

## Kvalitetsreformen ved MN-fakultetet: undervisningsformer, vurderingsformer og karakterbruk.

Mye av arbeidet med kvalitetsreformen har hittil handlet om struktur og styringsformer — hvilke bachelorprogram skal vi ha, hvordan skal de organiseres, hvilke fullmakter skal et programstyre ha osv. I dette arbeidet har det egentlige innholdet i reformen — nemlig studiekvaliteten — kommet i bakgrunnen. Forhåpentligvis har *Inspirasjonsuken for bedre studiekvalitet* 18.-25. oktober brakt den frem igjen i manges bevissthet, for nå er tiden inne til å fylle våre flotte diagrammer med innhold!

I dette notatet vil jeg drøfte tre viktige spørsmål: hvordan vi skal undervise, hvordan vi skal vurdere, og (litt mer spesielt og teknisk) hvordan vi skal bruke det nye karaktersystemet. Jeg har ingen endelige svar, men jeg har noen utdypende spørsmål og noen forslag. Hensikten med notatet er primært å stimulere til diskusjonen ved å sette andres forslag og forsøk inn i en felles ramme.

### Undervisningsformer

Karikaturen på universitetsundervisning: En foreleser holder enetale for studentene to ganger i uken fra september til november, deretter lukkes alle studentene inne i et eksamenslokale et passende antall timer, og til slutt tildeler foreleser og sensor hver student et desimaltall på grunnlag av eksamensbesvarelsene.

Det hender nok at vi selv er med å bringe denne forestillingen videre, men egentlig vet vi godt at virkeligheten ser annerledes ut. Vi har ikke bare forelesninger, men regneøvelser, kollokvier, lab-kurs, feltkurs, obligatoriske oppgaver, datalab'er, prøveeksamener, helgeregninger, orakeltjeneste og mye annet. I de fleste kurs er likevel forelesningene den sentrale undervisningskomponenten; det er de som bestemmer rytmen i kurset, definerer hva som er viktig og hva som er mindre viktig, fungerer som informasjonssentral, og binder de andre undervisningskomponentene sammen.

Forelesningen som undervisningsform er under sterkt press. Mange mener at den er ineffektiv, fremmedgjørende og foreldet, og det kommer av og til oppslag om forskning som

tyder på at forelesninger overhode ikke har noen effekt på studentenes læring. Man skal selvfølgelig være forsiktig med å opponere mot forskningsresultater med ren synsing, men jeg må innrømme at jeg har vanskelig å feste lit til de mest negative resultatene – til det synes jeg at jeg har lært alt for mye av forelesninger selv! Etter min erfaring er forelesninger uovertrufne til å vise hvordan fagfolk faktisk tenker, hva som er deres intuisjon, hva de oppfatter som hovedtemaer og hva som bare er sidetemaer – kort sagt: de viser hvordan vi bør strukturere lærestoffet inne i vårt eget hode. Forelesningene har to fordeler fremfor læreboken – de er dynamiske og øyeblikksrettede. Med ”dynamisk” mener jeg at forelesningsformen tillater oss å presentere argumentene slik de vokser frem – vi får gitt studentene et gløtt inn i vårt eget tankeverksted der de får se hvordan ideer griper inn i hverandre, og får et innblikk i hvilke assosiasjoner som leder oss videre. Med ”øyeblikksrettet” mener jeg at en forelesning er rettet mot studentene på det nivået de befinner seg akkurat nå, og at den derfor kan bygges opp helt ulikt ettersom den er en første introduksjon eller en oversikt over allerede velkjent stoff. Til sammenligning er en lærebok rettet mot gjenbruk – den skal fungere både ved første gangs gjennomlesning og ved senere oppslag og repetisjon, og blir derfor tyngre og mer omstendelig i formen.

Forelesninger har også sine åpenbare ulemper, men uansett hva vi måtte mene om dem som undervisningsform, tror jeg vi må regne med å leve med dem en stund til. Skal universitetet makte å opprettholde et bredt faglig tilbud fra første semester til doktorgrad, er vi avhengig av å utnytte den stordriftsfordelen som forelesningene gir – vi har rett og slett ikke kapasitet til å drive all vår undervisning i mer intensive former. Spørsmålet blir derfor ikke om vi skal ha forelesninger, men om hvordan de bør være og hvordan de skal bindes sammen med resten av undervisningen.

Ved Kjemisk institutt har man forsøkt å organisere forelesningene på en ny måte ved å legge den ene helt i begynnelsen av uken og den andre helt i slutten. Den første forelesningen er en ”forlesning” der ukens tema presenteres i brede trekk, mens den andre er en ”etterlesning” der man oppsummerer og systematiserer lærestoffet. Mellom de to forelesningene ligger de mer studentaktive undervisningselementene: gruppeundervisning, kollokvier, laboratoriarbeid. Forelesningsvolumet er kuttet ned slik at studentene ikke kan regne med å få gjennomgått alt lærestoffet i detalj. Det er ikke sikkert at dette opplegget passer i alle kurs, men det gir et eksempel på hvordan man kan fornye forelesningsformen, og spesielt hvordan man kan knytte forelesninger tettere til de andre undervisningsaktivitetene.

