

Rapport fra nasjonal kommisjon for vurdering av karaktersetting på masteroppgaver i statsvitenskap, avgitt 10. desember 2007

Innledning

Nasjonalt fagråd i statsvitenskap fungerer som karakterpanel innen disiplinen. Karakterpanelet skal følge med i lærestedenes bruk av bokstavkarakterene, og bidra til å utvikle en felles forståelse av karakterskalaen. Så langt har fagrådet diskutert karaktersettingen en gang i året på basis av statistisk materiale fra DBH. I brev av 27.04.07 fra Universitets- og høgskolerådet (UHR) ble fagrådet bedt om i inneværende år å utføre en mer kvalitativ undersøkelse av et begrenset antall masteroppgaver, med det formål å få et bedre bilde av "hvilke vurderinger og hvilken forståelse som ligger til grunn for karaktersettingen".

I sitt møte 25.09.07 gikk Nasjonalt fagråd inn for å opprette en nasjonal kommisjon for å vurdere 10 masteroppgaver nærmere. Kommisjonen skulle bestå av "representanter fra lærestedene som har master i statsvitenskap, inkludert Høgskolen i Bodø, og institusjonene har ansvar for å stille med en representant". Videre sies referatet i sak 13/2007 at "institusjonene må selv dekke honoraret til sitt komitémedlem, etter sine egne satser. Bjørn Erik Rasch samordner arbeidet." Punktet om honorering av kommisjonsmedlemmer er understreket også i et eget skriv fra UHR av 24.04.07, og bakgrunnen er at "arbeidet er viktig for å oppnå en felles forståelse av skalaen". Dette elementet skal bidra til at alle medlemmene av kommisjonen vier arbeidet tilstrekkelig tid og oppmerksomhet.

Kommisjonen har fått følgende sammensetning:

Professor **Bjørn Erik Rasch**, Universitet i Oslo, nestleder i Nasjonalt fagråd

Førsteamanuensis **Marcus Buck**, Universitetet i Tromsø

Førsteamanuensis **Tanja Ellingsen**, NTNU

Førsteamanuensis **Jan Froestad**, Institutt for administrasjon og organisasjonsvitenskap, Universitet i Bergen

Redaktør **Jan-Erik Grindheim**, *Stat og styring*; representant for Institutt for sammenliknende politikk, Universitetet i Bergen

Førsteamanuensis **Asbjørn Røiseland**, Høgskolen i Bodø

Førsteamanuensis **Morten Øgård**, Universitetet i Agder

Selv om institusjonene har stått helt fritt ved valg av medlemmer til kommisjonen, har den fått en så bred sammensetning at alle hovedområdene i faget må sies å være dekket.

Opplegg for undersøkelsen

I brev fra UHR av 27.04.07 er et opplegg for kvalitativ undersøkelse av karaktersetningen beskrevet. Det er gjort enkelte endringer i det opplegget som UHR beskriver (jf. fagrådets diskusjon under sak 13/2007 den 25.09.07). Grunnen er at resultatene fra undersøkelsen dermed kan bli av noe større verdi for fagmiljøet. Følgende fremgangsmåte er lagt til grunn:

1. Det er valgt ut *10 masteroppgaver* fra masterprogrammer i statsvitenskap ved 6 læresteder. (UHRs opplegg gikk ut på å velge 5 oppgaver fra ett studieprogram.)
2. Kommisjonen har konsentrert seg om oppgaver som er blitt bedømt til B, hvilket innebærer at det blir belyst om det finnes en rimelig felles forståelse av denne hyppig brukte karakteren – en felles forståelse både *innad i den bredt sammensatte kommisjonen og mellom de ulike lærestedene* hvor sensuren av masteroppgavene opprinnelig foregikk. (UHR antydte et opplegg der kommisjonen ikke skulle kjenne til karakterene på forhånd.)
3. Lærestedene fremskaffet lister over masteroppgaver i statsvitenskap som oppnådde karakteren B i vårsemesteret 2007 (samlet 54 oppgaver). Det ble deretter foretatt et *tilfeldig utvalg* av to masteroppgaver hver fra Admorg, Sampol, UiO og NTNU og én oppgave hver fra UiT og UiA. Ikke noe kommisjonsmedlem var involvert i oppgavevalget fra eget universitet. (Noen av instituttene har masterprogrammer som ikke inngår i undersøkelsen, som organisasjons- og ledelsesvitenskap ved UiT og PECOS ved UiO.)
4. Hvert enkelt kommisjonsmedlem har bedømt alle oppgavene (individuelt), og konkludert ved å plassere hver av dem på en fingradert skala; kommisjonen har *ikke* diskutert seg frem til en felles eller samlet karakter på hver av oppgavene, slik en bedømmelseskommisjon ellers gjør. Men hvert kommisjonsmedlem har sendt inn sitt endelige karakterforslag etter at en først hadde begrunnet oppgavene skriftlig, slik at det var mulig å justere eget syn i lyse av det andre hadde skrevet. Dette ble likevel i liten grad gjort. Det er dette materialet i anonymisert form som danner grunnlag for sluttrapporten. (Den "desentraliserte" fremgangsmåten er i all hovedsak i samsvar med UHRs anbefaling.)
5. Kommisjonen har gjort alt arbeid via e-post. Det gjelder alle trinnene fra diskusjon av kommisjonens arbeidsmåte og fremsettelse av begrunnede vurderinger av de enkelte oppgavene, til analyse av materialet og utformingen av denne sluttrapporten. (UHR

har anbefalt at kommisjonen skulle tre sammen for å analysere resultatene etter at hvert enkelt medlem hadde foretatt sin begrunnede vurdering, men sterkt tidspress og manglende ressurser har umuliggjort dette.) Medlemmenes detaljerte vurderinger knyttet til hver enkelt masteroppgave finnes i et eget fortrolig notat på 50 sider.

6. Rapporten er ført i pennen av Marcus Buck, Tanja Ellingsen og Bjørn Erik Rasch, men alle medlemmer har kommet med innspill i prosessen. Alle kommisjonsmedlemmene har sluttet seg til hovedsynspunktene i rapporten, og den kan dermed å anses å være enstemmig. Underveis er det også diskutert en rekke forhold som det ikke er funnet rom for å gå nærmere inn på i rapporten. Videre finnes det som nevnt et lengre, fortrolig underlagsnotat som inneholder detaljerte begrunnelser til hver enkelt oppgave fra medlemmer av kommisjonen, hvor det synliggjøres hvordan det er tenkt i forhold til alle oppgavene.

