

Liv Langfeldt og Elisabeth Hovdhaugen

Kriterie- og metodegrunnlag for (re)akkreditering av høyere utdanning



© NIFU STEP Studier av innovasjon, forskning og utdanning
Wergelandsveien 7, 0167 Oslo

Arbeidsnotat 1/2006
ISSN 1504-0887

For en presentasjon av NIFU STEPs øvrige utgivelser, se www.nifustep.no

Innhold

Sammendrag	3
1 Innledning	4
2 Fortolkningsrom	5
2.1 "Faglig nivå"	5
2.2 "Dokumenterte resultater"	9
3 Hvordan vurdere faglig nivå og dokumenterte resultater?.....	11
3.1 Kriterier	11
3.2 Metoder	14
3.3 Metoder og kriterier: Fordeler, ulemper og dilemmaer.....	16
3.4 Usikkerhet	23
4 Overordnede spørsmål og valgmuligheter	24
4.1 Kvalitetssikring i høyere utdanning: en avveining av mange hensyn	24
4.2 Å balansere ulike hensyn i faglige vurderinger.....	25
4.3 Hvordan ivareta fagligheten i vurderingsprosessen?.....	28
4.4 Avsluttende kommentar	30
Referanser	31

Sammendrag

I dette notatet drøftes styrker og svakheter ved ulike kriterie- og metodevalg som en må ta stilling til når praksis for akkreditering av studietilbud skal etableres. Fokus er på vurdering av faglig nivå og dokumenterte resultater.

I notatet deles de aktuelle faktorene og kriteriene ved vurdering av faglig nivå inn i tre hovedaspekter: (1) Det faglige innholdet i studietilbudet, (2) nivået på undervisningen/formidlingsaspektet i studietilbudet, og (3) hvilken faglig kontekst studietilbudet er en del av. Hver av disse aspektene har ulike tolkningsmuligheter og kan utdypes gjennom ulike kriterier og dokumenteres på ulike måter.

For å ivareta legitimitet, faglighet, likebehandling, forutsigbarhet og rettsikkerhet, krever akkrediteringsprosesser kriterier med høy ”treffsikkerhet” (høy validitet) og at vurderingene er pålitelige i den betydning at ulike målinger/vurderinger gir likt resultat (høy reliabilitet). I tillegg bør akkrediteringsprosesser ikke være for kostnadskrevenende.

Notatet tar bl.a. for seg ulike kvantitative data som indikatorer på kvaliteten på de aktuelle aspektene av en utdanning. Disse indikatorene har imidlertid uklar sammenheng med faglig nivå og trenger å kombineres med, og ses i sammen med, andre kriterier. Andre forhold krever fagfellevurdering og blir mer kostbare å vurdere, særlig om man skal sikre god reliabilitet ved å benytte et stort utvalg av fagfeller.

Videre drøftes muligheter og begrensinger ved å benytte fagfellevurdering og faglig skjønn i akkrediteringsprosessen. Forskningen rundt fagfellevurdering (peer review) viser at det faglige ståsted til den eller de som velges til å foreta en faglig vurdering ofte vil ha betydning for evalueringsutfallet. Fagfellevurdering er slik sett en prosess der ulike vurderinger og ulik ekspertise settes mot hverandre for å få et best mulig grunnlag for en beslutning. Hvilken ekspertise som gis adgang til prosessen kan derfor ha stor betydning for resultatet.

En generell konklusjon er at det er vanskelig å måle faglig nivå, at ethvert kriterium er usikkert og at enhver vurderingsprosess vil måtte kombinere ulike kriterier og tolke dem i lys av hverandre. I siste del av notatet diskuteres ulike fremgangsmåter/alternativer for å balansere ulike hensyn og redusere usikkerhet i vurderingene.

Blant alternativene som presenteres er (1) en to-trinns-prosess som skiller skjønnsbaserte vurderinger og vurderinger som ikke er skjønnsbaserte, (2) mer åpne/fleksible kriterier som gir rom for tilpasning til det enkelte evalueringsobjekt/fag, (3) utarbeidelse av egne sett av kriterier og standarder for de enkelte fag, eller (4) å etablere en praksis med forundersøkelser for hver akkrediteringsprosess der en tilpasser og tolker kriteriesettet og de generelle standardene i forhold til det enkelte fag som skal vurderes. Videre drøftes hvilket handlingsrom og hvilken myndighet fagfeller gis i vurderingene. Her vektlegges bl.a. at om en har ulike typer kriterier uten klare regler for hvordan de skal veies mot hverandre, vil sammensetningen av evalueringsutvalget kunne være den beslutningen som avgjør hvordan kriteriene ”vektes”.

1 Innledning

Bakgrunnen for notatet er en henvendelse fra NOKUT med ønske om å få drøftet tolkningsrom og dilemmaer knyttet til å vurdere faglig nivå og dokumenterte resultater i høyere utdanning. Utgangspunktet er at i "Forskrift om akkreditering, evaluering og godkjenning etter lov om universiteter og høyskoler og lov om private høyskoler" (UFD 2.1.2003) heter det at "Ved revisjon av etablerte studietilbud skal organet også legge vekt på utdanningens faglige nivå og dokumenterte resultater".

Spørsmålene som drøftes er:

- Hvordan kan begrepene faglig nivå og dokumenterte resultater fortolkes? Kan disse begrepene forstås på ulike måter, og hva er konsekvensene av de ulike forståelsesrammene?
- Hvilke kriterier er relevante å trekke inn for å kunne vurdere faglig nivå og resultater? Har noen kriterier mer relevans enn andre? Hvilket faglig eller erfaringsmessig grunnlag eksisterer som kan si noe om kriteriers eventuelle validitet og reliabilitet i akkrediteringssammenheng?
- Hvilke metodemessige konsekvenser har forskriftens vektlegging av faglig nivå og resultater? Gitt bruken av sakkyndige og fagekspertter i akkrediteringssammenheng, hvilken rolle bør klare kriterier og formelle standarder ha?

Formålet er å drøfte styrker og svakheter ved ulike forståelsesrammer, kriterie- og metodevalg og identifisere sentrale dilemmaer som NOKUT må ta stilling til når de skal etablere praksis for akkreditering av etablerte studietilbud.

2 Fortolkningsrom

I denne delen av notatet drøftes hvordan begrepene ”faglig nivå” og ”dokumenterte resultater” kan fortolkes. Hvilke ulike fortolkningsmåter er aktuelle og hva er konsekvensene av ulike fortolkningsmåter?

2.1 ”Faglig nivå”

”Nivå” henviser til graden av noe, mens ”faglig” i vår sammenheng henviser til innholdet i utdanningen, til det som skal læres. ”Faglig nivå” kan slik forstås som hvor avansert utdanningen er, eller hvilket faglige trinn den er på, eksempelvis om den er på nivå med det som normalt vil oppfattes som lavere grads utdanning eller på nivå med høyere grads utdanning.

Med en slik tolkning blir ”faglig nivå” et mer begrenset spørsmål enn eksempelvis et spørsmål om hvor høy *faglig kvalitet* det er på utdanningen. Spørsmålet om faglig kvalitet kan ses på som hvor godt studietilbudet er gitt hvilket faglig nivå det er på. Samtidig er det mulig å inkludere spørsmålet om faglig kvalitet i spørsmålet om faglig nivå, i og med at det kan hevdes at jo høyere faglig kvalitet utdanningen har, jo høyere nivå er den på (også innen sitt nivå). Studiekvalitet er på den annen side et mer generelt og mangetydig begrep som eksempelvis også kan omfatte forhold knyttet til studentenes trivsel og andre forhold som ikke er direkte knyttet til den faglige siden av et studium, men som likevel kan ha stor betydning for studentenes læring (Wiers-Jenssen et al. 2002).

Organer som akkrediterer høyere utdanning vurderer en rekke ulike forhold som kan relateres til utdanningens faglige nivå. Vurderingene til det britiske *Quality Assurance Agency for Higher Education* omfatter eksempelvis pensum, undervisning, læring, støtte og veiledning til studentene, læringsressurser, vurdering /evaluering, studentenes progresjon og prestasjon og kvalitetsforvaltning og forbedring.¹ Disse ulike faktorene og kriteriene faller slik vi ser det inn under tre ulike hovedaspekter: (1) Det faglige innholdet i studietilbudet (dvs. hva det undervises i og hva studentene er forutsatt å lære), (2) nivået på selve undervisningen, dvs. hvilke egenskaper formidlingsaspektet i studietilbudet har, og (3) hvilken faglig kontekst studietilbudet er en del av. Også kriteriene som benyttes av tilsvarende organer i andre land kan innpasses i denne tredelingen (ref. Stensaker og Harvey 2006). For å strukturere diskusjonen har vi valgt å dele den inn etter disse tre dimensjonene, og vise at hver av disse aspektene ved faglig nivå kan forstås på mange ulike måter.