Skal et slikt opplegg fungere, er man også avhengig av å ha struktur på de andre undervisningskomponentene. Etter min erfaring er det et vel så stort behov for å revitalisere gruppeundervisningen som det er for å omorganisere forelesningene. Gruppeundervisningen

skal hjelpe studentene i deres aktive arbeid med lærestoffet, men degenererer ofte til oppgavegjennomgåelse med studentene som passive tilskuere. Ofte skyldes dette at studentene ikke er innstilt på å være aktive – enten fordi de ikke har arbeidet med stoffet, eller fordi de ikke vil eksponere seg for sine medstudenter. En av største utfordringer er å lage rammer for gruppeundervisningen som leder studentene inn i en mer aktiv rolle. Igjen kan det være instruktivt å se på det nye opplegget ved Kjemisk institutt – her deles studentene inn i klasser som igjen deles inn i kollokviegrupper. Kollokviegruppene får ansvar for å presentere oppgaveløsninger i klassen. I prinsippet er det ikke noe nytt i dette opplegget – jeg møtte selv et lignende opplegg med ”minigrupper” som matematikkstudent for nesten 30 år siden, og har selv benyttet det i min egen gruppeundervisning. Matematikkopplegget gikk imidlertid i oppløsning – etter hvert som studentene fikk mer å gjøre utenom studiene, ble de mindre interessert i å være med i forpliktende samarbeid, og det ble stadig vanskeligere å finne tidspunkt som passet for alle. Det er imidlertid en viktig forskjell på datidens matematikkopplegg og dagens kjemiopplegg – matematikkens ”minigrupper” hadde aldri noen offisiell status, det var opp til gruppelæreren å organisere dem og følge dem opp, og de var aldri timeplanfestet. Dette gjorde dem uforpliktende og sårbare – og et lett offer for studentenes manglende interesse. I dagens kjemiopplegg er kollokviegruppene en del av den organiserte undervisningen, og studentene som vil være med, må forplikte seg til aktiv deltagelse. Det er mulig å stå utenfor opplegget, men da må man være innstilt på å arbeide med stoffet på egen hånd. Også ved Fysisk institutt har man prøvd et lignende opplegg med ”aktive” og ”passive” studenter.

Personlig tror jeg slike opplegg har mye for seg. Et alternativ kunne selvfølgelig være å gjøre den ”aktive” modellen obligatorisk for alle, men det er så mange studenter som har gode grunner til å ønske en mer individuell arbeidsform, at dette neppe er ønskelig. En av erfaringene fra matematikkundervisningen var at det å ”presse” en motvillig person med i en minigruppe, lett førte til at hele gruppen gikk i oppløsning. Jeg ser én stor ulempe med å gjøre det ”aktive” opplegget frivillig, og det er at de som faglig eller sosialt har aller størst behov for å være med, ofte velger å stå utenfor. Vi må derfor gjøre det ”aktive” tilbudet mest mulig inkluderende og la det fremstå som standardopplegget for heltidsstudenter. Jeg tror også at det er viktig at et slikt opplegg i størst mulig grad timeplanfestes og inkluderes i det generelle studieopplegget – for eksempel ved at de ukentlige oppgavesettene spesifiserer hva som skal gjøres på kollokviene, eller at obligatoriske oppgaver og underveiseksamener får en tilknytning til kollokviegruppene.

En undervisningsmetode som er i vinden for tiden er *problembasert læring* (PBL). Den finnes i forskjellige avskygninger, men hovedfilosofien er at studentene skal ta

utgangspunkt i konkrete faglige problemer og lære faget gjennom arbeidet med disse problemene. Selv om undervisningen gjerne begynner med en introduksjon til fagområdet, er det i hovedsak studentenes ansvar selv finne frem til den teorien som er nødvendig for å løse problemene. Ved Universitetet i Oslo er det i hovedsak de medisinske miljøene som har tatt i bruk PBL, og ved vårt fakultet er Farmasøytisk institutt i ferd med å legge om sin undervisning i denne retningen.

Selv innbiller jeg meg at PBL egner seg bedre for et profesjonsstudium med konkrete, praktiske problemstillinger og med et behagelig forhåndstall mellom studenter og lærere, enn for et teoretisk rettet allmennstudium med mange studenter og få lærere. Vår primære oppgave er å formidle de sentrale naturvitenskapelige teoriene, og jeg har ærlig talt problemer med å forestille meg at Ronny og Rita på fire måneder skal nå frem til Maxwells ligninger gjennom en sekvens av problemer. På den annen side kan vi godt tenke oss mer anvendte enkeltkurs der studentene primært skal lære å bruke teorien i praksis, og der PBL-metoder kan være med å skape den realismen vi ønsker oss. Jeg vil også peke på at informasjonsteknologien har gitt oss muligheten til å arbeide med mer realistiske og spennende problemstillinger også innenfor vanlige kurs.

Med kvalitetsreformen følger et krav om at vi skal vurdere studentene hyppigere og på mer variert måte. Det betyr at en større del av undervisningen blir knyttet til vurdering, og at vi større grad enn før blir tvunget til å tenke på sammenhengen mellom undervisningsopplegg og vurderingsopplegg. Det er derfor naturlig å kombinere den mer detaljerte diskusjonen av arbeidsformer med en diskusjon av ulike vurderingsformer.

### **Vurderingsformer**

Pedagogene skiller mellom det de kaller *summativ* og *formativ* vurdering av en students arbeid. Vurderingen er *summativ* dersom hensikten er å gi studenten en karakter (eller lignende) som vil inngå i den endelige kursvurderingen, mens den er *formativ* dersom hensikten bare er å gi studenten veiledning for det videre arbeidet med kurset. I utgangspunktet kan distinksjonen virke akademisk, men ser vi situasjonen fra studentens posisjon, får vi øye på en dramatisk forskjell. En student som leverer inn sitt arbeid for *summativ* vurdering, bør i størst mulig grad forsøke å skjule sine svakheter og fremheve sine styrker, mens en student som innleverer arbeidet for *formativ* vurdering, bør være så ærlig som mulig og kanskje legge spesiell vekt på sine svakheter. Legg også merke til at lærerens rolle er svært ulik i de to oppleggene - ved *formativ* vurdering er han en medspiller, men ved *summativ* vurdering kan han fort oppfattes som en motspiller. Når studentene leverer inn et

arbeid for vurdering, må det alltid være klart på forhånd om vurderingen er formativ eller summativ.