Karakterfordelingen 2006

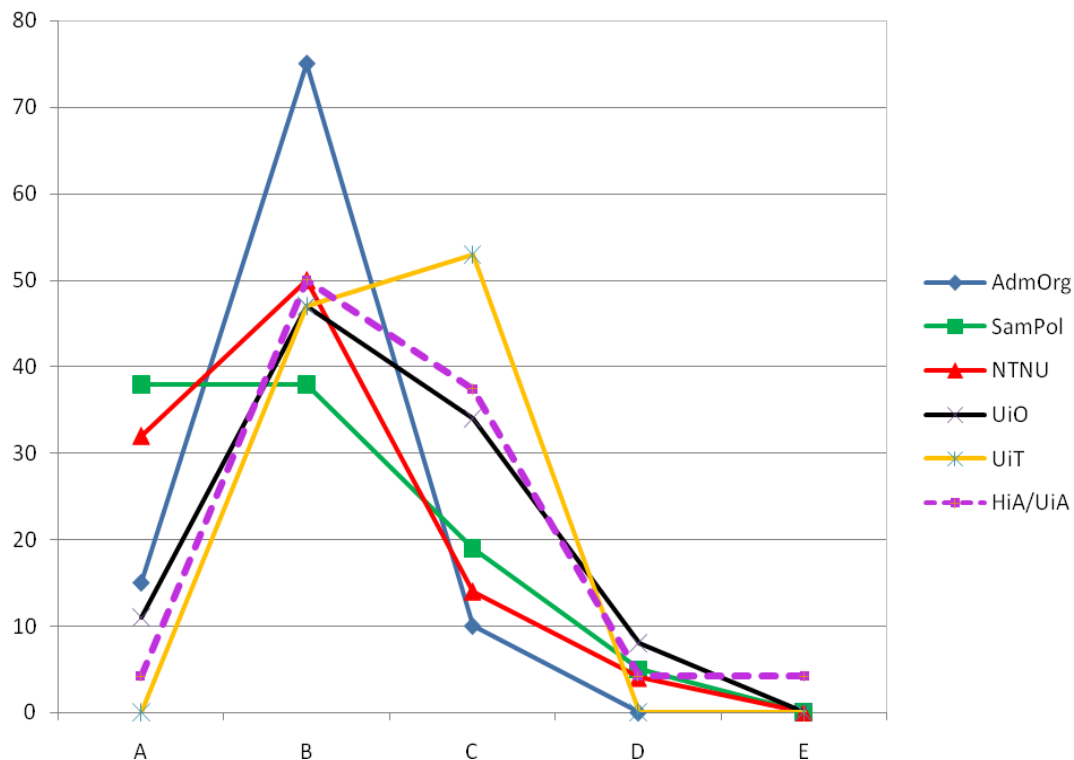
Som nevnt er det kun oppgaver som har fått karakteren B som er plukket ut (tilfeldig) og vurdert denne gangen. Dette har sammenheng med at det er en meget skjev karakterfordeling på masteroppgaver i faget, med svært mye gode karakterer. Ser vi på karakterstatistikken for 2006 (sak 13/07 i Nasjonalt fagråd), var B ("meget godt") den hyppigst brukte karakteren. Halvparten av masteroppgavene oppnådde denne karakteren, mens 15 prosent oppnådde karakteren A ("fremragende"). Dette innebar at nær $2/3$ av alle masteroppgavene sikret seg en av de to toppkarakterene. C-en ble brukt på rundt 30 prosent av oppgavene, og D bare i 5 prosent av tilfellene. Det var kun én E i det statistiske materialet som er tatt ut av DBH for 2006.¹

Det er en god del variasjon i karakterfordelingene lærestedene i mellom. Når det gjelder B-ene ligger Institutt for administrasjons- og organisasjonsvitenskap (AdmOrg) ved UiB på topp med 75 prosent, og Institutt for sammenlignende politikk (SamPol) ved samme universitet nederst med 38 prosent. De øvrige instituttene ligger i nærheten av 50 prosent B-er. Det er også markante forskjeller i bruken av karakteren A. På topp i anvendelsen av A ligger SamPol og NTNU med henholdsvis 38 og 32 prosent, mens AdmOrg har 15 prosent og UiO 11 prosent. Det ble ikke gitt A på masteroppgaver i statsvitenskap ved Universitetet i Tromsø i 2006, og det var bare én ved Høgskolen i Agder (nå UiA). Hvis en ser A og B i sammenheng,

¹ Det synes å være noen mindre feil eller mangler i dette materialet, men dette rokker ikke ved hovedmønsteret som er referert.

finder vi at en av disse karakterene er gitt på 90 prosent av alle oppgaver ved AdmOrg, på 82 prosent av oppgavene ved NTNU, på 76 prosent av oppgavene ved SamPol, på 58 prosent av oppgavene ved UiO, 54 prosent ved HiA og 47 prosent av oppgavene ved UiT. Dette er gjennomgående høye tall, men det er også betydelig variasjon.

Det er i alt 193 oppgaver som inngår i det statistiske materialet. Hele 91 av oppgavene er bedømt ved UiO. De øvrige lærestedene har hatt rundt 20 oppgaver hver (kun 15 ved UiT og 24 ved HiA). Vi har ikke laget en statistisk oversikt for 2007. Nasjonalt fagråd har tidligere behandlet karakterstatistikk for 2004 og 2005. Disse årene gir mye av det samme bildet som 2006 når det gjelder karakterfordelingen på masteroppgaver.



Figur 1. Karakterfordeling på masteroppgaver i statsvitenskap 2006.

Det spørsmålet som melder seg er om masteroppgavene i faget virkelig holder et så høyt faglig nivå som statistikken indikerer, eller om karakterskalaen er brukt ulikt av de kommisjonene som har vært involvert i sensuren på oppgavene. Formålet med undersøkelsen har vært å få et første – om enn ganske begrenset – inntrykk av om dette er tilfellet. Det er som nevnt valgt ut tilfeldig 10 masteroppgaver som er bedømt våren 2007,

og alle universiteter som i dag har masterprogrammer i statsvitenskap (eller en av fagets underdisipliner) er representert med oppgaver. Tyder gjennomgangen av det lille utvalget med masteroppgaver på at det ligger en felles forståelse til grunn for karaktersettingen i faget?

