskjell i karaktersetting.

Når vi ser nærmere på fagområdene, ser vi også noen forskjeller. For sykepleie og regnskap og revisjon ser vi at karakterene på nasjonal deleksamen og studieprogrammet har en høy korrelasjon (henholdsvis $r = 0,84$ og $r = 0,81$), der stigningstallet er vist med den gule og grønne linjen i Figur 6. For grunnskolelærerutdanningene er korrelasjonen mellom karakterer på nasjonal deleksamen og studieprogrammet noe svakere ($r = 0,46$), der stigningstallet er vist med den brune linjen i figuren. Videre ser vi at det er sykepleieeksamenene som ligger nærmest den røde linjen og har dermed best samsvar mellom karakterene på studieprogrammet og deleksamen.

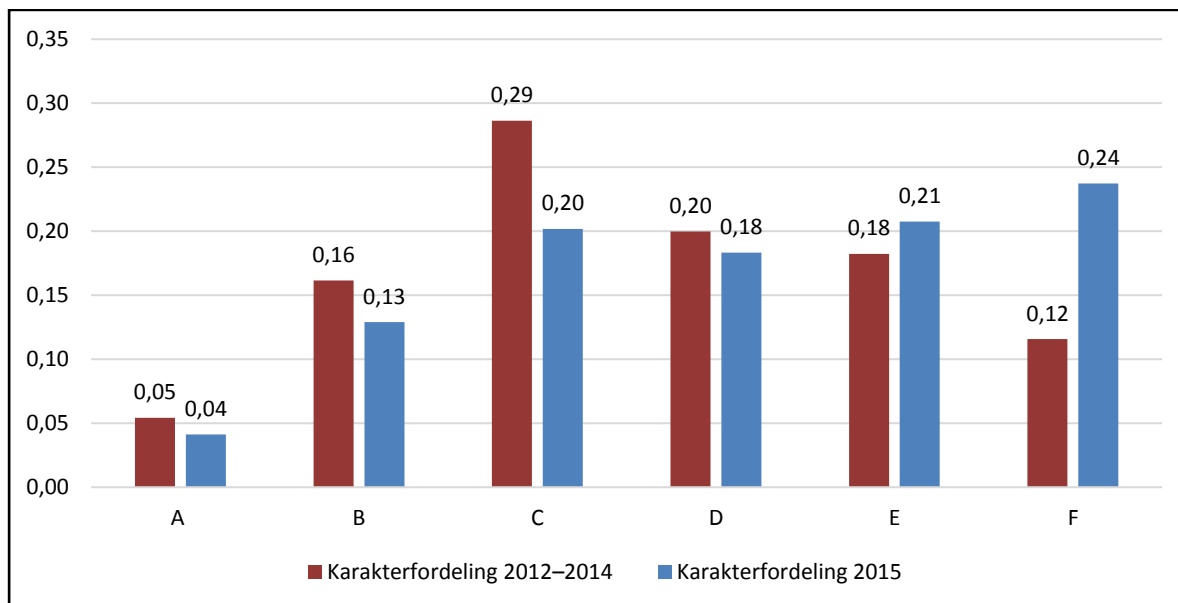


Figur 6. Spredningsplot av karakterer på nasjonal deleksamen og studieprogrammet for hvert studiested fordelt på tre fagfelt. Den røde linjen viser diagonalen der karaktersettingen er helt lik.

Nasjonale deleksamen gir oss også mulighet til å sammenligne resultater på eksamener som er sensurert lokalt med resultater på den nasjonale deleksamen.

Våren 2015 arrangerte NOKUT den første nasjonale deleksamenen i årsregnskap for bachelorstudenter i revisjon og regnskap. I årene før NOKUT fikk ansvar for å arrangere nasjonale deksamener hadde alle institusjoner som tilbyr bachelor i revisjon og regnskap brukt en felles utviklet eksamen i faget, men sensuren ble organisert lokalt. Den viktigste forskjellen våren 2015 i forhold til tidligere år var altså at sensuren nå var nasjonalt organisert og fullstendig blind. Hvordan påvirket dette institusjonenes resultater?