¹ http://www.qaa.ac.uk/reviews/subjectReview/handbook/srhbook_part2.asp

Utdanningens faglig innhold

Faglig innhold i en utdanning relaterer seg til det faglige stoffet det blir undervist i – både pensum og annen faglig innsikt som blir formidlet – og avhenger slik ikke utelukkende av definert pensum men også av undervisningspersonalets kompetanse (her omtalt som lærerne).

Både pensum og annen faglig innsikt som blir formidlet, kan defineres både ved faglig omfang og dybde. Mange studier har en rammeplan som definerer læringskravene til studiet, for eksempel sykepleierutdanningen, allmennlærerutdanningen og praktisk-pedagogisk utdanning. I den grad slike dokumenter definerer en klar ramme for hva som skal være pensum i et gitt fag (omfang og dybde) har man her en standard å forholde seg til når det gjelder faglig innhold. I andre fag, fremfor alt universitetsfag, eksisterer det ikke en slik plan for hva som skal inngå i utdanningen. Hva som skal inngå i et fag, hva som skal læres mer grundig eller bare overfladisk er dermed temaer som kan debatteres. Imidlertid er det mulig å anta at det finnes en form for konsensus i miljøene som underviser i faget om hva som skal inngå i den grunnleggende utdanningen. Dette illustreres blant annet gjennom at lærestedene anerkjenner hverandres utdanninger slik at det er mulig å komme med grunnutdanning fra et lærested og få lov til å fortsette utdanningen ved et annet lærested (evt. etter visse tilleggskurs/deleksamener). I kapittel 3 diskuterer vi bl.a. hvilken rolle fagfellevurdering kan spille ved kartlegging av nivået på innholdet i en utdanning – med eller uten rammeplan.

Videre er undervisningspersonalets kvalifikasjoner viktige for utdanningens faglige innhold. Hvor oppdatert lærerne er i forhold til fagets forskningsfront vil bli synliggjort gjennom hva som formidles. En kunnskapsrik lærer vil ha mer å tilføre utdanningen enn en lærer som er mindre kunnskapsrik. Her er med andre ord også bredde og dybde viktig, lærerne kan ha både ulik bredde og dybde i undervisningsfaget og i ulik grad være oppdatert på ny forskning innen spesifikke felt. Lærernes faglige innsikt vil også kunne gjenspeile seg i utformingen av pensum. Det vil si at lærernes oppfatning av hva som er relevant på for eksempel et innføringskurs vil kunne påvirke hva som er definert pensum og hva som ellers formidles i undervisningen.

Samtidig som alle disse forholdene vil påvirke faglig innhold vil det kunne være svært ulike meninger om hvor stor betydning de har, hvordan faglig innhold kan dokumenteres og ikke minst hva som bør være det faglige innholdet i en gitt utdanning. Dette kommer vi tilbake til under.

Faglig formidling i utdanningen

Det er også et stort tolkningsrom når det gjelder formidlingsaspektet. Lærernes faglige innsikt vil også kunne påvirke selve formidlingen av faget. En oppdatert og kunnskapsrik lærer vil eksempelvis kunne ha mer engasjement og faglig trygghet i sin formidling. Faglig

innsikt virker slik sammen med pedagogiske evner, en flink pedagog vil også ha mye kunnskap å formidle.

Undervisningsformer kan også ha betydning for faglig nivå. Avhengig av hva slags kunnskap og innsikt som skal formidles, vil det kunne være ulike metoder som er best egnet (eksempelvis forelesninger, gruppeundervisning, laboratorieundervisning eller individuell veiledning).

Likeledes kan ulike vurderings- og eksamensordninger være mer eller mindre egnet i ulike fag og også i ulik grad fremme et høyt faglig nivå. Vurderingsmetoder/eksamensordninger er også relevant som et mer indirekte kriterium ved vurdering av faglig formidling: Hva slags sensorer som benyttes og hvordan eksamen er lagt opp, er relevant når eksamensresultater benyttes som dokumentasjon på hvor god den faglige formidlingen er.

Faglig kontekst

Av de tre aspektene ved faglig nivå favner ”faglig kontekst” bredest. Svært mange ulike kontekstuelle forhold kan ses som relevante for en utdannings faglige nivå. Eksempelvis vil den faglige infrastrukturen og kobling til andre faglige aktiviteter – som FoU eller yrkesrettet praksis – kunne ses som viktige faktorer som har betydning for bedømmning av faglig nivå.

I hvilken grad kobling til andre faglige aktiviteter, som FoU eller yrkesrettet praksis, ses som et relevant aspekt når en vurderer faglig nivå, vil kunne variere mellom fag og etter om det er lavere eller høyere grads studier det dreier seg om.

Også selve fagmiljøets attraktivitet (eksempelvis popularitet som forsknings- og studiested) vil kunne ses som en indikator på faglig nivå. Slike mål kan reflektere hva den relevante ”opinionen” anser som de sterkeste og de svakeste studiene, men vil også kunne være kontroversielle. I den grad det faktisk finnes et samlet syn på den faglige attraktiviteten til de ulike utdanninger innen et fag, er det ikke sikkert at disse synspunktene har noe godt faglig fundament. Det er også mulig at en slik mening/rangering har et adekvat fundament, men likevel er irrelevant i forhold til en gitt evaluering (fordi fundamentet faller utenfor det evalueringen skal ta stilling til).

Generelt kan vi si at faglig kontekst er det ”bredeste” og minst opplagte aspektet i en vurdering av faglig nivå. Det vil kunne være betydelig uenighet om hvor relevante ulike kontekst-kriterier er når faglig nivå skal vurderes.

Tabell 1 Illustrasjon på tolkningsrom: Ulike aspekter av og mulige indikatorer på en utdannings faglige nivå

Hovedaspekter	Underaspekter (eksempler)	Vurderingsmåte/indikatorer (eksempler)
A: Faglig innhold	Pensums faglige omfang Pensums faglige dybde	S: Sammenlikne med omfang og dybde i tilsvarende studier S: Krav til bestått eksamen (krav til mestring av pensum)
	Lærernes faglige innsikt Lærernes faglige oppdatering	K: Formelle kvalifikasjoner K/S: Nasjonalt og internasjonalt faglig nettverk og samarbeid K/S: Vitenskapelig publisering K/S: Yrkesrettet praksis
B: Faglig formidling	Lærernes faglige innsikt Lærernes pedagogiske evner	K: Lærernes formelle faglige og pedagogiske kvalifikasjoner K: Eksamensresultater K: Studentevalueringer
	Undervisnings- og vurderingsformer	S/K: Sammenlikning tilsvarende studier K: Forholdstall studenter per lærer K: Eksamensordninger (sammenlikning tilsvarende studier) K: Sensorenes formelle kvalifikasjoner K: Bruk av eksterne sensorer
C: Faglig kontekst	Lærested: Rutiner og tiltak for å sikre nivå på studietilbud	S: Faglige krav til egne studietilbud S/K: Faglig samarbeid og innflytelse på tvers av studietilbud/lærenes faglige frihet
	Lærested og studium: Faglig miljø/ fagmiljøets attraktivitet	S: Faglig anerkjennelse (eksterne kollegers meninger) K: Antall lærere og elever på studiet (minimumsstørrelse) K: Studenttilgang/opptakskrav/ inntakskvalitet K: Popularitet som arbeidssted (antall kvalifiserte søkere per utlyst stilling)
	Lærested og studium: Faglig infrastruktur	S/K: Sammenlikning med infrastrukturer i andre tilsvarende studier K: Tilgang på spesifikk infrastruktur (fagavhengige krav)
	Lærested: faglig bredde	K: Antall andre studier på lærestedet, antall andre fagområder
	Lærested: Formelt faglig nivå	K: Status som universitet/vitenskapelig høyskole/høyskole K: Rett til å tildele grader (antall Bachelorstudier, Masterstudier og Doktorgradsstudier)
	Kobling til andre faglige aktiviteter: Forskning og utvikling Yrkesrettet praksis (inkl. konsulentvirksomhet etc)	S/K: Lærenes FoU virksomhet S/K: Annen FoU virksomhet i fagmiljøet S/K: Lærenes fagrelaterte praksis S/K: Annen fagrelatert praksis i fagmiljøet

Hovedaspekter: Faglige hoveddimensjoner ved en utdanning

Underaspekter: Hva det kan være aktuelt å inkludere i en vurdering av faglig nivå

Vurderingsmåte/indikatorer: S = faglig/skjønnsbasert vurdering

K = kvantitativt mål/mulig med ikke-skjønnsbaserte mål

I tabell 1 har vi satt opp de tre diskuterte hovedaspektene ved faglig nivå, samt eksempler på aktuelle underaspekter og indikatorer. ”Faglig innhold” og ”Faglig formidling” er aspekter som oftest vil ses som avgjørende for faglig nivå – jo høyere score jo høyere nivå. Aspekter under ”Faglig kontekst” vil ofte ses mer som minimumskrav – om man ikke har en viss standard her kan det være til hinder for faglig nivå. ”Faglig innhold” og ”Faglig formidling” er også aspekter der ekspertise i de involverte fagene er påkrevd for (i alle fall

som del av) en meningsfull vurdering. Når det gjelder ”Faglig kontekst” er det større rom for å gjøre vurderinger uten ekspertise i de aktuelle fagene.