Meningsutveksling underveis

Kommisjonen har altså vurdert 10 masteroppgaver. I praksis har det skjedd på den måten at det enkelte medlem har gjort seg opp en mening om oppgavene. Underveis er relativt utførlige begrunnelser utvekslet på e-post (og disse er samlet i et eget, fortrolig dokument). Det er godt mulig at den første begrunnelsen som er gitt i tilknytning til en oppgave, gitt at den har vært fyldig og detaljert, kan ha påvirket vurderingen og karaktersettingen til andre medlemmer av gruppen. Dette gjelder kanskje særlig i de tilfellene hvor det er kommet et tidlig innspill fra et kommisjonsmedlem med spesialkompetanse på det forskningsfeltet masteroppgaven behandler. Dette betyr at spredningen i vurderinger eller karaktersetting innad i komiteen kanskje med dette blir mindre enn den ville vært dersom kommisjonsmedlemmene hadde jobbet helt på egen hånd i første fase av arbeidet. På den annen side har utvekslingen av vurderingsmomenter - gitt den store spredningen i oppgavetemaer og analysemetoder - gitt et mye bedre fundament for å si om den opprinnelige karakteren var rimelig eller ikke, enn det som ville vært tilfelle hvis hvert medlem satte karakterer helt isolert (uten i det hele tatt å kjenne til hvordan andre medlemmer resonnerer). Alt i alt mener vi derfor at den tilnærmingen som ble valgt har vært hensiktsmessig.

Avvik fra ordinær sensur

Medlemmene av kommisjonen har bedømt oppgavene som er valgt ut etter beste evne, og søkt å angripe dette på samme måte som ved ordinær masteroppgavesensur. Samtidig har vårt arbeid klare avvik i forhold til vanlig sensur.

For det *første* betyr ikke konklusjonene vi trekker noe som helst for kandidatene det gjelder. De har allerede fått sin karakter, og blir ikke gjort kjent med at oppgavene deres er drøftet på nytt i det hele tatt. Slik sett er vårt arbeid langt mer "uforpliktende" enn sensur normalt er.

For det *andre* møter vi verken veileder eller kandidat. Ved alle lærestedene er det formøter mellom veileder og eksamenskommisjon. Flere av lærestedene har også muntlig som del av eksamen på masteroppgaveemnet. Det innebærer ikke bare at kommisjonen møter kandidaten personlig, blant annet for å fremføre begrunnelse av karaktersettingen direkte,

men det betyr også at kandidaten vil ha mulighet til å tilbakevise kritiske punkter sensorene måtte ha lagt vekt på ved bedømmelsen; sensorene kan for eksempel ha misforstått noe, og i utgangspunktet vært innstilt på å trekke for dette.

For det *trede* har vi basert oss på skriftlighet i bedømmelsesprosessen; medlemmer har skrevet ned og utvekslet utførlige, detaljerte begrunnelser. Hver og en har hatt en helt annen mulighet til å tenke over innspill, vurderingsmomenter og begrunnelser enn det en vil ha i en ordinær eksamenskommissjon som diskuterer seg frem til en karakter i løpet av 15-20 minutter uten at noen av synspunktene foreligger skriftlig i forkant.

For det *fjerde* har vår gruppe bestått av langt flere personer enn en ordinær eksamenskommissjon. På masteroppgaver har alle lærestedene i dag eksamenskommissjoner med to sensorer, hvorav minst én er ekstern. Vi er altså en gruppe på i alt 7 medlemmer, hvor de fleste har argumentert utførlig (skriftlig) i tilknytning hver eneste oppgave.

For det *femte* står karakterkommissjonen i en helt annen situasjon enn ordinære sensorer i og med at vi vurderer så mange som 10 masteroppgaver på en gang, og har en unik mulighet til vurdere oppgaver i forhold til hverandre. På sett og vis er ordinær sensur ekstremt desentralisert, og mange av sensorene er ikke så ofte med på bedømmelser, og får følgelig heller ikke utviklet noe godt grunnlag for sammenligning av masteroppgaveprestasjoner.

For det *sjette* er det slik at de fleste sensorer ved ordinær sensur er plukket ut på grunnlag av den spesialkompetanse de innehar på feltet studenten behandler. I vårt tilfelle har kommisjonsmedlemmer lest og bedømt oppgaver som ikke ligger innenfor eget spesialfelt. Dette kan tenkes å ha hatt betydning for bedømmelsen. I utgangspunktet skulle en imidlertid tro at det vil være lettere å se svakheter og mangler ved oppgaver som ligger innenfor eget forskningsområde (og dette gjør kanskje resultatene som rapporteres nedenfor enda mer bemerkelsesverdige). På den annen side kan det også tenkes at en sensor som ikke er helt inne i det forskningsfeltet det er tale om, vil ha større vanskeligheter med å få øye på og verdsette substansielle bidrag, og i stedet legge for stor vekt på slike ting som problemstillingens klarhet, metodebruk og fremstilling av stoffet.

Bedømmelsen av masteroppgavene som er valgt ut: Finnes en felles forståelse av hva karakteren B innebærer?

Kommisjonsmedlemmenes karakterforslag er oppsummert i tabellen under. En detaljert tabell som får frem nyansene i medlemmenes vurderinger finnes til slutt i rapporten (appendiks).

I Tabell 1 fremgår det hvor oppgaven opprinnelig ble levert og sensurert, og hvor mange studiepoeng oppgaven formelt har. Plussene angir at studentene kan ha hatt ett eller to

”støtteemner” til masteroppgaven (separate emner), det vil si emner som innebærer at studentene gis anledning til å arbeide med masteroppgaveprosjektet i lengre tid enn det de formelle studiepoengene indikerer.