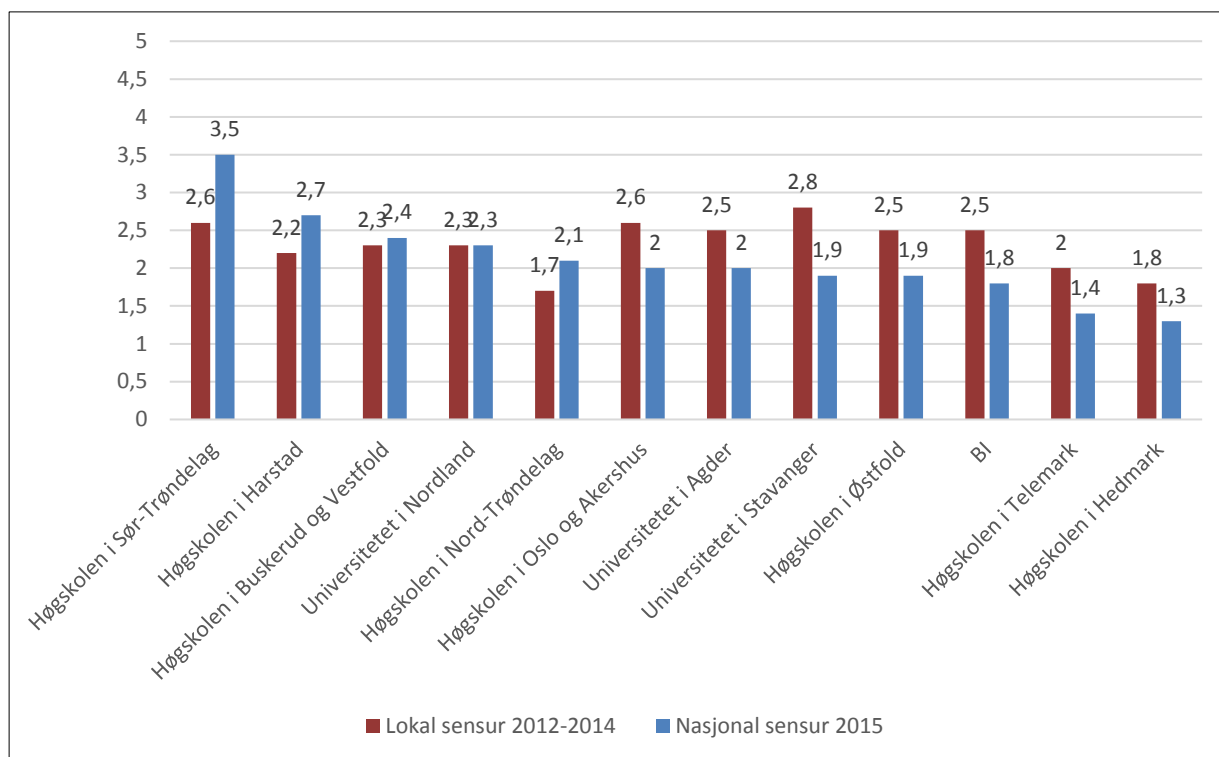
I figur 7, ser vi på den nasjonale karakterfordelingen. I perioden 2012–2014 var gjennomsnittskarakteren på eksamen i årsregnskap 2,4 (en sterk D), mens våren 2015 var den 1,9 (D).



Figur 7. Nasjonal karakterfordeling i nasjonal eksamen i årsregnskap. Figuren viser andelen studenter som fikk A, B, C, D, E og F på den nasjonale eksamen med lokal sensur i perioden 2012-2014 og med nasjonal sensur i 2015.