I Stortingsmelding 27 omtales både formativ og summativ vurdering, men uten at det skilles klart mellom de to formene. Det understrekes sterkt at studentene skal få hyppig tilbakemelding på det de gjør, men det vektlegges også at den endelige vurderingen skal bygge på mer enn en slutteksamen. I prinsippet er det ingen ting i veien for å ha både formativ og summativ underveisvurdering i et kurs, men i praksis er det minst tre komplikasjoner. Den ene er at den summative vurderingen fort tar luven fra den formative – studentene oppfatter oppgavene med karakter som viktigere enn dem uten. Den andre komplikasjonen er rett og slett at det kan bli for mye vurdering – skal studentene ta tre kurs i parallell med både summativ og formativ vurdering, vikler vi oss fort inn i et vurderingsmareritt for både studenter og lærere. Den tredje komplikasjonen er at studentene kanskje ikke stoler på at lærerne virkelig skiller den formative vurderingen fra den summative, og at de derfor ikke tør blottstille seg i den formative delen. Mange kjenner kanskje igjen denne situasjonen fra videregående skole der læreren både skal være en faglig veileder og sette standpunkt karakter basert på innsatsen underveis.

Skillet mellom summativ og formativ vurdering er skarpt, men vi kan tenke oss ordninger som balanserer på kanten mellom de to. Et eksempel er obligatoriske oppgaver som bedømmes til bestått/ikke bestått, men der hensikten er at alle skal bestå bare de er villige til å legge inn nok arbeid. Her kan den formative vurderingen bestå i at studentene får flere muligheter til å levere inn (deler av) besvarelsen, og at de får råd om hvordan de bør rette opp eller videreføre arbeidet for å få det godkjent. En annen mulighet er å koble innholdet i den formative og summative delen slik at en god besvarelse av den formative delen gir et godt utgangspunkt for den summative. Hvis den summative vurderingen er en skoleeksamen uten hjelpemidler, kan dette rett og slett bety at noen av oppgavene fra den formative vurderingen dukker opp på nytt i den summative, men vi kan også tenke oss løsere forbindelser. Alle slike ordninger bør være avklart med studentene på forhånd siden hensikten er å få dem til å ta den formative vurderingen på alvor.

Den formative vurderingen kan spenne fra korte råd om det studentene gjør i undervisningssituasjonen til mer utfyllende kommentarer til lengre skriftlige eller muntlige fremstillinger. Skrivetrening står sentralt i kvalitetsreformen, og i våre fag bør dette i hvert fall omfatte klare og velbegrunnede oppgavebesvarelser med presis og korrekt bruk av fagterminologi.

Hvilke muligheter har vi så for den summative underveisvurderingen? Jeg skal drøfte fire alternativer: (skriftlig) skoleeksamen, muntlig eksamen, hjemmeeksamen, og mappevurderingen.

Skoleeksamen: Det er flere grunner til at vi kan ønske å ha eksamen underveis. Hvis vi bare ønsker å få flere vurderingspunkter eller å holde studentene litt i ørene tidlig i semesteret, kan en enkel skoleeksamen være det vi trenger. For å unngå organisatoriske vanskeligheter bør denne være så enkel at den kan holdes i undervisningslokalet. Matematisk institutt har eksperimentert med flervalgsoppgaver der svaralternativene permuteres for å unngå problemer med kinking. Fordelen med denne eksamensformen er at den er lite arbeidskrevende og gir studentene rask tilbakemelding. Ulempen er at den ikke gir tilbakemelding på annet enn hva som er riktig og hva som er galt. Den bør derfor ikke brukes som eneste eksamensform.

Muntlig eksamen betyr ikke nødvendigvis tradisjonell teorihøring. Et alternativ er å la studentene fremføre et faglig foredrag etterfulgt av spørsmål, eller å la dem gjennomgå en oppgave de har forberedt på forhånd. Fordelen ved denne eksamensformen er at den gir mulighet for en dialog med studentene, ulempen er at den er uhyre tidkrevende. En mulighet er å bruke muntlig eksamen til å ta stikkprøver ved andre eksamensformer der det kan være tvil om studentene har løst oppgavene selvstendig (dette har f.eks. vært praktisert i forbindelse med obligatoriske oppgaver i informatikk). Matematisk institutt har gode erfaringer med muntlige eksamener der studentene får i oppdrag å gjennomgå en oppgave fra et oppgavesett de har forberedt.

Hjemmeeksamen: Ved denne eksamensformen får studentene ofte god tid til å besvare oppgavene (f.eks. én uke), og de har fri tilgang til hjelpemidler. Fordelen med denne oppgaveformen er at man kan teste studentene på komplekse oppgaver, og samtidig stille høye krav til presentasjonen. Ulempen er at det kan være vanskelig å lage rettferdige spilleregler, og det kan også være problematisk å avgjøre i hvilken grad studentene har løst oppgavene selv. Det siste gjelder spesielt eksakte fag der en oppgave ofte har bare én eller to fornuftige løsningsmetoder, og der plagiat derfor er vanskelig å avsløre.