De oppgavene som er plukket utgjør for denne undersøkelsen utgjør 17-20 prosent av alle masteroppgaver som fikk karakteren B våren 2007. Oppgavene er plukket tilfeldig fra lister som til sammen hadde 54 oppgaver, men alle listene er ikke helt fullstendige og det synes å ha kommet med enkelte oppgaver som i realiteten ble levert høsten 2006. Tross denne lille usikkerheten om det samlede antall oppgaver i undersøkelsesperioden, er det ingen tvil om at de 10 oppgavene som er vurdert utgjør en ikke ubetydelig andel av B-oppgavene. Samtidig er selvsagt 10 oppgaver et lite antall, og det er ikke forsvarlig å trekke bastante konklusjoner – særlig ikke om tilstanden ved de enkelte lærestedene (og spesielt ikke ved de instituttene som har kun én oppgave blant de 10 som vurderes).

Tabell 1: Karakterkommisjonens vurderinger (antall medlemmer som har valgt det aktuelle karaktertrinnet). Se appendiks (Tabell A2) for detaljer.

Oppgavekode m. studiepoeng	A	B	C	D
1. UiT 60	-	-	4	1
2. Sampol 60+	-	1 ²⁾	5	-
3. Offadm 60++	-	-	6	1
4. Offadm 60++ ¹⁾	-	1 ²⁾	4	-
5. NTNU 45	1	5	-	-
6. NTNU 45	-	4	2	-
8. UiO 30++	-	-	6	-
9. UiO 30++	1	6	-	-
10. UiA 40	-	-	3	2
11. Offadm 60++	-	-	3	4

¹⁾ Det viste det seg etter at 5 uttalelser var avgitt at et medlem av karakterkommisjonen hadde vært sensor på oppgaven. Dette kommisjonsmedlemmet har ikke vurdert oppgaven på nytt.

²⁾ Vurdert som meget svak B av kommisjonsmedlemmet det gjaldt.

På grunnlag av materialet som rapporteres i Tabell 1 (og Tabell A2) vil kommisjonen bemerke følgende:

1. Det er *markante kvalitetsforskjeller* på masteroppgavene som er vurdert, selv om alle oppgavene har oppnådd B ved sensuren som er gjennomført.
2. Det er 3 av de 10 oppgavene som de aller fleste kommisjonsmedlemmene har vurdert som klare B-er. To av oppgavene, én fra NTNU og én fra UiO, vurderes som en meget solid B av alle medlemmer (i begge tilfeller er det foreslått svak A som karakter også). Den tredje B-oppgaven (også fra NTNU) har et par av

kommisjonsmedlemmene betegnet som meget sterk C, mens de øvrige 5 mener at det er tale om en klar og solid B.

3. Flertallet av masteroppgavene som er vurdert holder etter kommisjonsmedlemmenes mening *ikke* en standard som tilsier bruk av karakteren B, gitt ordlyden i den fagspesifikke beskrivelsen av dette karaktertrinnet. Opp mot halvparten av oppgavene (i alle fall 4 av de 10) kan heller ikke sies å være i nærheten av å oppfylle de krav som bør stilles til en B-oppgave.
4. Karakteren D er trukket frem av enkelte kommisjonsmedlemmer (ikke alltid de samme) i forbindelse med 4 av de 10 oppgavene (jf. Tabell A2). Én av oppgavene som er til vurdering blir plassert på D av et flertall av kommisjonsmedlemmene; dette flertallet mener at svakhetene ved oppgaven er så klare og vesentlige at bruk av C-en ikke er rimelig (dette er utdypet i underlagsnotatet med begrunnelser).
5. De fire oppgavene som har fått svakest vurderinger tilhører alle i hovedsak fagområdet Offentlig politikk og administrasjon. Kommisjonen har samlet hatt relativt god kompetanse for å vurdere oppgaver som faller inn under området. Dette kan være en tilfældighet. Men det er også mulig at det indikerer en mange på felles forståelse av karakterene ulike fagområder eller subdisipliner i mellom. Som kjent er det ikke så vanlig å sette sammen eksamenskommisjoner som går på tvers av de tradisjonelle områdene i faget, og dette kan i teorien begrense mulighetene for å utvikle en felles forståelse av karakterskalaen.
6. Som nevnt er kommisjonens grunnlag for å si noe om de enkelte lærestedenes sensurpraksis meget svakt. Tilfældigheter har gjort at det er blitt vurdert 3 oppgaver fra Offadm (UiB), som er det instituttet som i det statistiske materialet for 2006 hadde klart størst andel B-er (dvs. 75 prosent). Ingen av disse oppgavene har kommisjonsmedlemmer flest vurdert til B, og én av oppgavene mener flertallet at ligger på D. Ingen av de 4 oppgavene fra UiB som er med blant de 10 utvalgte ligger på B-nivå (en femte oppgave fra UiB – oppgave 12 i Tabell A2 – er vurdert av to kommisjonsmedlemmer til å ligge på B). NTNU er det eneste lærestedet der begge oppgavene som kommisjonen har vurdert havner på B. Ingen av oppgavene ved de minste miljøene (UiT og UiA) har havnet på B i kommisjonens vurdering; ved vurderingen av disse oppgavene hadde enkelte kommisjonsmedlemmer trukket frem D-en, men de fleste havnet på (en heller svak) C.
7. Kvalitetsforskjellene på oppgavene som er vurdert er markante selv om en ikke i det hele tatt tar hensyn til normeringen i form av antall studiepoeng. Faktisk er det

slik at de tre oppgavene som har klart høyest kvalitet i vårt materiale (og som har forsvart B-en), alle er levert på programmer der masteroppgaven har under 60 studiepoeng (og hvor den forberedende kursdelen av masterprogrammet har et noe større omfang i stedet). Det fremgår av underlagsnotatet med utførlige begrunnelser at de tre oppgavene er best som analyser eller forskningsrapporter betraktet, også om en ser helt bort fra de er skrevet og levert innenfor en strammere tidsramme enn øvrige oppgaver. (De tre oppgavene ble også levert til normert tid.)