Som vi ser fra figuren, er karakterene i perioden 2012–2014 nærmest normalfordelt, mens det for 2015 er en mer lineær fordeling fra karakteren A til F. Vi ser også at det var en relativt stor reduksjon i karakteren C (9 prosentpoeng) og en enda større økning i karakteren F (12 prosentpoeng). Den første nasjonale deleksamenen viste altså at karakterene ble ganske mye svakere når all sensur var blind og organisert på nasjonalt nivå. Dette er resultater som stemmer godt overens med for eksempel UHRs karakterrapporter.

Hva hvis vi ser på forskjellen på institusjonsnivå. Er det slik at lokal sensur gir «bedre» resultater enn nasjonal sensur? I figur 8 bruker vi samme tall for å se på dette.



Figur 8. Gjennomsnittskaracter per institusjon. Figuren viser institusjonenes gjennomsnittskaracter på eksamenene med lokal sensur og nasjonal sensur.

Figur 8 viser at de institusjonene som gjorde det best på den nasjonale deleksamenen i 2015 har vært strengere i sin lokale sensur tidligere år. De institusjonene som gjorde det svakest på den nasjonale eksamenen i 2015 er de som relativt sett har vært de snilleste med lokal sensur. Disse resultatene støtter altså godt opp under resultatene over hvor vi bruker gjennomsnittskaracterer i alle fag som sammenligningsgrunnlag.²

Analysene fra nasjonal deleksamen bekrefter tidligere forskning som konkluderer med at det er stor variasjon i karactersettingen ved norske universiteter og høyskoler, men hva forklarer disse variasjonene?

2.4 Hva sier forskningslitteraturen om årsaker til variasjon i karactersetting?

I SØF-rapporten går Strøm m.fl. (2013: 12-17) gjennom både internasjonal og norsk forskningslitteratur på feltet. Kort oppsummert peker den internasjonale forskningen på tre ulike typer av fenomener som årsaker av variasjon i karactersetting kan sorteres under: karakterinflasjon, innføring av resultatbasert finansieringssystem og relativ karactersetting.

² For lignende sammenligninger for sykepleiestudenter og lærerstudenter se tidligere rapporter om resultater fra nasjonal deleksamen.

Karakterinflasjon skjer når gjennomsnittskarakteren nasjonalt stiger over tid, og er et fenomen som er observert og studert blant annet i USA. Det er ulike forklaringer på hvorfor dette skjer. Én er at universiteter gir studenter gode karakterer for at de lettere får gode jobber. Når studenter får gode jobber får institusjonene et bedre rykte. Alle institusjoner bidrar til dette, også prestisjeuniversiteter. En annen forklaring er at det er universiteter med svakere studenter som bidrar til inflasjonen. Disse institusjonene konkurrerer mot hverandre om å gi best karakterer for at arbeidsgiverne skal tro at kvaliteten er høy.

Innføring av resultatbaserte finansieringssystemer kan påvirke institusjonenes karakterpraksis. Strøm m.fl. (2013: 13-14) viser til en undersøkelse i Italia som sier at dette har endret karakterpraksis ved institusjonene der. Samtidig peker de også på at Tyskland også har innført et resultatbasert finansieringssystem uten at det ser ut til at det har påvirket karakterpraksis der.

Forskningen finner også at **relativ karaktersetting** er utbredt i mange land. I slike tilfeller vil årsaken til variasjon i karaktersetting være variasjon i inntakskvaliteten på institusjonene/studiene. Institusjoner med høy inntakskvalitet vil ha en strengere karakterpraksis enn institusjoner med lav inntakskvalitet. Strøm m.fl. (2013: 14-15) viser også til flere studier som sier at streng karaktersetting kan ha positiv effekt på studentenes prestasjoner, men at effekten varierer avhengig av hvor flinke studentene er i utgangspunktet. Det kan være positivt for sterke studenter som i større grad blir motivert til ytterligere innsats hvis det er vanskelig å få gode karakterer, mens svake studenter har en tendens til å bli demotivert av streng karaktersetting. Hvis man tar utgangspunkt i den gjennomsnittlige innsatsen til studentene, er det imidlertid en fare for at den kan bli redusert når karaktersettingen er relativ fordi alle kan oppnå samme karakter med mindre innsats.