Et viktig spørsmål å ta stilling til ved denne eksamensformen er i hvilken grad studentene har lov til å bruke hjelpemidler, samarbeide eller få hjelp av utenforstående. Siden vi ellers i undervisningen er positive til at studentene lærer å samarbeide og bruke hjelpemidler, kan det virke absurd å forby dette på en hjemmeeksamen. På den annen side kan det skape åpenbare urettferdigheter ved at karakteren man får avhenger av hvilke bekjenskaper man har (de mest ytterliggående vil hevde at evnen til å stifte de riktige

bekjentskapene er en viktig egenskap som bør honoreres, men jeg vil innvende at det å ha en fetter som er flink i matematikk neppe kan regnes som spesielt honorabelt).

Matematisk institutt har eksperimentert med hjemmeeksamener i sine begynnerkurs. Her har spillereglene vært at studentene har lov til å samarbeide og få så mye hjelp de vil, men at den besvarelsen de til slutt leverer inn, skal være skrevet av dem selv og representere den kunnskapen de selv sitter inne med. Disse reglene har den fordel at de er rimelig kontrollerbare, for eksempel ved å bruke muntlig eksamen til å ta stikkprøver. Reglene kan også virke logiske dersom man tror at det er sluttresultatet som er viktig og ikke hvordan man har nådd det, men i matematikk er dette med respekt å melde det reneste tøv – man lærer mye mer av å løse en oppgave selv enn ved å få andre til å forklare den for seg.

Til tross for problemene med å lage en rettferdig ramme, føler jeg at hjemmeeksamener faglig og pedagogisk har så mange positive sider at vi bør fortsette å bruke dem. På grunn av rettferdighetsproblemene er det imidlertid best å bruke dem i kombinasjon med andre vurderingsformer.

Mappevurdering: Jeg vet ikke om det er riktig, men jeg har alltid innbilt meg at mappevurdering først oppstod innenfor kunstutdannelse: Studentene arbeidet med flere teknikker, og ved studiets slutt plukket de ut de arbeidene de var mest tilfreds med innenfor hver teknikk og ble bedømt på grunnlag av dem. Metoden kan høres besnærende ut, men er ikke så lett å overføre direkte til våre fag – det er liksom ikke den samme susen over en mappe full av integraler vi er fornøyde med å ha fått til, eller kjemiske analyser der feilen er på promillenivå. Én grunn er kanskje at realfagene har en så klar progresjon at et integral vi var stolte over å få til i september, er blitt en trivialitet i november. En annen grunn kan jeg kanskje illustrere ved å vri på en historie som blir fortalt om Norbert Wiener. Wiener ble en gang spurt om hvordan det kunne ha seg at han som var slik en glimrende matematiker, kunne være en så elendig sjakkspiller. Wiener svarte at det var da opplagt – i sjakk ble man målt etter sitt dårligste trekk, mens man i matematikk ble målt etter sitt beste! Til tross for Wieners autoritet, våger jeg å påstå det motsatte om forholdet mellom kunstutdannelse og realfagsutdannelse - i kunstutdannelsen kan det være rimelig å måle folk etter det beste de kan få til, i realfagsutdannelse må vi også ta hensyn til de manglene de har - det er viktig å beherske *alle* de grunnleggende teknikkene før man fortsetter til neste nivå!

Til saken – hva er egentlig mappevurdering? For den som virkelig ønsker å grave seg ned i dette spørsmålet, anbefaler jeg UNIPEDs temanummer om mappevurdering (nr. 2, 2002) som gir en grundig, informativ og fasettert innføring. Her er min kortversjon: Nøkkelordene i mappevurdering er *samling*, *utvelgelse* og *refleksjon* (det lyder utvilsomt bedre på engelsk: *collection*, *selection*, *reflection*). Først produserer studentene en samling

tekster, oppgavebesvarelser, bilder eller hva som måtte være relevant i det enkelte tilfellet. Deretter velger de ut de elementene som skal puttes i mappen for vurdering. Ofte vil det være visse krav til hvordan mappen skal settes sammen – i en kunstutdanning kan det f.eks. være krav til at mappen ikke skal fylles opp med akvareller, men at samtlige teknikker skal være rimelig representert. Til slutt lager studentene en reflekterende begrunnelse som både forklarer valget av mappeelementer og gir en kort vurdering av hvert enkelt mappeelement. Når sensorene vurderer mappen, tar de både hensyn til mappeelementene og til begrunnelsen. Ifølge denne definisjonen er altså selvvurdering et sentralt element i en ekte mappevurdering.

Når vi hittil har snakket om mappevurdering ved vårt fakultet, har vi nok vært langt mindre ambisiøse enn denne definisjonen tilsier. Ofte har det bare vært snakk om å samle sammen enkeltstående eksamener, muligens med rett til å fjerne den svakeste. Siden eksamenene allerede har vært karaktersatt, har utvelgelsesprosessen neppe satt studentenes vurderingsevne på de største prøver! Vi har sannsynligvis kurs der en fullverdig mappevurdering kan ha en naturlig plass, men jeg vil advare mot å gjøre en så omfattende vurderingsform til standard. Jeg tror vi er bedre tjent med å senke ambisjonsnivået noe, og heller finne andre måter til å oppøve studentenes vurderingsevne.

Uansett hvilke eksamensformer vi velger, er det viktig at vi tar hensyn til studentenes totale belastning. Vi må huske på at studentene tar to kurs i tillegg til vårt eget. Hvis vi ønsker fire underveisevurderinger i vårt kurs, vil studentene bli vurdert 12 ganger i løpet av semestre dersom våre kolleger på de andre kursene er like vurderingshissige. Det er kanskje i meste laget om studentene også skal få tid til å arbeide med lærestoffet? Det er viktig at forelesere som underviser parallelle kurs tar en liten prat slik at det totale vurderingspresset ikke blir for stort og slik at eksamenene blir rimelig fordelt gjennom semesteret.