8. Det er en viss variasjon i karaktersettingen ulike kommisjonsmedlemmer i mellom, noe en særlig får inntrykk av ved å se på bruken av den (riktignok vilkårlige) fingraderte karakterskalaen i Tabell A2. Ikke i noe tilfeller er det imidlertid tale om *mer* enn ett karaktertrinn forskjell (A og B på oppgavene 5 og 9; B og C på oppgavene 2, 4 og 6; C og D på oppgavene 1, 3, 10 og 11); oppgave 8 er vurdert til C av alle. Spennvidden i vurderinger synes ikke noe sted å være større enn at kommisjonen trolig kunne ha diskutert seg frem til én felles karakter uten særlig problemer (dersom det hadde vært ønskelig). Gitt den store variasjonen i kvaliteten på oppgavene som vurderes, som altså alle har oppnådd karakteren B, er det kanskje overraskende at det ikke har vært større, systematiske forskjeller i vurderingene blant kommisjonsmedlemmene. Riktignok er det slik at et par kommisjonsmedlemmer i snitt har gitt litt bedre karakterer enn de øvrige (særlig om avkryssninger på den fingraderte karakterskalaen i Tabell A2 legges til grunn)², og noen kommisjonsmedlemmer har vært noe mer tilbøyelig til å bruke et større spenn av karakterskalaen enn andre. Hovedinntrykket er likevel relativt høy grad av samsvar innad i kommisjonen. Det har ikke vært tid til å analysere det 50 siders underlagsnotatet, som inneholder detaljerte begrunnelser, med sikte på å avdekke om det er større variasjon i premissgrunnlag, anvendelse av vurderingskriterier og begrunnelsesmåter enn det karaktersettingen tyder på; medlemmer av kommisjonen kan jo komme frem til samme resultat når det gjelder karakter selv om de legger vekt på nokså ulike ting ved bedømmelsen. (De neste avsnittene av rapporten belyser likevel noe av denne tematikken nærmere.)

² De to har satt en god C eller bedre på alle oppgavene de har bedømt bortsett fra én (dvs. den ene av de to har vurdert oppgave 11 til D, mens den andre er den i kommisjonen som har satt best karakter på den samme oppgaven). Begge disse medlemmene har tilknytning til samme universitet.

Karakterbeskrivelse og sjekkliste: Er kriteriegrunnet for vurderingen tilfredsstillende?

Karakterkommisjonens vurderinger av de 10 (eller 12) masteroppgavene som er med i utvalget har vært basert på (1) den gjeldende karakterbeskrivelsen, samt (2) sjekkliste med vurderingskriterier som finnes i "Nasjoale retningslinjer for bedømmelse av masteroppgaver i statsvitenskap", behandlet i Nasjonalt fagråd 25.09.2007. Begge disse kriteriesettene er gjengitt nedenfor.

Boks 1: Den gjeldende fagspesifikke beskrivelsen av karaktertrinnene, vedtatt av Nasjonalt fagråd for statsvitenskap 27.10.2005.

Karakter	Masteroppgaven
A Fremragende	Fremragende oppgave med en original analyse, der problemstillingen er svært godt forankret i litteraturen på fagfeltet. Oppgaven utmerker seg også klart når det gjelder metodebruk og presentasjon av stoffet.
B Meget god	Meget god analyse med en klar problemstilling som er meget godt forankret i litteraturen på fagfeltet. Oppgaven er også meget god når det gjelder metodebruk og presentasjon av stoffet.
C God	En god oppgave på alle viktige punkter. Klar problemstilling og en godt gjennomført analyse som er godt forankret i litteraturen på feltet. Oppgaven er også god når det gjelder metodebruk og presentasjon av stoffet.
D Brukbar	Stort sett god oppgave med en brukbar analyse, men som også har noen vesentlige svakheter når det gjelder problemformulering, forskningsdesign, gjennomføring av analysen, metodiske ferdigheter eller presentasjon.
E Tilstrekkelig	En oppgave der analysen innfrir minimumskravene for vitenskapelig fremstilling og drøfting, men som har vesentlige mangler når det gjelder problemformulering, forskningsdesign, gjennomføring av analysen, metodiske ferdigheter eller presentasjon.
F Ikke bestått	En oppgave der analysen ikke innfrir minimumskravene for vitenskapelig fremstilling og drøfting, og som har meget vesentlige mangler når det gjelder problemformulering, forskningsdesign, gjennomføring av analysen, metodiske ferdigheter, sitatbruk eller presentasjon, eller som ikke tilfredsstillen en forskningsetisk minimumsstandard.

Boks 2: Sjekkliste med vurderingsmomenter som er antatt å kunne være til hjelp for eksamenskommisjoner i arbeidet med å sette karakter.

1. Er det en faglig sett vanskelig eller lett oppgave studenten har satt seg fore å behandle? (Spørsmålet må vurderes i forhold til antall studiepoeng på masteroppgavearbeidet.)
2. Har studenten vist evne til å avgrense temaet, slik at det blir håndterbart innenfor de rammene som gjelder for den aktuelle typen oppgave, og er det formulert en klar og presis problemstilling?
3. Har studenten oversikt over tidligere forskning på området, og vist evne til å forankre problemstillingen i denne? (Her høynes kravene med økt studiepoengstildeling.)
4. Har studenten vist evne til selvstendighet (og originalitet) ved valg av problemstilling og/eller gjennomføring av analysen? (Det forventes i utgangspunktet høyere grad av selvstendighet når studiepoengtildelingen øker.)
5. I hvilken grad er det valgt et egnet forskningsdesign — og er det begrunnet på en overbevisende måte innenfor den ramme som oppgaven setter?
6. I hvilken grad er metode(n)e som er valgt velegnede — og er de metodiske valgene begrunnet på en overbevisende måte innenfor den rammen som oppgaven setter?
7. Hvilken innsats har studenten lagt ned i innsamling og/eller tilrettelegging av materialet som benyttes i analysen? (Kravene øker med økt uttelling i studiepoeng.)
8. Har studenten vist evne til å bygge opp og gjennomføre analysen på en klar, systematisk og effektiv måte, med relevante og stringente resonnementer? I hvilken grad har studenten sørget for å underbygge eller belegge påstander som fremsettes?
9. I hvilken grad er resultatene av analysen holdbare? Er det etablert funn eller konklusjoner som forskersamfunnet kan ha tillit til? (Spørsmålet om holdbarhet gjelder like mye uansett studiepoenguttelling, men normalt vil en likevel vente seg mer hva resultater eller funn angår i oppgaver med høyt antall studiepoeng.)
10. Har studenten vist evne til å skrive klart, svare på problemstillingen som er satt opp og argumentere saklig (benytte en akademisk form)? (Kravene til språklig klarhet, og fremstillingen av stoffet mer generelt, gjelder alle typer masteroppgaver uansett studiepoengstildeling.)
11. I hvilken grad er bruken av referanser i forbindelse med sitater og kilder korrekt, herunder en fullstendig og riktig ført litteraturliste? (Dette gjelder alle typer masteroppgaver. Toleransen for slurv eller småfeil er særlig liten når det gjelder oppgaver som er levert på overtid.)