Det bør her også nevnes at oversikten utelater noen aspekter som vi tror normalt ikke vil oppfattes som direkte knyttet til faglig nivå. Blant annet er utdanningens samfunnsmessige attraktivitet, dens relevans og nytte, indikert eksempelvis ved kandidatens lønnsnivå og karriereprogresjon, ikke inkludert. Dette er likevel aspekter som vil kunne inkluderes under vårt neste punkt ”dokumenterte resultater”.

2.2 ”Dokumenterte resultater”

I likhet med faglig nivå kan ”dokumenterte resultater” i en utdanning tolkes på mange ulike måter. Her skal vi avgrense oss til ulike typer dokumentasjon/skriftlig informasjon som vil være aktuell som underlag for å vurdere faglig nivå. Siste kolonne i tabell 1 illustrerer noe av mulighetsrommet mht hva det kan være aktuelt å dokumentere. Under kommenterer og utdyper vi dette mulighetsrommet i form av aktuelle resultatindikatorer for hver av de tre aspektene: faglig innhold, formidling og kontekst. Vi illustrerer slik de mange ulike typer ”dokumenterte resultater” som kan ses som relevante for å vurdere faglig nivå. En ytterligere utdyping av de mest relevante indikatorene finnes i kapittel 3.

Dokumentasjon på faglig innhold

Mulig dokumentasjon på faglig innhold favner vidt. En kan eksempelvis be lærestedet om dokumentasjon på at pensum er i tråd med rammeplaner (der slike foreligger), eller at studiet blir godkjent som grunnlag for videre studier ved andre læresteder, og at innholdet slik vurderes å samsvare med innholdet i tilsvarende utdanninger. Tolker vi ”dokumenterte resultater” snevert er det imidlertid kun det siste (godkjenning ved andre læresteder) som er *resultat* av utdanningen.

En annen typisk resultatorientert indikator på at en utdanning har adekvat innhold ville være dokumentasjon på at kandidatene er ettertraktet på det fagspesifikke arbeidsmarkedet – eksempelvis at de er mer etterspurt enn kandidater fra konkurrerende læresteder, har høyere lønnsnivå eller bedre karriereprogresjon.

Det aspektet av faglig innhold som går på lærernes faglige innsikt og oppdatering kan også dokumenteres på mange ulike måter. Enkle kvantitative mål kan eksempelvis inkludere hvor mye tid som er avsatt til faglig oppdatering, inkl. FoU-arbeid, deltakelse i faglige nettverk, fagkonferanser og lignende, eller dokumentasjon på hvilke slike aktiviteter (og omfanget av dem) som de har deltatt i.

Dokumentasjon på faglig formidling

Den faglige formidlingen kan likeledes dokumenteres på ulike måter. Lærernes faglige innsikt og pedagogiske evner kan eksempelvis dokumenteres ved deres formelle kvalifikasjoner (andel med førstekompetanse/professorkompetanse, andel med ulike pedagogiske meritter), ved deres FoU virksomhet (eksempelvis kvantifisering av publiseringsvirksomheten) og eventuelt deres praksiserfaring. Dette er imidlertid mer innsatsfaktorer i utdanningen enn resultater av den – og slik aktuelt som indikatorer på faglig nivå, men mindre aktuelt ved dokumentasjon av resultater. Lærernes FoU-virksomhet kan likevel tolkes som et resultat av deres faglige virksomhet som er/bør være så nært knyttet til undervisningen at den er relevant når resultater av undervisningens faglige nivå skal dokumenteres. Tolker vi ”dokumenterte resultater” mer snevert står vi tilbake med aktuelle indikatorer av typen: studieprogresjon, andel som fullfører, eksamensresultater og studentevalueringer. Styrker og svakheter ved slike indikatorer kommer vi tilbake til i kapittel 3.

Dokumentasjon på faglig kontekst

For å dokumentere faglig kontekst er det mulig å bruke en rekke ulike indikatorer, bl.a. kvantitative mål på FoU-virksomhet, spørreundersøkelser eller søkertall som indikerer fagmiljøets attraktivitet, og sensorers formelle kompetanse (ref. siste kolonne i tabell 1). Med en snever tolkning av ”dokumenterte resultater” er det imidlertid i liten grad aktuelt å se faglig kontekst som et resultat av utdanningens faglige nivå. Unntaket er fagmiljøets attraktivitet. Eksempelvis indikatorer på faglig anerkjennelse og popularitet (som studenttilgang og kvalifiserte søkere på ulyste vitenskapelige stillinger) vil her kunne være aktuelle.

3 Hvordan vurdere faglig nivå og dokumenterte resultater?

I dette kapittelet drøfter vi sentrale kriterier på faglig nivå og mulighetene for å dokumentere og vurdere hvordan en utdanning scorer på disse kriteriene. I forhold til diskusjonen over snevrer vi inn mulighetsrommet noe, og konsentrerer oss om de kriterier og metoder som i dag benyttes av NOKUT og det som ellers synes å være aktuelle kriterier og metoder for å vurdere faglig nivå og resultater i en utdanning.

3.1 Kriterier

Som et utgangspunkt for gjennomgangen har vi i tabell 2, 3 og 4 sortert NOKUTs sentrale kriterier relatert til faglig nivå (etter mønster av begrepsapparatet i tabell 1). Vi har her plukket ut sentrale punkter i NOKUTs gjeldende forskrift om kriterier og standarder, uten å ha hatt ambisjoner om en full gjennomgang – både fordi forskriften for tiden er under omarbeiding og fordi vi har sett det som vår oppgave å drøfte overordnede kriterier og ikke detaljer (og går slik eksempelvis ikke inn på kriterier for lavere grads versus høyere grads utdanning).²

Tabell 2 NOKUTs sentrale kriterier relatert til faglig innhold

Underaspekt av faglig innhold	Eksempler fra NOKUTs forskrift
Studieplan/Pensums faglige omfang og dybde	<ul style="list-style-type: none">▪ Studiets innhold må være på nivå med tilsvarende studier på universitets- og høghskolenivå▪ Det må være en tydelig progresjon i studiet▪ Studiet må gi kunnskap om og erfaring med bruk av vitenskapelig metode, og være egnet til å utvikle forståelse, refleksjon og modning▪ Studentene skal gis innføring i FoU-virksomhet▪ Institusjonen må vise at området for doktorgradsstudiet har tilstrekkelig faglig bredde og dybde og indre faglig sammenheng
Lærernes faglige innsikt	<ul style="list-style-type: none">▪ Institusjonen skal ha egenkompetanse innenfor sentrale fagområder som inngår (institusjonsakkrediteringer)▪ Institusjonen må ha fast ansatte med kompetanse innenfor studiets kjerneområder▪ Minst 20 % av fagmiljøet bør normalt ha førstestillingskompetanse▪ Institusjonen må ha tilsatte i hovedstilling med kompetanse innenfor det som internasjonalt regnes som kjerneområder innenfor det aktuelle doktorgradsområdet.▪ Fagtilsatte tilknyttet doktorgradsstudiet må ha kompetanse på førsteamanuensisnivå, derav minst 50 % med professorkompetanse▪ For erfaringsbaserte mastergradsstudier vektlegges også fagtilsattes erfaring fra feltet, fagmiljøets nettverk mot praksisfeltet, ledelse og deltakelse i forskningsprosjekter, forsøks- og utviklingsarbeid og erfaring fra tilhørende grunnutdanning på feltet. (foreslått fjernet i ny forskrift)
Lærernes faglige oppdatering	<ul style="list-style-type: none">▪ Institusjonen skal vise at den er i stand til å utvikle sitt fagpersonale og sin undervisning i samsvar med studentenes behov, fagets og samfunnets utvikling (institusjonsakkrediteringer)

² Eksemplene i tabell 2-4 er fra "Forskrift om kriterier for akkreditering av institusjoner og standarder og kriterier for akkreditering av studietilbud i norsk høgre utdanning" (NOKUT 5.5.2003). Kriterier for vurdering av studietilbud på ulike nivå/ulike grader er inkludert. Forslag til ny forskrift ble sendt på høring 17.10.2005 og skal etter planen tre i kraft 1.1.2006.