### **Karaktersetting på laveregradskurs**

Som vi alle vet, fører kvalitetsreformen med seg et nytt karaktersystem. Vi skal bruke en skala med ståkarakterene A-E og strykkarakteren F (i tillegg kan vi bruke bestått/ikke bestått). Denne karakterskalaen har fått både en ”kvalitativ” og en ”kvantitativ” beskrivelse. Den generelle *kvalitative* beskrivelsen er gitt ved følgende tabell (der det er fagmiljøenes oppgave å fylle ut den siste kolonnen).

<b>Symbol</b>	<b>Betegnelse</b>	<b>Generell, kvalitativ beskrivelse av vurderingskriterier</b>	<b>Fagspesifikk beskrivelse av vurderingskriterier</b>
<b>A</b>	Fremragende	Fremragende prestasjonsomklart utmerker seg. Viser stor grad av selvstendighet.	

<b>B</b>	Meget god	Meget god prestasjon som ligger over gjennomsnittet. Viser evne til selvstendighet.	
<b>C</b>	God	Gjennomsnittlig prestasjon som er tilfredsstillende på de fleste områder.	
<b>D</b>	Nokså god	Prestasjon under gjennomsnittet, med en del vesentlige mangler.	
<b>E</b>	Tilstrekkelig	Prestasjon som tilfredstiller minimumskravene, men heller ikke mer.	
<b>F</b>	Ikke bestått	Prestasjon som ikke tilfredstiller minimumskravene	

Den *kvantitative* beskrivelsen ser slik ut:

<b>Symbol</b>	<b>Betegnelse</b>	<b>ECTS-fordeling</b>	<b>Rammeverdier</b>
<b>A</b>	Fremragende	10%	8-12 %
<b>B</b>	Meget god	25%	20-30%
<b>C</b>	God	30%	24-36%
<b>D</b>	Nokså god	25%	20-30%
<b>E</b>	Tilstrekkelig	10%	8-12%
<b>F</b>	Ikke bestått	—	—

Det sies om rammeverdiene at ”statistisk normalfordeling ikke skal brukes i hvert enkelt emne eller fag i hvert enkelt semester. Fordelingen må framkomme over tid. Departementet har ikke angitt hvilket tidsperspektiv de ser for seg når det gjelder dette punktet, men en periode på minst 5 år er ikke utenkelig. Når alle eksamener i et større fagmiljø blir sett under ett innenfor en slik tidsramme, og for flere eksamener samlet, bør fordelingen etter Studieutvalgets mening ligge innenfor [rammeverdiene].” (Omtalen av det nye karaktersystemet er basert på brev fra Universitets- og høyskolerådet datert 29/5-2002.)

Det er ikke helt klart hvordan man tenker seg samspillet mellom den kvalitative og den kvantitative beskrivelsen. I praksis blir det vel gjerne slik at man bruker den kvantitative som hovedmodell der man har store kull og lett sammenlignbare eksamensbesvarelser, mens man tyr til den kvalitative når disse betingelsene ikke er oppfylt.

Ved fakultetet har vi allerede en del erfaring med den nye karakterskalaen, og tabellen nedenfor gir en oversikt over karaktergivingen (noen kurs med svært få studenter er utelatt). Den første kolonnen angir hvor mange studenter som møtte til eksamen, den andre kolonnen hvor mange som bestod, og den tredje angir strykprosenten. De seks neste kolonnene angir henholdsvis hvor mange studenter som fikk hvilken karakter (øverste linje), og hvor mange prosent dette utgjør av de beståtte besvarelsene (nederste linje). Den ”ideelle” fordelingen i nederste linje er altså 10%, 25%, 30%, 25%, 10%. Gir vi karakterene tallverdiene A=5, B=4,

C=3, D=2, E=1, kan vi regne ut en snittkarakter, og det er denne som er angitt i siste kolonne ("ideell" karakterfordeling gir snittkarakteren 3.0). De fleste resultatene i tabellen er fra våren 2002, men noen av matematikkresultatene er fra høsten 2001. Den nederste linjen i tabellen angir det samlede resultatet.