Når vi dreier blikket mot Norge så er det lite som tyder på at vi har hatt en nasjonal karakterinflasjon her etter innføring av nytt karaktersystem i 2003. I UHRs rapport *Karakterbruk i UH-sektoren 2015* (2012) finner de kun små endringer i karakterpraksis nasjonalt fra 2004 til 2015. Tall fra NSD/DBH viser også kun minimale endringer fram til 2015.

Når det gjelder endringer som følge av innføring av nytt finansieringssystem, så finner Frølich og Strøm (2008), gjennom en spørreundersøkelse sendt til ansatte ved universiteter og høyskoler i 2005, en større grad av intensjon blant ansatte om å la flere studenter stå på eksamen på høyskoler enn på universiteter. Denne intensjonen er høyest på høyskoler med lavt antall primærstøttere per studieplass. Møen og Tjelta (2010) ser på karaktersetting i økonomisk-administrative studier og finner at studenter fra høyskoler med lavt inntakskrav og studenttall systematisk gjør det dårligere på den siste delen av siviløkonomistudiet ved NHH, enn studenter med samme karakter som kommer fra høyskoler med høyere inntakskrav og større studenttall. At lave inntakskrav kan føre til at studentene får bedre

karakterer enn forventet finner man også indikasjoner på i en analyse NOKUT har gjort på data fra Studiebarometeret (Hauge m.fl. 2017). Her finner man mange av de samme institusjonseffektene som i SØF-rapporten, men når man inkluderer gjennomsnittlig antall karakterpoeng på studieprogramnivå som avhengig variabel i analysen så reduseres forskjellen i karaktersetting mellom institusjonene. At inntakskvaliteten på et studium påvirker karakterene til studentene er et tegn på at det forekommer relativ karaktersetting.

3 Oppsummering og veien videre

Diskusjoner rundt karaktersetting i norsk høyere utdanning har opptatt sektoren i til dels stor grad siden kvalitetsreformen i 2003. Forskning på karaktersetting i Norge og analysene i dette notatet viser at en A ikke er en A. Norske retningslinjer tilsier at karakterer i norsk høyere utdanning skal være kriteriebasert og ikke normbasert. Det vil si at vurderingen av studentenes prestasjoner skal knyttes opp mot bestemte faglige kriterier og ikke opp mot hvor gode andre studenter er. Kriteriebasert vurdering kan i praksis bety at man innenfor noen studier, institusjoner og kull får veldig store andeler A eller F. Det er ingen garanti for at kriteriebasert vurdering vil føre til en normalfordeling av karakterer. Dersom fordelingen av studenter på studieprogram og institusjoner hadde vært helt tilfeldig på bachelornivå, så kan man i teorien tenke seg at karakterfordelingen på bachelornivå hadde blitt normalfordelt selv med kriteriebasert vurdering. Slik er ikke virkeligheten. Det er til dels stor og systematisk forskjell på inntakskvaliteten på institusjoner og studieretninger i Norge i dag. Noen institusjoner og studieretninger rekrutterer sterkere studenter enn andre institusjoner og studieretninger. Det betyr at når man har kriteriebasert vurdering, burde de institusjonene og studieretningene som har rekrutterer de sterkeste studentene ha de beste resultatene og de institusjonene og studieretningene som rekrutterer de svakeste studentene ha de dårligste resultatene, målt i karakterer. I virkeligheten er det ikke det som skjer. Analysene i dette notatet og i forskningen rundt norsk karaktersetting generelt viser at situasjonen nesten er snudd på hodet. Institusjoner og studieretninger som rekrutterer svakere studenter har like gode eller bedre resultater enn institusjoner og studieretninger som rekrutterer sterke studenter. For å si det på en tabloid måte: En del institusjoner gjør gull av gråstein.