Når det gjelder studiets innhold utdyper NOKUTs forskrift kravene på ulike måter: Innholdet skal være på nivå med tilsvarende studier, nivået skal være økende (progresjon i studiet), og det skal være tilstrekkelig bredde og dybde. I tillegg er det mer spesifikke krav til at studiene skal omfatte innføring i vitenskapelig metode og FoU-virksomhet (tabell 2)

Under ”lærenes faglige innsikt” legges det vekt på forholdsvis klare og formelle kriterier som det ikke kreves faglig skjønn for å vurdere: Formalkompetanse og fast ansettelse. Andre aktuelle kriterier kunne inkludere ”realkompetanse” i faget, noe som i høyere grad ville kreve faglig skjønn i vurderingene.

Tabell 3 NOKUTs sentrale kriterier relatert til faglig formidling

Aspekt av faglig formidling	Eksempler fra NOKUTs forskrift
Lærernes faglige innsikt og pedagogiske evner	<ul style="list-style-type: none"> ▪ Faglige og pedagogiske kvalifikasjoner vurderes etter de kriterier som til enhver tid er gjeldende for tilsetning og opprykk i vitenskapelige/kunstneriske stillinger.
Undervisningsformer	<ul style="list-style-type: none"> ▪ Undervisnings- og læringsformene må beskrives, og organiseringen av læringsmiljøet må være egnet for å nå målsettingen for studiet. ▪ Opplæringen skal være lagt opp slik at stipendiaten får satt sitt doktorgradsarbeid inn i en bred faglig sammenheng.
Vurderings- og kontrollmetoder	<ul style="list-style-type: none"> ▪ Eksamens- og vurderingsordningen må oppfylle de faglige og formelle kravene som stilles til studier på universitets- og høyskolenivå, og må være relevant i forhold til den målsetting som er gitt for studiet. ▪ Vurderingsordningene må være relevante i forhold til studiets mål og arbeidsformer. ▪ Sensorer må ha samme faglige kompetanse som det som kreves for å undervise på studiet.

Som det fremgår av tabell 3 er det relativt lite i NOKUTs forskrift som er rettet mot kriterier for å vurdere faglig formidling, og ingen ting rettet mot resultatene av formidlingen. Dette utfylles imidlertid i dokumenter rettet mot revisjon av akkreditering. I et av dokumentene som lå til grunn for revidering av sykepleieutdanninger (2005) er eksempelvis følgende kriterier nevnt:

- Studentenes tilfredshet
- Kandidatenes tilfredshet
- Arbeidsgivernes tilfredshet
- Data om gjennomstrømning
- Vurdering av eksamensresultater

Dette er kriterier rettet mot ulike aspekter av faglig formidling: Det skal vurderes i hvilken grad utdanningen svarer til forventningene til henholdsvis studentene, de ferdig utdannede kandidatene og deres arbeidsgivere, samt i hvilken grad, hvor raskt og med hvilket eksamensresultat studentene fullfører utdanningen.³ Når det gjelder å vurdere resultater av faglig formidling favner altså NOKUTs praksis bredt. Som vi kommer tilbake til under, kan de aktuelle resultatene av faglig formidling her vurderes på en rekke ulike måter.

³ Til sammenlikning krever UFD at NOKUT ved revidering av studietilbud skal ”vurdere kvaliteten i undervisningen, for eksempel gjennom studentenes vurderinger av undervisningen, studentenes eksamensresultater og ved ekstern uavhengig evaluering av fagmiljøet” (Rundskriv R-001-03).

Tabell 4 NOKUTs sentrale kriterier relatert til faglig kontekst

Aspekt av faglig kontekst	Eksempler fra NOKUTs forskrift
Rutiner og tiltak for å sikre nivå på studietilbud	<ul style="list-style-type: none"> ▪ Mastergradsstudiet må ha rutiner som sikrer og videreutvikler tilbudets kvalitet. ▪ Det må foreligge rutiner for sikring og utvikling av kvaliteten i doktorgradsstudiet.
Faglig miljø/fagmiljøets attraktivitet	<ul style="list-style-type: none"> ▪ Stabilt fagmiljø knyttet til studiet ▪ Størrelsen på fagmiljøet må være tilpasset den undervisning, veiledning og forskning som må utføres for å videreutvikle studiet og sikre den faglige kvaliteten i studietilbudet.
Faglig infrastruktur	<ul style="list-style-type: none"> ▪ Fagbibliotek som oppfyller visse krav ▪ Egnede undervisningslokaler m.m. ▪ Institusjonen skal være knyttet til nasjonale og internasjonale nettverk i forbindelse med høyere utdanning og forskning og delta i det nasjonale samvirke når det gjelder forskerutdanning. (institusjonsakkrediteringer) ▪ Institusjonens nasjonale og internasjonale samarbeid og nettverk som er relevant for doktorgradsstudiet, må dokumenteres.
Formelt faglig nivå og bredde	<ul style="list-style-type: none"> ▪ Rett til å tildele grad på gitt nivå for x antall studier (institusjonsakkrediteringer)
Kobling til andre faglige aktiviteter	<ul style="list-style-type: none"> ▪ Institusjonen skal også ha forskning og formidling som primærvirksomhet (institusjonsakkrediteringer) ▪ Institusjonen skal ha FoU relatert til sine fagområder/stabil forskningsvirksomhet av høy kvalitet/kunne vise til dokumenterte resultater fra FoU-virksomhet (institusjonsakkrediteringer) ▪ Det bør være samsvar mellom fagmiljøets kompetanse, FoU-virksomhet og fagområdet for studiet ▪ Fagmiljøet som er knyttet til mastergradsstudiet, må drive aktivt FoU-arbeid innenfor studiets fagområde ▪ Institusjonen må ha et aktivt og stabilt forskningsmiljø med god faglig bredde på et internasjonalt nivå innenfor det aktuelle fagområdet.

En rekke av kriteriene i NOKUTs forskrift faller inn under det vi har kalt ”faglig kontekst”. Når det gjelder ”faglig infrastruktur” er det flere materielle krav (fagbibliotek, egnede lokaler m.m.), samt mer generelle faglige krav om å være del av nasjonalt og internasjonalt fagnettverk. Med hensyn til ”kobling til andre faglige aktiviteter” er det stor vekt på FoU-virksomhet. Når det gjelder ”fagmiljø” er det generelle krav om stabilitet og tilstrekkelig størrelse, men ingen kriterier som går på hvor anerkjent eller faglig attraktivt miljøet er. Dette er i tråd med at forskriften gjelder akkreditering og det som er sagt over om at kriterier under faglig kontekst ofte vil være minimumskrav. Et mer fleksibelt alternativ ville være et sett av kriterier der lav score på stabilitet og størrelse kunne oppveies av høy score på anerkjennelse og attraktivitet.

Det britiske Quality Assurance Agency for Higher Education (QAA) har på den annen side fagspesifikke standarder for alle fag/disipliner. De britiske prosedyrene for akkreditering og godkjenning av studier er langt mer gjennomregulerte enn eksempelvis NOKUTs prosedyrer. QAA gir også ut en egen håndbok som tar for seg alle delene i godkjenningen⁴. Standardene er publisert på QAAs nettsider. I motsetning til QAA har NOKUT ett felles sett av kriterier for alle fag. I del 3.3. og i kapittel 4 drøfter vi nærmere muligheten for mer fleksible kriterier og kriterier som gir stort rom for skjønn, versus klarere standarder og mer kvantitative mål som stabilitet og størrelse.