Kurs	Møt t	Be- tått	Stryk %	A (=5)	B (=4)	C (=3)	D (=2)	E (=1)	F (=0)	Snitt- karak.
<b>FY 040</b>	66	60	9.1%	5	22	6	19	8	6	2.95
				8.3%	36.7%	10.0%	31.7%	13.3%	—	
<b>FY 041</b>	19	18	5.3%	6	3	5	3	1	1	3.56
				33.3%	16.7%	27.8%	16.7%	5.6%	—	
<b>FY-ME 100</b>	76	73	3.9%	8	19	24	16	6	3	3.10
				11.0%	26.0%	32.9%	21.9%	8.2%	—	
<b>G 120</b>	13	13	0%	2	6	3	2	0	0	3.62
				15.4%	46.2%	23.1%	15.4%	0%	—	
<b>G 140</b>	13	13	0%	3	6	3	1	0	0	3.85
				23.1%	46.2%	23.1%	7.7%	0%	—	
<b>INF 101</b>	204	185	9.3%	28	56	54	33	14	19	3.28
				15.1%	30.2%	29.2%	17.8%	7.6%	—	
<b>INF 102</b>	365	334	8.5%	26	93	136	76	3	31	3.19
				7.8%	27.8%	40.7%	22.8%	0.9%	—	
<b>INF 130</b>	36	32	11.1%	4	6	13	6	3	4	3.06
				12.5%	18.8%	40.6%	18.8%	9.4%	—	
<b>KJ 101</b>	131	108	21.3%	14	31	37	20	6	23	3.25
				13.0%	28.7%	34.3%	18.5%	5.6%	—	
<b>MAT 100A (H01)</b>	42	38	9.5%	6	19	7	5	1	4	3.63
				15.8%	50.0%	18.4%	13.2%	2.6%	—	
<b>MAT 100B (H01)</b>	135	107	20.7%	13	34	35	18	7	28	3.26
				12.1%	31.8%	32.7%	16.8%	6.5%	—	
<b>MAT100C (H01)</b>	76	58	23.7%	7	18	19	7	7	18	3.19
				12.1%	31.0%	32.8%	12.1%	12.1%	—	
<b>MAT 100C (V02)</b>	57	47	17.5%	4	12	20	11	0	10	3.19
				8.5%	25.5%	42.5%	23.4%	0%	—	
<b>MAT 110A</b>	58	50	13.8%	4	21	18	7	0	8	3.44
				8%	42%	36%	14%	0%	—	
<b>MAT 110B</b>	86	80	7.0%	10	36	23	11	0	6	3.56
				12.5%	45%	28.8%	13.8%	0%	—	
<b>TOTALT</b>	<b>1377</b>	<b>121 6</b>	<b>11.7 %</b>	<b>140</b>	<b>382</b>	<b>403</b>	<b>235</b>	<b>56</b>	<b>161</b>	<b>3.26</b>
				<b>11.5 %</b>	<b>31.4 %</b>	<b>33.1 %</b>	<b>19.3 %</b>	<b>4.6%</b>	—	

Som man ser, er karaktergivingen noe "snill" — de beste karakterene A, B og C er overrepresentert med henholdsvis 11.7%, 31.4% og 33.1% (mot "standardens" 10%, 25% og 30%), mens de dårligste karakterene D og E er underrepresentert med henholdsvis 19.3% og 4.6% (mot "standardens" 25% og 10%). For de fleste karakterene ligger våre resultater også

utenfor rammeverdiene A: 8 - 12%, B: 20-30%, C: 25 - 35%, D: 20 - 30 %, E: 8 - 12%. Gjennomsnittskarakter er 3.26 mot "standardens" 3.0. Bare ett kurs (FY 040) har en gjennomsnittskarakter som er dårligere enn 3.0, mens alle de andre kursene ligger over "normalen".

Ser vi bort fra geologene som har svært få studenter, er det matematikerne som er de største "snillistene" i oversikten ovenfor. Dette skyldes hovedsakelig at man har hatt problemer med å kalibrere karaktersettingen på hjemmeeksamenene – her har det blitt mange A'er og B'er. Kanskje bør man bli litt flinkere til å minne seg selv om at A faktisk står for "fremragende" og at det er "B" som står for "meget godt"? (Jeg vil likevel tillate meg et lite hjertesukk på vegne av faglærerne: Det er til tider tungt å skulle gjennomføre en pedagogisk oppgradering parallelt med innføringen av et nytt karaktersystem med strenger normalisering - man føler at man får studentene til å yte stadig mer, uten å kunne gi dem tilbakebetaling i form av bedre karakterer!)

På slutteksamen har vi i matematikk stort sett fulgt følgende strategi. Vi gir poeng på hver oppgave og setter strykgrensen til 40% av totalscoren (dette tilsvarer strykgrensen i det gamle systemet). Bestått delen av skalaen deler vi inn i fem intervaller, ett for hver ståkarakter, med de midterste intervallene litt bredere enn kantintervallene. Til slutt kontrollerer vi skalaen ved å se hva slags besvarelser som nå havner i de ulike intervallene. Kanskje fører denne sjekken til at vi justerer karaktergrensene litt.

Med de oppgavetradisjonene vi har i matematikk, ser denne fremgangsmåten ut til å gi en karakterfordeling som ligger tett opp til den vi ønsker. Andre fag har kanskje andre oppgavetradisjoner (f.eks. en annen fordeling av lette og vanskelige oppgaver) som gjør at de må bruke skalaen annerledes. I mer deskriptive fag er det kanskje ikke naturlig å bruke en poengskala i det hele tatt, og man må da finne frem til andre kriterier for hvordan de nye karakterene skal brukes. Før man lager slike kriterier kan det være lurt å skaffe seg en oversikt over hvordan man har brukt det gamle karaktersystemet – det er ikke sikkert karakterfordelingen er slik man forestiller seg!

### **Karaktersetting på hovedfagsoppgaver**

I avsnittet ovenfor har jeg beskrevet en tilnærming til kurskarakterene der den kvantitative beskrivelsen spiller hovedrollen. For masteroppgaver har jeg vanskelig for å se at denne beskrivelsen kan spille en like sentral rolle. Én grunn er at de færreste av oss kommer borti så mange masteroppgaver at det gir mening å drive statistikk av denne typen, en annen er at de

forskjellige hovedfagene har så ulik rekrutteringsprofil at det lett kan bli urimelig å normalisere innenfor et enkelt fag. Dette betyr imidlertid ikke at vi helt skal se bort fra den kvantitative beskrivelsen — vi kan f.eks. bruke den som en ledesnor i utarbeidelsen av de kvalitative kriteriene. Ett enkelt eksempel: Ifølge den kvantitative beskrivelsen skal rundt 10% av studentene ha karakteren A. Det betyr at A skal være en relativt eksklusiv toppkarakter, og vi bør derfor ikke lage kvalitative beskrivelser som leder til at halvparten av studentene får denne karakteren! Tilsvarende er karakterene D-E tenkt å dekke rundt 35% av dem som består, og vi bør derfor lage kriteriene slik at det er mulig å få dårligere karakter enn C!