I etterkant av selve vurderingene av masteroppgavene har kommisjonens medlemmer også blitt bedt om å foreta en analyse av resultatene. I dette avsnittet følger vi opp dette ved å diskutere i hvilken grad kommisjonens medlemmer har opplevd dagens vurderingsgrunnlag (karakterbeskrivelsen og den ferske sjekklisten) som nyttige, dekkende og formålstjenlige redskaper. Kommisjonsmedlemmene ble derfor (enkeltvis) bedt om å ta stilling til hvorvidt den gjeldende karakterbeskrivelsen syntes å være tilfredsstillende, og hvis den ikke ble ansett å være det, hva medlemmene eventuelt ønsket å endre på eller sette i stedet. Særlig gjaldt dette beskrivelsen av karakteren B versus karakteren C, sett i lys av at denne komiteens hovedbudskap synes å være at for mange oppgaver har fått karakteren B og at flere av disse heller burde hatt karakteren C (og minst én oppgave til og med D).

Resultatet av denne diskusjonen synes generelt å være at kommisjonens medlemmer stort sett er stort godt fornøyd både med karakterbeskrivelsen og sjekklisten. Når det gjelder den fagspesifikke karakterbeskrivelsen ser kommisjonen et behov for noen små justeringer i

beskrivelsen av karakteren A og karakteren B for å gjøre den mer konsistent og anvendelig.³ Beskrivelsen av de øvrige karaktertrinnene ønskes imidlertid opprettholdt, selv om det kom fram av diskusjonen internt i gruppen at beskrivelsen av karakteren C i praksis kanskje ikke er helt dekkende. Det vil si: Masteroppgaver som får karakteren C – til tross for at de er jevnt gode prestasjoner – har også noen mangler og svakheter ved seg. Hvordan en eventuelt skulle åpne opp for dette i beskrivelsen av C-en, og samtidig holde fast ved at karakteren C skal anses som en god karakter, har imidlertid ikke kommisjonen greid å løse. Vi oppfordrer imidlertid Nasjonalt fagråd om å bringe disse momentene med seg videre.

Når det gjelder sjekklisten vurderte komiteens medlemmer også denne til å være ganske nyttig. Særlig ble det fremhevet av flere kommisjonsmedlemmer at denne listen av vurderingsmomenter burde vedlegges alle oppgaver ved utsendelse til bedømmelseskommisjoner, og at dette kunne være av spesiell nytte for ferske sensorer. I tillegg vil en slik sjekkliste være informativ for studentene selv, som ved at denne sjekklisten offentliggjøres i større grad vil få en reell forståelse av hva sensorer ser etter når de setter karakter.

Til tross for at sjekklisten opplevdes som nyttig, ble det imidlertid tidlig klart at kommisjonens medlemmer i sitt bedømmelsesarbeid fortrinnsvis brukte en kortere liste av "sjekkpunkter", nemlig de som kan trekkes direkte ut av beskrivelsen av karaktertrinnene. Det gjelder følgende punkter (hentet ut av beskrivelsen for karakteren B):

1. Problemstilling (klar)
2. Litteraturforankring (meget god)
3. Metodebruk (meget god)
4. Analyse/analytiske ferdigheter (meget god)
5. Presentasjon av stoffet (meget god)

Hvorvidt alle fem dimensjonene må kunne beskrives som "meget gode" for å få karakteren B er imidlertid litt åpent, men ordene er brukt eksplisitt når det gjelder fire av de fem dimensjonene. Etter kommisjonens mening bør følgelig også minst fire av de fem dimensjonene være på plass som "meget gode" for å forsvare karakteren B. På dette

³ Vi har fremmet et separat forslag med justerte formuleringer direkte til Nasjonalt fagråd. Vi gjengir formuleringene her. Karakteren A: "En fremragende oppgave med en original analyse, der problemstillingen og analysen er svært godt forankret i litteraturen på fagfeltet. Oppgaven utmerker seg også med svært god metodebruk, og presentasjon av stoffet." Karakteren B: "En meget god oppgave med en selvstendig analyse, der problemstillingen og analysen er meget godt forankret i litteraturen på fagfeltet. Oppgaven er også meget god når det gjelder metodebruk og presentasjon av stoffet."

punktet er det syndet når det gjelder flere av oppgavene vi har vurdert, i den forstand at det er gitt B på oppgaver som er "meget gode" på få eller ingen av de dimensjonene som er sentrale i beskrivelsen av karaktertrinnet.

Selv om kommisjonens medlemmer generelt syntes at den lange sjekklisten av punkter fra Nasjonalt fagråd er ganske nyttig, ble det imidlertid poengtert at en slik kortere liste som den ovenfor mer anvendelig – eller lettere å anvende i praksis. Vi tror det bør overveies å systematisere karakterbeskrivelsen i form av et (avkrysnings)skjema; selv om karakterfastsettelsen skal og bør hvile på et selvstendig, faglig skjønn og en helhetsvurdering, kan skjemaer av denne typen være en støtte som bidrar til at sensorenes skjønnsutøvelse blir mest mulig systematisk og likeartet (eller likebehandlende). Et eksempel på dette er vist i Tabell A3 i appendikset, og den bygger utelukkende på gjeldende beskrivelse av karaktertrinnene (som alle sensorer uansett er bundet av).

Studiepoeng (dvs. forventet tidsbruk), undersøkelsens omfang og karakterfastsetting: Er det mulig å holde ulike typer masteroppgaver fra hverandre ved bedømmelsen?

Vi vil innledningsvis understreke at selv om at kvalitetsforskjellene på oppgavene kommisjonen har vurdert er markante helt uavhengig av normeringen i form av studiepoeng, mener vi at det prinsipielt må tas hensyn til forventet tidsbruk ved sensur.

Vår målsetting med dette punktet har dermed vært å drøfte om det er grunnlag for å etablere en kognitiv ramme for sensur som i større grad enn i dag sikrer likebehandling med hensyn til karakterfastsetting på tvers av de ulike masterprogrammene. Dette problemet har sitt utgangspunkt i at de ulike masterprogrammene har ulik uttelling på masteoppgavene målt i studiepoeng. Dette innebærer at kandidatene ved normal progresjon har ulik tid til disposisjon for å ferdigstille sine oppgaver.