⁴ <http://www.qaa.ac.uk/reviews/subjectReview/handbook/contents.asp>

3.2 Metoder

Det er flere fremgangsmåter som kan være aktuelle når faglig nivå og dokumenterte resultater skal vurderes. I dette delkapitlet gjennomgår vi de mest relevante alternativene (inklusive NOKUTs praksis). I neste delkapittel drøftes sterke og svake sider ved de ulike metodene som kan benyttes i forhold til å vurdere ulike aspekter av faglig nivå.

Fagfellevurdering

Ved fagfellevurdering er evalueringskompetanse definert ut fra fagfellesskap med de evaluerte. Fagfellevurdering brukes når evaluatorene trenger tilsvarende kompetanse som de som skal evalueres (kompetanse på samme fagfelt). Studier av fagfellevurdering av forskning viser at det er betydelig uenighet mellom fagfeller – ulike evaluatorene bedømmer den samme forskningen ulikt. En studie som sammenfatter en rekke studier av fagfellevurderinger for tidsskrifter konkluderte med at samsvaret mellom vurderinger fra ulike fagfeller bare er litt bedre enn et terningkast (Lindsey 1988). I flere studier er det evaluatorens faglige ståsted som fremheves som den sentrale forklaringsfaktor for utfallet av evalueringer (se f.eks. Chubin & Hackett 1990; Travis & Collins 1991; Cole et al. 1981). Det er med andre ord det faglige ståsted til den eller de som velges til å foreta bedømmelsen som forklarer evalueringsutfallet. Fra et mer overordnet perspektiv kan vi se fagfellevurdering som en prosess der ulike vurderinger og ulik ekspertise settes mot hverandre for å få et best mulig grunnlag for en beslutning. Da er ikke uenighet i seg selv et problem, men når ulike eksperter har ulike vurderinger, innebærer dette at både utvelgelsen av evaluatorene og beslutningsprosessene i komiteer som bedømmer forskning, blir kritiske faktorer for utfallet.

Rutiner og prosesser for fagfellevurdering kan altså være mer eller mindre gode med hensyn til å fange opp ulike vurderinger og syn, og produsere en legitim og adekvat konklusjon. Den vanlige arbeidsmåten for fagfellebaserte evalueringsutvalg er å gjennomgå skriftlig dokumentasjon om evalueringsobjektets faglige virksomhet (eksempelvis lese og bedømme vitenskaplige artikler), gjennomføre miljøbesøk og gjennom utvalgsmøter å diskutere seg fram til innholdet og konklusjonene i en felles rapport. Ut fra studier av fagfellevurdering kan vi si at hvordan dette gjøres kan ha stor betydning for konklusjonene i evalueringsrapporten.

Hybridutvalg

Det er også en rekke andre typer ekspertutvalg som delvis bygger på prinsippet om fagfellevurdering, delvis på annen type representasjon: enten fordi oppgaven ikke krever at utvalget har kompetanse som dekker alle fagfeltene som skal evalueres, eller fordi oppgaven krever annen type kompetanse enn den som finnes hos evalueringsobjektet. NOKUTs sakkyndige komiteer faller inn under denne typen ekspertutvalg som delvis bygger på fagfelleprinsippet, delvis på annen type representasjon og kompetanse.

Felles for ren fagfellevurdering og hybridutvalg er at målet er at evaluatorene skal være objektive og upartiske, de skal ikke være interessenter. Samtidig vil de involverte være klar over at sammensetningen av utvalget kan være avgjørende for utfallet og at det kan ligge maktkamper bak oppnevningene. Det *faglige grunnlaget* for vurderingene er i stor grad faglig skjønn og taus kunnskap. Beslutningene tas i grupper med flat struktur og alle medlemmene har i realiteten vetorett – om de ikke er enig med flertallet kan de kreve en minoritetsmerknad. Samtidig er det et press for å nå konsensus, og forhandlinger og kompromisser kan prege møtene. Som regel unngås minoritetsmerknader, bl.a. fordi det ikke ville gi evalueringen like stor tyngde (Langfeldt 2002).

Alternativer til fagfellevurdering

I sammenhenger som ikke krever fagfellekompetanse, kan en også bruke en rekke andre evalueringsmodeller som ikke er basert på ekspertkomiteer, men på utredninger og analyser utført av profesjonelle evaluatore/konsulenter. Slik bistand kan også være et supplement til ekspertutvalg, idet det fremskaffes dokumentasjon og analyser som underlag for ekspertutvalgets arbeid (relevant statistikk, spørreundersøkelser etc.).

NOKUTs praksis ved re-akkreditering av studietilbud

NOKUT benytter sakkyndige komiteer ved alle akkrediteringer. Disse faller, som nevnt over, i kategorien hybridutvalg. Ved re-akkreditering av studietilbud består utvalgene av 2-4 sakkyndige med kompetanse innen det fagområdet som skal evalueres, evt. et annet relevant fagområde, samt en student. Minst ett medlem skal komme fra utlandet (ikke påkrevd ved studietilbud på lavere grads nivå). Av habilitetshensyn oppnevnes ingen tilknyttet det aktuelle studietilbudet eller institusjonen som skal evalueres. Krav til formalkompetanse er avhenging av hvilket nivå som vurderes.⁵ Sekretærfunksjonen ivaretas av en prosjektgruppe i NOKUTs sekretariat, som også sikrer fremdrift og koordinering av evalueringen. Utvalget får bakgrunnsinformasjon i form av en egnevaluering fra institusjonen, samt annen aktuelle dokumentasjon, inklusive spørreundersøkelser og/eller intervjuer av studenter, kandidater og arbeidsgivere. Utvalget gjennomgår bakgrunnsmateriale, foretar institusjonsbesøk og diskuterer seg fram til en felles vurdering som nedfelles i en skriftlig rapport som behandles i NOKUTs styre.

NOKUTs ”hybridutvalg” har altså hovedvekt på fagfeller (vitenskaplig ansatte ved tilsvarende utdanninger), samt et mindre innslag av brukere av utdanningen (én student per utvalg). Også i dokumentene som danner rammene for vurderingsarbeidet har NOKUT valgt en kombinasjonsløsning – en kombinasjon av vekt på faglig/skjønnsbasert vurdering og på bruk av mer entydige/ikke skjønnsbaserte indikatorer (ref. del 3.1).

⁵ Professorkompetanse ved evaluering av doktorgradsutdanning, førstestillingskompetanse ved evaluering av mastergradsutdanning, høyskolelektor eller høyere ved evaluering av lavere grads studietilbud.

3.3 Metoder og kriterier: Fordeler, ulemper og dilemmaer

I dette delkapitlet drøfter vi en del av de mest aktuelle metoder og kriterier for å vurdere faglig nivå. En kort oppsummering av styrker og svakheter og mulige løsninger finnes i tabell 5.

Pensums faglige nivå

Hva som skal være pensum på et fag kan det være svært ulike meninger om, spesielt for utdanninger uten rammeplaner. Om det foreligger en rammeplan for faget bør en gruppe uavhengige fagfeller forholdsvis enkelt kunne avgjøre om pensum er i tråd med denne. Om det ikke foreligger rammeplan kan det likevel ha utviklet seg en mer eller mindre eksplisitt konsensus for hva som skal inngå i faget på ulike nivå. Fagfellevurdering kan være et middel for å utvikle en slik konsensus dersom den ikke allerede eksisterer. Som nevnt kan fagfellers vurderinger være farget av deres faglige ståsted, og i den grad det er ulike meninger om hva som er viktig pensumstoff og hva som er krav til faglig omfang og dybde i en gitt type utdanning, vil sammensetningen av evaluatorgruppen være avgjørende for utfallet av evalueringen. Slik usikkerhet kan reduseres på ulike måter – ved å gi klarere føringer for vurderingene (eksempelvis gi et klart referansenivå for pensum) eller ved å oppnevne flere evaluatore (et større evalueringsutvalg). Begge løsninger har klare ulemper. Klarere føringer for vurderingene krever at oppdragsgiver tar stilling til spørsmål det er ulike meninger om i fagmiljøet, noe som kan innebære at evalueringens konklusjoner oppfattes som kontroversielle og mindre legitime. Store evalueringsutvalg er mer kostbare, og kan også være mer tidkrevende og man har heller ingen garanti for at utvalget klarer å komme fram til enighet. Et alternativ til store evalueringsutvalg er å bruke flere individuelle eksperter som bidrar med hver sine skriftlige uttalelser om pensums nivå – uttalelser som deretter ligger til grunn for samlede vurderinger foretatt av et mindre evalueringsutvalg.