Situasjonen som er beskrevet i dette notatet er en situasjonsbeskrivelse som har vart i mange år, og det til tross for at problemstillingen har blitt tatt opp med jevne mellomrom. Det spørsmålet vi må stille oss er hvor alvorlig problemet er, hvem dette er et problem for, hvor mye tid og ressurser vi skal bruke for å løse problemet og til slutt om det er bedre å bruke tid og ressurser på andre ting som øker kvaliteten i utdanningen og sikrer studentenes læring.

Over tid har flere virkemidler blitt tatt i bruk for å redusere tilfeldigheten i karaktersettingen og for å sikre at en A er en A. Det har lenge vært krav om tilsynssensorer, man bruker eksterne sensorer og UHR har drevet med karakterpaneler i flere år. I sin siste karakterrappport foreslår UHR mer bruk av tilsynssensorer og sensorfelleskap samt mer bruk av sensorveiledninger. I så henseende har Kunnskapsdepartementet sendt et lovforslag om krav til sensorveiledninger på høring. Felles for disse forslagene er at de i stor grad kan hjelpe til å redusere tilfeldigheten i karaktersetting på lokalt plan, men de kan i mindre grad hjelpe til å sikre at å få A på en institusjon krever det samme som å få A på en annen institusjon.

Dersom man faktisk ønsker å sikre at alle studenter skal vurderes helt likt etter de samme kriteriene, så må man organisere alle vurderinger på et nasjonalt plan med blind sensur. Dette er neppe ønskelig. Det vil være negativt for studentenes læring, det vil gå utover den akademiske friheten og det vil koste både mye tid og mye penger.

Før vi setter i gang nye tiltak for å sikre at en A er en A, uansett, må vi derfor spørre oss selv hvor alvorlig situasjonen er, hvem problemet rammer og hvor mye tid og ressurser vi skal bruke.

4 Litteratur

- Achen, A. C. og P. N. Courant. (2009). "What are grades made of?" *Journal of Economic Perspectives*, 23 (3), 77-92.
- De Paola, M. (2011). Easy grading practices and supply-demand factors: evidence from Italy, *Empirical Economics*, 41, 227-246.
- Frølich, N. og Strøm, B. (2008): Higher education funding and incentives: Evidence from the Norwegian funding reform. *European Journal of Education*, 43, s. 555-576.
- Hauge, M.S., Skåtun, K.C. og Lid, S.E. (2017): Grades as predictors of student performance. *Paper presentert på EAIR 2017*. Porto.
- Møen, J. og Tjelta, M. (2010): Grading standards, student ability and errors in college admission, *Scandinavian Journal of Educational Research* 54, s. 221-237.
- Strøm B., Falch, T., Gunnes, T. og Haraldsvik, M. (2013). Analyse av karakterbruk og kvalitet i høyere utdanning. SØF-rapport 3/13. [online] http://www.sof.ntnu.no/SOF-R%2003_13.pdf
- UHR. (2015): Karakterbruk i UH-sektoren 2015. [online] http://www.uhr.no/documents/Karakterrapport_2015_endelig_versjon.pdf