Det synes derfor rimelig at det stilles noe lavere krav til for eksempel en oppgave på 30 studiepoeng enn til en oppgave på 60 studiepoeng med hensyn til omfang. Vi vil her understreke at det *ikke* dreier seg om å stille lavere *kvalitetskrav*. Det må være de samme kravene til klarhet, stringens, metodebruk, presentasjon av stoffet, og lignende. Dette er momenter som er nevnt i den kvalitative beskrivelsen av karaktertrinnene, og i forhold til disse ser vi ingen grunn til å skille mellom oppgaver med ulik studiepoengstildeling.

Vi vil imidlertid vente oss mer av en masteroppgave med flere studiepoeng når det gjelder dimensjonering av undersøkelsen, hvor omfattende materialet som analyseres er og hvor grundig analysen er, og – muligens – graden av selvstendighet ved problemvalg, analyse og fremstilling. Dette er momenter som ikke direkte nevnes i beskrivelsen av karaktertrinnene. Indirekte er det imidlertid en sammenheng mellom omfang og kvalitet fordi en mer

omfattende undersøkelse/analyse av et fenomen under ellers like vilkår gjerne fremstår som "bedre" enn en mindre omfattende undersøkelse. Det er derfor meget viktig at alle eksamenskommisjoner er kjent med at vilkårene er *ulike* på de eksisterende masterprogrammene.

Det er imidlertid ikke lett å avgjøre hvordan disse ulikhetene bør gi seg utslag i praksis ved sensurering. For det første, for at det skal være mulig å sammenlikne prestasjonene, må den *faktiske* tiden kandidatene har anvendt tas i betraktning. Ellers vil det slå ufordelaktig ut for den mest omfattende oppgavetypen: Hvis det er noe lettere å få en god karakter på en oppgave med få studiepoeng ("liten" oppgave), mens studentene faktisk har brukt like lang tid på gjennomføringen, får den minste oppgaven en bedre karaktermessig uttelling. Det vil ikke innebære likebehandling. Derfor må det foretas en vurdering av forholdet mellom normert tid og faktisk anvendt tid. Den enkleste måten å måle dette på er selvfølgelig å ta utgangspunkt i hvorvidt kandidaten har levert på normert tid. Det kan dog reises vektige motforestillinger mot utlukkende å basere seg på dette kriteriet. For eksempel, hva med studenter som opplever at uforutsette forhold forsinker fremdriften? Det kan være forhold knyttet til feltarbeid, datasett som blir forsinket i forhold til forventet levering (f.eks. fra NSD) eller andre forhold studenten egentlig ikke direkte kan lastes for? Det synes åpenbart at slike momenter må tas i betraktning, både av rene rettferdighetshensyn og av hensyn til hele masterprogrammet. Det er kommisjonens klare oppfatning at vi må unngå å etablere en ramme som er så rigid at studentene kan tenkes å vegre seg mot å gi seg i kast med prosjekter som innebærer feltarbeid, tilrettelegging av større datasett eller innsamling av eget, originalt materiale.⁴

For det andre finner kommisjonen at det i praksis er vanskelig å anvende dette med ulikheter i studiepoeng på en systematisk måte. Det kan argumenteres for at programmene i ulik grad har poenggivende emner som kan regnes som "støtteemner" til oppgaven (prosjektseminar, oppgaveseminar eller lignende) slik at studentene i praksis har lenger tid til rådighet enn antall studiepoeng tilsier. Dette kan vanskeliggjøre en direkte sammenlikning også mellom oppgaver som har formelt likt antall studiepoeng. I tillegg viser en kort egenvurdering fra panelets medlemmer om hvorvidt antall studiepoeng *faktisk* ble tatt i betraktning under arbeidet, og hvorvidt de *bør* tas i betraktning, til dels klare avvik mellom teori og praksis. Mens studiepoeng i den faktiske bedømmingen bare i liten eller noen grad ble rapportert vektlagt, mener flere at dette burde vektlegges i større grad – spesielt med hensyn til:

- a) hvorvidt det er en faglig sett vanskelig eller lett oppgave studenten har satt seg fore å behandle,

⁴ En evaluering av masterprogrammet ved UiO viser dog at det stort sett ikke er denne typen forhold som har ført til forsinkelser i forhold til normert tid.

- b) hvorvidt studenten har vist evne til selvstendighet (og originalitet) ved valg av problemstilling og/eller gjennomføring av analysen, og
- c) hvilken innsats studenten har lagt ned i innsamling og/eller tilrettelegging av materialet som benyttes i analysen.

Det er selvsagt mulig at den begrensede vekten kommisjonsmedlemmene tilla antall studiepoeng ved bedømmelsen av oppgavene kommisjonen har hatt til vurdering, skyldes at de "minste" oppgavene var de klart beste. På bakgrunn av denne gjennomgangen tror vi at det uansett både vil være *svært vanskelig og lite formålstjenlig å etablere et sett detaljerte retningslinjer på dette området*. Derimot, for i noen grad å ivareta prinsippet om likebehandling med hensyn til tidsbruk, kan det være fornuftig å ta utgangspunkt i følgende enkle "regel":

Hvis en masteroppgave på mindre enn 60 studiepoeng er levert til normert tid, bør det få betydning for vurderingen på enkelte punkter (vi må forvente en litt mindre dimensjonert undersøkelse og analyse på grunn av en knappere tidsramme). Men dersom en slik masteroppgave er levert på overtid, bør samme kriterier som for en oppgave på 60 studiepoeng legges til grunn.

For at en slik regel skal kunne anvendes i praksis må imidlertid alle mastergradskommisjoner få informasjon om hvor lang tid kandidaten faktisk har brukt i forhold til normert tid (dvs. antall semester registrert som fulltidsstudent).⁵

Konklusjon

Kommisjonens vurderinger av de 10 oppgavene som er valgt ut må sies å være omfattende, detaljert og grundig, noe som blant annet dokumenteres i det fortrolige underlagsnotatet som er laget. Det er utvilsomt en styrke i forhold til konklusjonene som trekkes om kvaliteten på hver enkelt oppgave. Samtidig er det klart at både det faktum at kommisjonen har vært mye større enn en ordinær eksamenskommisjon, og det at den har jobbet på en annen måte enn eksamenskommisjoner gjør, kan ha påvirket resultatene vi er kommet frem til; grundigere vurderinger kan lett bli til strengere vurderinger, selv om kommisjonens medlemmer har bestrebet seg på å vurdere oppgavene på nøyaktig samme måte som de ville gjort ved en ordinær bedømmelse.