En tredje mulighet er å gi mer utfyllende, men åpne, rettleddninger i stedet for klare føringer og et sett av obligatoriske kriterier⁶. Når det gjelder å vurdere pensums faglige nivå kan ulike indikatorer på at utdanningens innhold er i tråd med et etablert syn på adekvat faglig nivå, inngå i en slik rettleddning. Eksempelvis kan det at et lavere grads studium blir/ikke blir godkjent for inntak til høyere grads studier ved andre læresteder, inngå i en rettleddning, men kan vanskelig ses som et absolutt kriterium – at en utdanning ikke godkjennes for videre påbygging et annet sted er ikke nødvendigvis et uttrykk for ulikt faglig nivå.

⁶ Med ”rettledning” menes da et sett av kriterier som kan legges til grunn for diskusjon i utvalget, og utvalget kan så begrunne sine valg og sin kriteriebruk i evalueringsrapporten. Med ”obligatoriske kriterier” har evalueringsutvalget langt mindre spillerom, og slik også mindre behov for å begrunne valg.

Dokumentasjon av faglig oppdatering

Det kan være vesentlige forskjeller mellom fag mht hvilke typer kanaler som er aktuelle for faglig oppdatering og også forskjeller mht hvor adekvat ulike former for FoU-virksomhet er for å måle lærernes faglige oppdatering. Dertil er det også store forskjeller mht hvordan FoU virksomhet kan måles, bl.a. pga ulike publiseringstradisjoner i ulike kanaler. Det kan derfor være vanskelig å lage noen felles kvantitative indikatorer for å dokumentere faglig oppdatering. Om målingen i stedet konsentreres om *tid* avsatt til faglig oppdatering (inkl. FoU-arbeid) blir det enklere å sette opp mer generelle krav. Om de kvantitative målambisjonene reduseres til i hvilken grad lærerne har en viss andel av sin tid avsatt til faglig oppdatering får vi et enkelt mål som burde være ukontroversielt (minstekrav til tid som må oppfylles).

Faglig formidling

Flere ulike forhold er aktuelle ved vurdering av faglig formidling. Her skal vi se spesielt på lærernes formalkompetanse, studentevalueringer og eksamensresultater.

Lærere og sensorenes formalkompetanse: Med formalkompetanse refereres det her til stillingsnivå. For å bli førsteamanuensis kreves det at man oppfyller gitte vitenskapelige kriterier (arbeid tilsvarende doktorgrad), eller tilsvarende kunstneriske meritter, samt dokumentert relevant praktisk-pedagogisk kompetanse. For førstelektorstilling er kvalifikasjonene samlet sett på nivå med førsteamanuensis, men grunnlaget for kvalifikasjonene er bredere (KUF 1995⁷). For professorkompetanse kreves arbeid tilsvarende to doktorgrader (samt klarere krav til at arbeidet holder høy internasjonal standard).

Kriterier som går på oppnådd formalkompetanse er klare og entydige og det kan enkelt avgjøres om kravet er oppfylt eller ikke, både av institusjonene selv og av NOKUT uten noen form for fagfellevurdering – fordi fagfeller allerede har gjort de vurderingene som ligger bak tildelingen av formalkompetansen⁸. Det kan likevel stilles spørsmålsteget ved hvor påkrevd formalkompetanse er for god faglig formidling innen ulike fag. Slik kan entydige og absolutte krav om faglig kompetanse (formalkompetanse) også ses som et problem. Med vurderingsmetoder basert på fagfellers skjønn (ren fagfellevurdering eller hybridutvalg) vil en kunne få mer direkte vurderinger av lærernes faglige innsikt og pedagogisk evner, og dermed vurderinger av det faglige nivået på undervisningen som er mer tilpasset den enkelte utdanningen som vurderes. Ulempen med dette alternativet – foruten at fagfellevurdering er kostnadskrevende – er lavere reliabilitet. Ulike grupper av fagfeller kan ofte komme frem til ulike konklusjoner.

⁷ Ny forskrift om kriterier for ansettelse og opprykk er for tiden ute på høring.

⁸ Formalkompetanse oppnås ved vurdering i tilknytning til tilsetning (konkurransen om stilling) eller ved vurdering i tilknytning til søknad om opprykk. Opprykksordningen er drøftet i Kyvik et al. 2003.

Eksamensresultater er den mest åpenbare måten å dokumentere faglig formidling på. Ferdig ”kodete” data foreligger i form av sensorenes bedømmelser av studentenes faglige nivå. Spørsmålet er i hvilken grad eksamensresultater er en god indikator på hva som er formidlet og hva studentene har lært, og i hvilken grad dataene er sammenliknbare. Ulike fag og ulike utdanninger kan ha ulik vanskelighetsgrad/nivå, men fortsatt bruke samme skala, uten at eksamensdataene gir noen informasjon om slike ulikheter. Et annet viktig moment er at eksamensresultater henger sammen med det faglige nivået på studentene som tas inn på studiet. Eksamen måler hva studentene *kan*, ikke bare hva de har lært innen rammen av studiene. Karakterer måler med andre ord både innsats og kvalitet på studentenes prestasjon, samtidig som det også måler evner og talent (Arnesen og Try 2001). Videre er det sammenheng mellom det å ha gode karakterer fra videregående opplæring og det å ha ambisjon om å gjøre det godt til eksamen, det vil si at studenter som har gode karakterer fra videregående opplæring vil ha ambisjon om å også gjøre det godt i høyere utdanning (Hovdhaugen 2004). Forskning viser også at karakternivået fra utdanning på lavere nivåer er en god indikator for hvilket karakternivå studenten vil oppnå i høyere utdanning (se for eksempel Grøgaard 1995 og Helland 2004). Kvaliteten på studentene som rekrutteres til utdanningen vil altså gjenspeile seg i utdanningens eksamensresultater.

Når karakter henger sammen med inntakskvalitet, vil utdanninger kunne få dårlige eksamensresultater til tross for at de har gode lærerkrefter og høyt nivå på den faglige formidlingen.⁹ Omvendt vil også utdanninger med lavt nivå på den faglige formidlingen kunne oppnå gode eksamensresultater, uavhengig av kvaliteten på lærerkreftene.

Ofte vil det imidlertid kunne være en selvforsterkende effekt: Sterke lærekrefter og høyt nivå på formidlingen tiltrekker seg studenter på et høyt faglig nivå og gir gode eksamensresultater, og også muligheten for å ha strengere opptakskrav på studiet. I hvilken grad inntakskvalitet bør være en del av vurderingen av det faglige nivået på en utdanning vil det dermed kunne være delte meninger om. Dette kommer vi tilbake til under ”attraktivitet” (under).

Summen av dette viser at eksamensresultater er svært komplekst som kriterium på faglig formidling. Strykprosent på studiet er et annet mål på eksamensresultater som kan være problematisk å tolke. Høy strykprosent kan eksempelvis ses som en indikator på høyt faglig nivå hos de studentene som har bestått, og dermed god kvalitet på utdanningen – gitt at årsaken til høy strykprosent er at det praktiseres strengere sensur eller at eksamensoppgavene ligger på et høyere vanskelighetsnivå enn på tilsvarende utdanninger. På den annen side kan høy strykprosent komme av lavt faglig nivå på undervisningen, av utilstrekkelig pedagogisk kompetanse hos lærerne/dårlig tilpasset undervisning, eller av lavt faglig nivå på studentene da de begynte på studiet. Tilsvarende vil lav strykprosent

kunne være en indikasjon på høyt faglig nivå og god undervisning (Hovdhaugen 2005). Tid brukt på studier er en annen faktor som kan påvirke strykprosenten, og denne kan det også være vanskelig å måle effekten av.¹⁰ Når eksamensresultater skal benyttes som dokumentasjon på faglig nivå, trenger man derfor å se disse i sammen med andre data: Data om inntakskvalitet, sensurordning (antall sensorer og bruk av eksterne sensorer), lærekreftenes kompetanse og gjerne også studentevalueringer.

Også når det gjelder *studieprogresjon og fullføringsandel* spiller mange ulike forhold inn, og verken studieprogresjon eller fullføringsandel trenger å være relatert til kvaliteten på den faglige formidlingen. Ulike typer studier har ulik fullføringsandel og studieprogresjon. I profesjonsstudier, som sykepleieutdanning og førskoleutdanning er andelen som fullfører på normert tid forholdsvis god, mens den er lavere i mer løst strukturerte studier (Aamodt 2001). Heller ikke når fullføringsandel eller progresjon avviker fra "normalen" for sitt studium, er tolkningen gitt. Synkende fullføringsandel kan eksempelvis være relatert til dårligere inntakskvalitet eller være en effekt av dårligere faglig kvalitet på utdanningen.