Med utgangspunkt i denne tankegangen har jeg laget et forslag til kvalitative karakterbeskrivelser for masteroppgaver i matematikk. Denne beskrivelsen har vært sirkulert blant matematikere og ser ikke ut til å ha vakt voldsom bestyrrelse så langt. Jeg presenterer derfor hovedtrekkene her som et eksempel på hvordan man muligens kan tenke også i andre fag.

La oss begynne med en statistikk som jeg for noen få år siden laget over hovedfagskarakterene til alle som hadde avlagt cand.scient.-eksamen i matematikk. Tabellen nedenfor viser den prosentvise karakterfordelingen på hovedfagsoppgaven (den nederste linjen angir den kumulative fordelingen).

1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1-2.	2.6-3.	3.1-3.	3.6-4.0
											5	0	5	
0	2.3	12.	18.	13.	7.3	10.	6.2	7.3	7.9	2.8	7.3	2.9	0	0.6
		9	0	5		7								
0	2.3	15.	33.	46.	54.	64.	70.	78.	86.	88.	96.4	99.4	99.4	100
		2	2	7	0	7	9	2	1	9				

Vi legger merke til at en tredjedel av studentene har fått 1.3 eller bedre, mens bare 10% har fått dårligere enn 2.0. Medianen ligger rundt 1.45, mens gjennomsnittet er 1.61.

Jeg tror ikke vi skal ta sikte på å lage en konverteringsnøkkel som automatisk regner om mellom det nye og det gamle karaktersystemet (det ville blant annet medføre at vi tar med oss de skjevhetene som kan ha utviklet seg mellom fagområder), men dersom vi skulle ha gjort det, tyder tabellen ovenfor på at en slik nøkkel for matematikk kanskje burde ha sett omtrent slik ut: A: 1.0-1.2, B: 1.3-1.4, C: 1.5-1.9, D: 2.0-2.5, E: 2.6-4.0 (her har jeg tatt noe høyde for at matematikkstudentene generelt er svært godt kvalifisert, og at vi derfor bør tillate oss en litt skjev fordeling). Selv om vi ikke skal basere oss på en konverteringsnøkkel, føler jeg at vi kan ta denne inndelingen som et utgangspunkt for våre kvalitative beskrivelser av de

nye karakterene. Hva er det som skiller en student i intervallet 1.0-1.2 fra en i intervallet 1.3-1.4? Etter min oppfatning har vi brukt karakterene 1.1 og 1.2 til å skille ut de aller største talentene, de som har tatt styringen over sitt eget hovedfagsprosjekt, som selv har generert ideer og løsningsmetoder, og som det nærmest ville være en skandale om vi ikke greide å skaffe en stipendiatstilling til. Karakterene 1.3-1.4(1.5) har vi brukt på studenter som vi svært gjerne vil ha som doktorgradsstudenter, men som vi i større grad føler må konkurrere med andre om stipendiatstillingene — de har sluppet til i gode år med mange stipender og få søkere, og måttet gå tomhendt i dårligere år. Erfaringen har sikkert vist at mange av dem som har sluppet til etter å ha fått 1.3 eller 1.4 på hovedfagsoppgaven, har gjort det bedre senere i livet enn noen som har fått 1.1 eller 1.2, men slike utslag vil man få med alle karaktersystemer — det kan skyldes at folk modnes forskjellig, eller at ikke alle hovedoppgaver gir like gode muligheter for initiativ og kreativitet.

Ut i fra disse betraktningene har jeg prøvd å lage karakterbeskrivelsene slik at A´en fanger opp de absolutt beste — de som vi under ingen omstendigheter vil gå glipp av — mens B´en fanger opp de vi svært gjerne vil ha som doktorgradsstudenter, men som i en verden med begrensede ressurser må konkurrere med hverandre om stipendiatstillingene. C´en forsøker å fange opp de dyktige håndverkerne, de som gjør sine saker solid og skikkelig, men som ikke har den fantasien eller driven som gjør det naturlig at de går videre med et doktorgradsstudium. Jeg har ikke ønsket å diskvalifisere disse studentene fra doktorgradsstudiet, men har valgt å bruke formuleringen ”kandidaten kan være kvalifisert for et doktorgradsstudium etter en helhetsvurdering”. Det er to grunner til dette — for det første finnes det eksempler på at studenter i denne kategorien har fått doktorgradsmuligheter utenfor ”ren” matematikk og greid seg aldeles utmerket, og for det andre får studenter med tilsvarende kvalifikasjoner jevnlig doktorgradsstipend i andre fag. Selv om denne studentgruppen har få muligheter til å nå opp i dagens konkurransesituasjon i ”ren” matematikk, virker det urettferdig å diskvalifisere dem i utgangspunktet. Karakteren D tenker jeg meg brukt på den studentgruppen som også arbeider solid og skikkelig, men der mangel på oversikt og/eller teknisk usikkerhet begynner å bli påtagelig. E`en har jeg reservert for de studentene som er helt på kanten til ikke å komme igjennom. I et veiledet studium uten tidsfrist vil det neppe være mange av disse kandidatene, men karakterens eksistens markerer at D-studentene står med klar margin. (Av de hovedfagsoppgavene jeg har sett, er det bare én jeg med sikkerhet ville ha gitt E etter kriteriene nedenfor, og den tilhørte en student som hadde brutt kontakten med veileder og levert inn oppgaven uten å ha fått tilbakemelding).