Det har ikke vært mulig å gjennomføre en inngående vurdering av mer enn 10 oppgaver. Antallet oppgaver er lavt, og utgjør i underkant av 20 prosent av de oppgavene som ble

⁵ Hvilket betyr at permisjoner, sykemeldinger osv. vil kunne komme til fratrekk i tiden.

bedømt til karakteren B våren 2007. Kommisjonens hovedfunn, nemlig at kun 3 av de 10 oppgavene holder den standard en bør forvente av en B-oppgave, er imidlertid så klart at det er grunnlag for å si at det i dag ikke synes å være en felles forståelse til grunn for karaktersettingen innenfor faget. Det betyr at karakterene som gis heller ikke er direkte sammenlignbare; undersøkelsen viser i alle fall betydelig variasjon i kvaliteten på masteroppgaver som alle oppnådde B ved sensuren for et halv års tid siden.

Denne kommisjonens oppgave har vært å gjennomføre en kvalitativ undersøkelse – en stikkprøve – av karaktersettingen på masteroppgaver. Vår oppgave er ikke primært å komme med anbefalinger.

Kommisjonen vil bemerke at det i og for seg er enkelt å beskrive forskjeller i karaktersetting slik det er gjort i denne rapporten, men verre å gjøre noe med dem. Det skyldes blant annet hensynet til likebehandlig eller rettferdighet både i forhold til tid (tidligere, nåværende og fremtidige studenter ved samme utdanningsprogram) og rom (studenter ved andre institusjoner eller på andre programmer). Derfor er det ikke gjort i en håndvending å endre praksis og "flytte" gjennomsnittskarakteren nedover på skalaen. I denne sammenheng vil karakterkommisjonen likevel foreslå at alle de statsvitenskaplige masterutdanningene forpliktet seg til å sende ut nødvendig informasjon til sensorene som benyttes. Det gjelder særlig karakterbeskrivelsene (gjerne utformet som et skjema som kan gi bedre støtte ved helhetsvurderingen enn det beskrivelsen av karakterene alene kanskje kan), og det gjelder nasjonale retningslinjer (med sjekkliste) for bedømmelse. Dette kan forhåpentligvis være en effektiv start på en utjamningsprosess som sikkert kommer til å ta noen år, forutsatt – selvfølgelig – at involverte parter ser på det å oppnå mer enhetlig karaktersetting som et mål det er verdt å forfølge; det er vanskelig å tenke seg at en kan oppnå sammenlignbare masteroppgavekarakterer innenfor faget uten aktiv, målrettet innsats fra de enkelte lærestedenes side.

APPENDIKS

Tabell A1. Karakterfordelingen for masteroppgaver i statsvitenskap (eller underdisipliner) 2006. Tabellen utgjør tallgrunnlaget for Figur 1. Kilde: DBH og Nasjonalt fagråd i statsvitenskap (sak 13/2007, behandlet i fagrådsmøte 25.09.07).

Institusjon	A	%	B	%	C	%	D	%	E	%	Antall bestått
UiB-AdmOrg	3	15,0	15	75,0	2	10,0	0	0	0	0	20
NTNU	7	31,8	11	50,0	3	13,6	1	4,5	0	0	22
UiT	0	0,0	7	46,7	8	53,3	0	0,0	0	0	15
UiO	10	11,0	43	47,3	31	34,1	7	7,7	0	0	91
UiB-SamPol	8	38,1	8	38,1	4	19,0	1	4,8	0	0	21
HiA-OffAdm	1	4,2	12	50,0	9	37,4	1	4,2	1	4,2	24
SUM	29	15,0	96	49,8	57	29,5	10	5,2	1	0,5	193

Tabell A2: Kommisjonsmedlemmenes vurdering av de utplukkede masteroppgavene på en svært fingradert karakterskala. Tallene uttrykker hvor mange kommisjonsmedlemmer som har plassert en bestemt masteroppgave på vedkommende karaktertrinn. Når det gjelder kolonnen for karakterer svakere enn C5, er karaktertrinnet som det enkelte medlem har foreslått ført opp.

Oppgaver Studiepoeng	Bedre enn A2	A2	A1	B10	B9	B8	B7	B6	B5	B4	B3	B2	B1	C10	C9	C8	C7	C6	C5	Svakere enn C5
1. UiT 60															1				1	C2; C2; D9
2. Sam 60+													1	1			2	1		C2
3. Off 60++															2	2				C4; C3; D7
4. Off 60++¹⁾													1			1			1	C4; C4
5. NTNU 45			1		1	2			1	1										
6. NTNU 45					1		2		1					1	1					
7. Sam 60+																				
8. UiO 30++															1	2		1	1	C2
9. UiO 30++			1	2	1	1	2													
10. UiA 40																	1			C2; C1; D8; D5
11. Off 60++																1	1			C3; D7; D7; D5; D5
12.Sam 60+¹⁾									1		1									

¹⁾ Når det gjelder oppgave 4, viste det seg etter at 5 uttalelser var avgitt at et medlem av kommisjonen hadde vært sensor på oppgaven. Når det gjelder oppgave 12, var det klart for de to som har avgitt uttalelse at ett medlem av kommisjonen hadde vært veileder. Vurderingene i disse to tilfellene skal derfor ikke vektlegges i samme grad som øvrige oppgaver.

Tabell A3: Eksempel på tabell som kan bidra til mer systematisk skjønnsutøvelse ved bedømmelse av masteroppgaver i statsvitenskap.

Vurderings- dimensjoner fra beskrivelsen av karaktertrinnene	Svært godt (fremragende)	Meget godt	Godt	Brukbart	Tilstrekkelig
Problemstilling					
Analyse					
Litteraturforankring					
Metodebruk					
Presentasjon av stoffet					
SAMLET:					
HELHETSVURDERING MED VEKTLEGGING AV DE ULIKE DIMENSJONENES BETYDNING:					