Studentevalueringer er en alternativ indikator på lærernes faglige innsikt og pedagogiske evner. Ulempen med studentevalueringer er at det i motsetning til eksamensresultater krever ny datainnsamling, og slik kan bli både tidkrevende og kostbart. Slik datainnsamling krever også høy kompetanse på utforming av spørreundersøkelser og god svarprosent for å gi valide og reliable data. På den annen side kan en slik undersøkelse fange inn mange ulike aspekter ved faglig nivå (Wiers-Jenssen et al. 2002). I tillegg til vurderinger av den faglige formidlingen, kan det være interessant å få studentenes vurderinger av fagmiljøet og læringsmiljøet, samt informasjon om bakgrunnsfaktorer som det vil være aktuelt å se resultatene i forhold til (sosioøkonomisk status, motivasjon, tid som brukes på studiet, progresjon, etc.). Også her vil resultatene selvsagt kunne avhenge av inntakskvalitet. For eksempel vil svake og sterke studenter kunne ha helt ulike oppfatninger om undervisningen, om det er progresjon i studiet og om lærernes evne til å formidle pensum og deres vilje og evne til å gi faglig veiledning.

Et annet forhold er at mange læresteder selv gjennomfører studentevalueringer. De oppfattes som nødvendige, men gir ikke automatisk økt studiekvalitet. Studentevalueringer benyttes i liten grad til å forbedre undervisningen, dels på grunn av at ansvaret for kvaliteten flyttes fra lærestedsnivå til individnivå (foreleser/lærere), og dels at evalueringen bare forholder seg til undervisningen slik den ble gjennomført uten å ta hensyn til de premisser som førte til akkurat den undervisningsformen. "Studentevalueringer av

⁹ En ytterligere kompliserende faktor her er at det vil kunne være uenighet om en optimal tilpassning av undervisningen til studentmassen kunne gitt et bedre resultat, og i hvilken grad en god faglig formidling også skal innebære tilpasning til endinger i studentenes forkunnskaper.

¹⁰ Det er store variasjoner mellom ulike studieprogrammer i hvor mye tid som brukes på studier (Wiers-Jenssen og Aamodt 2002). Berg (1995) viste imidlertid at de studentene som bruker mest tid på studiene sine gjør det best til eksamen.

undervisning må derfor inngå i en større prosess hvor lærestedet tar et overordnet ansvar” (Stensaker 1998: 19).

Faglig attraktivitet

Det er mange ulike kvantitative mål som kan gi en indikator på en utdannings faglige attraktivitet: Søkertall (i forhold til konkurrerende studier), hvor attraktive kandidatene er (etterspørsel, karriereprogresjon, lønnsnivå), hvor attraktivt fagmiljøet er som arbeidssted (eksempelvis søkere per utlyst vitenskaplig stilling i forhold til konkurrerende studier). Problemet er at attraktivitet ikke trenger å være direkte relatert til faglig nivå. Miljøets historie, tradisjoner og geografiske plassering spiller også inn. Dertil kommer en selvforsterkende effekt (Matteus-effekt, ref. Merton 1968 og 1988). De utdanninger som har oppnådd ry for å være attraktive har langt lettere for å tiltrekke seg sterke studenter og gode lærekrefter, enn andre læresteder, og blir dermed mer attraktive.

Ser vi spesielt på eksempelvis søkertall som en indikator på faglig nivå, ser vi også at ulike mål gir ulik informasjon. Antall søkere per plass eller poengkrav for å komme inn på utdanningen er begge aggregerte mål på søkertilstrømning til studiet. Antall søkere per plass er et problematisk mål på inntakskvalitet fordi antallet vil endre seg dersom det bli flere eller færre plasser på utdanningen, uten at søkertilstrømningen øker/minsker tilsvarende (Hovdhaugen 2005). Tilsvarende viser utdanningens poengkrav bare hva den siste personen som kom inn på utdanningen hadde i gjennomsnittlig poengsum, ikke hva som er gjennomsnittlig karakternivå til de studentene som går på utdanningen. Det kreves generelt gode karakterer for å komme inn på utdanninger der det er stor konkurranse om plassene. Søkere per plass og poengkrav på utdanningen sier dermed noe om konkurransen om plassene på ulike typer utdanninger og dermed om studiets attraktivitet, men lite om inntakskvaliteten på studentene på utdanningen. Gjennomsnittlig karakternivå på de som begynner i utdanningen kan på den annen side gi et bedre bilde av inntakskvaliteten.

I flere sammenhenger produseres det ulike rangeringer av høyere utdanningsinstitusjoner. Studier viser at disse reflekterer renommé og prestisje, og i liten grad sier noe om det læringsmiljø studentene møter på lærestedet (Ewell 2004, Saraff et al. 2005). Kuh og Pascarella (2004) konkluderte med at resultatene av U.S. News and World Report sin årlige ranking av ”America’s best colleges” i stor grad kan forutsis ved å se på det gjennomsnittlige SAT-score for lærestedets studenter, det vil si at rankingen egentlig bare viser inntakskvaliteten på studentene. Men selektivitet er i følge amerikanske studier ikke relatert til student-engasjement eller ønskede læringsmål (Pike 2003). I tillegg er institusjonell selektivitet en svak indikator på om studentene faktisk kommer til å få god undervisning og et godt læringsmiljø under studiene. Det er andre dimensjoner ved lavere grads utdanning som er viktigere for at studentene skal lykkes enn utdanningens eller lærestedets renommé og selektivitet (Pascarella 2001).

På den annen side kan faglig renommé og prestisje være viktig å inkludere i en mer helhetlig vurdering av faglig nivå. Om man kommer fram til resultater som strider sterkt mot utdanningenes score på attraktivitet, vil det være behov for gå dypere i materialet og begrunne det meget godt om man for eksempel finner at en utdanning som anses å være svært faglig attraktiv ikke fyller kravene til godkjenning, mens andre, mindre attraktive utdanninger, holder et høyere faglig nivå og kan godkjennes.

Fagmiljø

Forholdstallet mellom studenter og lærere er en av mange mulig indikatorer på hvor godt et fagmiljø er. Imidlertid er dette forholdstallet avhengig både av hvilken type undervisning som bedrives ved lærestedet og av fag og nivå. Noen fag er mer lærerintensive enn andre. En foreleser kan undervise mange hundre studenter, mens for lærere som leder en laboratorieøvelse er et mer begrenset antall studenter det vanlige, blant annet på grunn av ressurs hensyn (lokaler, sikkerhet). Dermed sier forholdstallet studenter – lærere mer om hvilken type undervisning som bedrives på studiet enn kvaliteten på undervisningen.

FoU-virksomhet

Det er mange muligheter til å bruke ulike kvantitative mål på FoU-virksomhet til å dokumentere FoU-resultater, eksempelvis hvor hyppig miljøet publiserer FoU-resultater i ulike kanaler, om de er involvert i internasjonalt FoU-samarbeid og hvor hyppig publikasjonene blir sitert av andre forskere. Kvantitative mål kan imidlertid være misvisende, bl.a. fordi det er store forskjeller mellom fag (også mellom underdisipliner) mht hva som kan måles og hvordan. Dertil kommer at sammenhengen mellom kvantitet og kvalitet i forskning er usikker, og det er derfor betydelig usikkerhet knyttet til å bruke bibliometriske data som grunnlag for styring av eller i UoH-sektoren. Ikke minst er faren for å presse frem et endret kommunikasjonsmønster i forskningsverdenen stor – press på gitte typer formidling som kanskje ikke passer like godt for alle fag eller alle deler av et fag (Hansen og Jørgensen 1995).

Det er heller ingen klar kobling mellom nivået på FoU-virksomheten og nivået på undervisningen. Om det blir for stort press på FoU-siden kan det tvert i mot gå ut over deler av undervisningen. Eksempelvis krever internasjonal publisering som regel meget høy grad av spesialisering innen emner som er aktuelle for et internasjonalt publikum og lærernes tid til breddeorientering og oppdatering på ”lokale” tema kan reduseres. Om kriteriet er at produksjonen er over en viss minstandard og målene er tilpasset publiseringskanaler og publiseringspraksis i det enkelte fag og dataene tolkes av fagfeller, kan publiseringsvirksomhet imidlertid fungere som dokumentasjon på et ”aktivt og stabilt forskningsmiljø”.