Et par kommentarer til før jeg kommer med forslagene til kriterier. Den første er at disse kriteriene vil være offentlig tilgjengelig og derfor bør være rimelig forståelige også for

dem som ikke har sensurert eller veiledet hovedfagsstudenter. De må altså skrives i et mindre internt språk enn det jeg hittil har benyttet. Den andre kommentaren er at kriterier av denne typen aldri passer helt på alle studenter — det vil alltid finnes studenter som passer i kategori B etter visse kriterier og i kategori D etter andre. Et typisk eksempel er at evnen til å forstå matematiske resonneringer ikke alltid står i rimelig samsvar med evnen til å formulere denne forståelsen skriftlig!

Forslaget til beskrivelser finnes i tabellen nedenfor. Legg merk til at den (og betraktningene ovenfor) gjelder en lang masteroppgave uten tidsfrist, altså en arvtager til den klassiske hovedoppgaven. Kriteriene for en kort oppgave med tidsfrist bør nok se noe annerledes ut (jeg har laget forslag for slike oppgaver også, men gjengir ikke det her).

<b>Forslag til karakterbeskrivelser for lang masteroppgave i matematikk uten tidsfrist</b>		
<b>Symbol</b>	<b>Betegnelse</b>	<b>Beskrivelse</b>
<b>A</b>	Fremragende	Kandidaten har fullstendig oversikt over den teorien som brukes. Hun/han viser en usedvanlig grad av modenhet og fantasi, og har selv generert viktige problemstillinger og løsningsmetoder i oppgaven. Fremstillingen er klar og presis. Kandidaten viser et forskertalent som man opplagt bør ta vare på.
<b>B</b>	Meget god	Kandidaten har meget god oversikt over teorien som brukes, og kan anvende den selvstendig. Hun/han har meget gode tekniske ferdigheter og kan gjennomføre kompliserte resonneringer på egen hånd. Fremstillingen er klar og presis. Kandidaten er klart kvalifisert for et doktorgradstudium.
<b>C</b>	God	Kandidaten har god oversikt over teorien som brukes, og kan til en viss grad anvende den selvstendig. Hun/han har gode tekniske ferdigheter og kan gjennomføre matematiske resonneringer på egen hånd. Fremstillingen er gjennomgående god, men kan ha mindre mangler. Kandidaten kan være kvalifisert for et doktorgradstudium etter en helhetsvurdering.
<b>D</b>	Nokså god	Kandidaten kjenner hovedtrekkene i teorien som brukes, men viser tegn på manglende oversikt. Hun/han kan gjennomføre tekniske rutineargumenter, men trenger hjelp til å binde teorien sammen. Fremstillingen er forståelig, men kan være ustrukturert og røpe manglende innsikt på enkelte områder. Kandidaten vil normalt ikke være kvalifisert for et doktorgradstudium.
<b>E</b>	Tilstrekkelig	Kandidaten kjenner hovedtrekkene i teorien som brukes, men viser sviktende forståelse og oversikt på flere punkter. Hun/hun gjennomfører noen matematiske resonneringer, men det finnes flere eksempler på feil og misforståelser. Kandidaten makter i liten grad å arbeide selvstendig med stoffet. Fremstillingen er stort sett akseptabel, men er til tider uryddig og vanskelig å forstå. Kandidaten er ikke kvalifisert for et doktorgradstudium.

<b>F</b>	Ikke bestått	Kandidaten har bare et overfladisk kjennskap til teorien som brukes, og store deler av fremstillingen er preget av klare feil og misforståelser. Han/hun viser stor usikkerhet i matematiske argumenter og beregninger. Fremstillingen er ustrukturert og vanskelig å følge.
----------	--------------	--

Det er kanskje dumt å argumentere for hardt for et forslag som bare er ute til lufting, men jeg har lyst til å påpeke noen fordeler som dette forslaget har sammenlignet med andre jeg har prøvd.

1. Kriteriene virker rimelig klare og operasjonelle (i hvert fall for meg!)
2. Vi får brukt store deler av karakterskalaen og det på en måte som ikke er helt i utakt med den ”kvantitative” beskrivelsen.
3. Vi slipper et altfor hardt press på karakteren A — B blir normal karakteren for en meget god oppgave, og A blir en bonus for de aller beste. Man kan utmerket godt få en stipendiatstilling selv om man ikke har fått A på masteroppgaven.
4. Det blir rimelig samsvar mellom karakterbetegnelsen og karakterbeskrivelsen — det er de studentene som nå havner i kategori A som vi (jeg?) tenker på som ”fremragende”. I min terminologi vil en 1.3 student som regel være ”meget god” og ikke ”fremragende” (hva hun/han blir senere i livet får være en annen sak!).

Et tungtveiende motargument er at NTNU allerede opererer med en konverteringsnøkkel mellom nye og gamle karakterer som er snillere enn det systemet jeg har beskrevet her. Det betyr at vi er helt avhengig av en nasjonal samordning. En innstramning av praksisen bør kobles til overgangen fra cand.scient. til master (det er ikke så mange som får en cand.scient. med bokstavkarakterer, så vi får hele tåle at disse kanskje får en liten fordel).

Forslaget ovenfor har som nevnt vært sirkulert blant matematikere uten å vekke opphisselse (så vidt jeg vet!). Jeg er spent på om disse beskrivelsene gir mening også for representanter fra andre fag, og om de i så fall kan danne mønster for tilsvarende beskrivelser der.