5 Vedlegg

5.1 Supplement tabell 1. Resultater fra sykepleie.

Institusjon	Antall med data	Karakterp oeng vgs	karakter ND	Karakter studprog	Differanse
HSH - Stord	81	4.1	1.8	2.7	0.9
HSH - Haugesund	94	4	2.1	3	0.9
UiT - Harstad	15	3.8	1.3	2.1	0.8
UiS	260	4.2	1.9	2.6	0.7
LDH	225	4.5	2.9	3.5	0.6
Nord - Mo i Rana	73	3.8	1.4	2	0.6
NTNU - Gjøvik	169	4	2.1	2.7	0.6
HiOA - Sandvika	29	4.1	1.8	2.4	0.6
Nord - Bodø	157	3.9	1.7	2.3	0.6
HSN - Porsgrunn	125	4	2.1	2.6	0.5
Nord - Vesterålen	17	3.9	2.2	2.6	0.4
UiT - Narvik	48	3.7	1.6	2	0.4
NTNU - Ålesund	120	4.2	2.4	2.7	0.3
HiM	159	4	2.1	2.4	0.3
HSN - Drammen	169	4.3	2	2.3	0.3
HiOA - Pilestredet	400	4.2	2.4	2.6	0.2
HiSF	68	4.2	3.2	3.4	0.2
HSN - Vestfold	118	4.1	2.3	2.4	0.1
Nord - Levanger	164	4.1	2.6	2.7	0.1
UiT - Hammerfest	56	3.5	2.2	2.3	0.1
HD	121	4.2	2.7	2.7	0
HiB	172	4.6	3.5	3.5	0
HiHm	219	3.9	2.8	2.8	0
UiA	251	4.2	3	3	0
UiT - Tromsø	110	4.4	3	3	0
VID - Oslo	173	4.2	2.8	2.7	-0.1
NTNU - Trondheim	187	4.7	3.5	3.4	-0.1
HiOA - Kjeller	154	4.2	2.7	2.4	-0.3
HiØ	145	4	2.6	2.3	-0.3

Resultater fra nasjonal deleksamen i sykepleie. Institusjonene med størst karakterdifferanse mellom studieprogrammet og nasjonal deleksamen er rangert øverst. Fargekodene indikere en differanse på mer enn 1 (rød), 0,5-0,9 (oransje), 0,1-0,4 (gul), 0 (lysegrønn), og -0,1—0,4 (mørkegrønn).

5.2 Supplement tabell 2. Resultater fra bachelor i regnskap og revisjon.

Institusjon	Antall med data	karakterpoeng vgs	karakter ND	Karakter studieprogram	Differanse
NORD - Bodø	26	3.9	1.7	2.7	1
UiS	48	4.2	2	3	1
HiØ	41	3.6	1.7	2.4	0.7
UiT - Harstad	13	3.7	2.3	2.9	0.6
HSN - Vestfold	63	4.1	2.2	2.7	0.5
INN - Hedmark	81	3.8	1.8	2.3	0.5
HSN - Telemark	31	3.6	2.3	2.7	0.4
HiOA	35	4.7	3	3.2	0.2
NTNU - Trondheim	35	4.6	3.1	3.1	0

Resultater fra nasjonal deleksamen i bachelor i regnskap og revisjon. Institusjonene med størst karakterdifferanse mellom studieprogrammet og nasjonal deleksamen er rangert øverst. Fargekodene indikere en differanse på mer enn 1 (rød), 0,5-0,9 (oransje), 0,1-0,4 (gul), 0 (lysegrønn), og -0,1—0,4 (mørkegrønn).

5.3 Supplement tabell 3. Resultater fra grunnskolelærerutdanningen.

Institusjon	Antall med data	karakterpoeng vgs	karakter ND	Karakter studieprogram	Differanse
HSN - Drammen	30	4	2.2	3.4	1.2
HVL - Stord	32	4	2.1	3.3	1.2
NLA - Bergen	50	4.1	2.3	3.2	0.9
HSN - Porsgrunn	24	4.1	2.5	3.2	0.7
NORD - Bodø	19	4.2	2.2	2.9	0.7
HSN - Vestfold	28	4.5	3	3.6	0.6
HVL - Bergen	190	4.1	2.2	2.8	0.6
HiOA	163	4.3	2.8	3.1	0.3
NORD - Levanger	34	4.1	2.6	2.9	0.3
NTNU	139	4.3	2.7	3	0.3
UiA	168	4.3	2.8	3	0.2
UiT - Tromsø	27	4.1	3	3.1	0.1

Resultater fra nasjonal deleksamen i grunnskolelærerutdanning. Institusjonene med størst karakterdifferanse mellom studieprogrammet og nasjonal deleksamen er rangert øverst. Fargekodene

indikere en differanse på mer enn 1 (rød), 0,5-0,9 (oransje), 0,1-0,4 (gul), 0 (lysegrønn), og -0,1—0,4 (mørkegrønn).