Tabell 5 Dokumentasjon av faglig nivå: Alternativer, fordeler og ulemper

Dokumentasjon av:	Data-/vurderingsmetode	Påkrevd ekspertise	Fordeler	Ulemper/svakheter	Mulige løsninger
Pensums faglige nivå	Om pensum er i tråd med rammeplan for faget	Fagfeller	Relatert til fagspesifikke krav	Krever at det foreligger en rammeplan	Flere eksperter, mer utfyllende og samtidig mer fleksible kriterier. Evt. individuelle ekspertuttalelser
	Fagfellevurdering; sammenlikning av pensum	Mange fagfeller	Tar hensyn til ulike faglige ståsted; kan bygge opp felles vurderingsramme	Kostbart; tidkrevende, ingen garanti for enighet, kan være problemer med habilitet	
	Sjekk om studiet er godkjent for påbygning til andre (akkrediterte) utdanninger.	Ingen	Lav kostnad	Usikker indikator; Ingen klar sammenheng mellom indikator og faglig nivå	
Faglig oppdatering	Kvantitative indikatorer på læreres/sensorers faglige oppdatering	Fagfeller (for å lage kriterier)	Grad av faglig oppdatering vil være relatert nivået på undervisning og sensur.	Krever innsamling av nye data; Store forskjeller mellom fag	Lavere kvantitative målingsambisjoner: Eks. minstekrav til tid avsatt til faglig oppdatering.
Formal-kompetanse	Se på hvor stor andel av undervisningspersonalet som har førstestillingskompetanse, etc.	Ingen	Lav kostnad	Uklar sammenheng med nivå på faglig formidling; Fagforskjeller	Bruke fagekspertise og ha mindre entydige krav/også vurdere realkompetanse
Eksamens-resultater	Gjennomsnittskarakter; strykporsent etc.	Ingen/Analysekompetanse	Lav kostnad	Kompleks indikator/vanskelig å tolke. Påvirkes bl.a. av inntakskvalitet	Kombinere med, og tolke mot, andre data
Student-evalueringer	Spørreundersøkelser til studenter og kandidater	Utforming og analyse av surveyer	Kan fange inn flere aspekter ved faglig nivå.	Kostnader; Påvirkes av inntakskvalitet	Kombinere med, og tolke mot, andre data
Studieprogresjon og fullføringsandel	Kvantitative mål	Ingen	Lav kostnad	Kompleks indikator/vanskelig å tolke. Påvirkes bl.a. av inntakskvalitet	Kombinere med, og tolke mot, andre data
Attraktivitet	Ulike kvantitative mål	Ingen/Utforming og analyse av surveyer	Kan gi indikasjon på generell oppfatning av faglig nivå	Uklar sammenheng med faglig nivå; Matteuseffekt	Kombinere med, og tolke mot, andre data
Fagmiljø/størrelse	Antall fast ansatte; Forholdstall lærer student etc.	Ingen	Lav kostnad	Uklar sammenheng med faglig nivå; Forholdstall sier mer om type utdanning enn om faglig nivå.	Bruke fagekspertise til å vurdere hver enkelt utdanning
FoU-virksomhet	Tallfesting av FoU-resultater	Bibliometri	Tilgjengelige data; Kan dokumentere aktivt og stabilt FoU-miljø	Store forskjeller mellom fag mht hva som kan måles og hvordan. Ingen nødvendig kobling med FoU-virksomhet og faglig nivå på undervisningen.	Fagspesifikke tolkninger; Minstekrav relatert til gitte fag

3.4 Usikkerhet

Tabell 5 oppsummerer gjennomgangen av ulike muligheter for å dokumentere faglig nivå. Det er ulike styrker og svakheter knyttet til ulike kriterie- og metodevalg. Det er eksempelvis en rekke ulike kvantitative indikatorer som kan benyttes til lav kostnad. Disse har imidlertid uklare sammenheng med faglig nivå og trenger å kombineres med, og ses i sammenheng med, andre data. Andre forhold krever fagfellekompetanse for å kunne vurderes og blir slik mer kostbare, særlig om man skal sikre god reliabilitet ved å benytte et stort utvalg av fagfeller.

Vår drøfting av de enkelte aspektene er nødvendigvis ufullstendig. En full drøfting av styrker og svakheter ved enkeltkriterier ville krevd helt andre rammer. Hovedpoenget i vår sammenheng er imidlertid at det er vanskelig å måle faglig nivå og at ethvert kriterium er usikkert. Løsningen blir slik å kombinere ulike kriterier og tolke dem i lys av hverandre.

Men det er ikke gitt på *hvilken måte* ulike typer kriterier best kan kombineres for å minimere usikkerheten. Noen av kriteriene vi har gjennomgått er eksempelvis relatert til resultater av utdanningen, mens andre er innsatsfaktorer i utdanningen: Eksamensresultater, studieprogresjon/-fullføring, godkjenning ved inntak til andre studier, studentevalueringer, attraktivitet (målt ved søkertall, etterspørsel, karriereprogresjon, lønnsnivå) og FoU-virksomhet/publisering er ulike resultatmål. Pensum, formalkompetanse, faglig oppdatering og fagmiljø er på den annen side innsatsfaktorer. Begge typer kriterier er mål på faglig nivå. Men hvordan tolker vi det hvis et miljø scorer lavt på formelle innsatsfaktorer (eksempelvis lite fagmiljø og lav formal-kompetanse), men høyt på dokumenterte resultater? Her vil det kunne være ulike meninger om hvordan ulike kriteriene skal veies mot hverandre og om en utdanning holder tilstrekkelig faglig nivå eller ikke.

I neste kapittel drøfter vi mulige tilnæringsmåter for å håndtere denne usikkerheten.

4 Overordnede spørsmål og valgmuligheter

I dette kapitlet drøftes de ulike hensyn og formål som et nasjonalt system for akkreditering av høyere utdanning skal ivareta. Alternative modeller som balanserer de ulike hensynene diskuteres (4.2) og det redegjøres spesielt for ulike måter å fremme et høyt faglig nivå på vurderingene (4.3).

4.1 Kvalitetssikring i høyere utdanning: en avveining av mange hensyn

Gjennom opprettelsen av et nasjonalt system for kvalitetssikring i høyere utdanning, og etableringen av NOKUT som det ansvarlige organ for å gjennomføre denne kvalitetssikringen, kan man være fristet til å hevde at man etter mange år med eksperimentering og forsøk endelig har satt fokus på kvalitet i norsk høyere utdanning. Samtidig kan man dermed overse at etableringen av et nasjonalt system for kvalitetssikring i høyere utdanning også har andre formål enn ”bare” å sikre kvaliteten. Stensaker (2004) har tidligere påpekt at det norske systemet for kvalitetssikring på mange måter er en avveining mellom ulike hensyn, der stikkord som rettsikkerhet, uavhengighet, økonomi, og internasjonalisering også er sentrale.

Uavhengigheten ble på mange måter fremhevet som et viktig behov ved opprettelsen av NOKUT, der etableringen av organet skulle tydeliggjøre et sterkere skille mellom fag og politikk i høyere utdanning. I det gamle systemet var det departementet som var saksbehandler når nye studier skulle godkjennes. Begrunnelsene for avgjørelser fattet av departementet kunne imidlertid mistenkeliggjøres i dette systemet, der argumentene ofte gikk på at det var politiske heller enn faglige hensyn som bestemte avslag og godkjenninger.

En tilgrensende verdi ved opprettelsen av NOKUT var i tillegg å etablere større grad av *rettsikkerhet* og likebehandling. Fordi man i det gamle systemet i stor grad baserte seg på peer review-mekanismer, og i mindre grad skriftlige kriterier (profesjonsutdanninger unntatt), er det mulig å hevde at forutsigbarheten i systemet var dårlig, og at rettsikkerheten til de som søkte om godkjenning for nye studier dermed var forholdsvis svak. Hvis ikke kriteriene er tydelige på forhånd, kan det også være vanskeligere å klage i etterkant. Poenget om rettsikkerhet kan ikke minst sies å gjelde for private studieprogram i høyere utdanning, som var underlagt egen lov, og som dermed også hadde en annen søknadsprosedyre enn offentlige utdanninger.

Når NOKUT ble etablert var en betingelse at dette organet skulle være lite og ubyråkratisk, og ikke være et kostnadsdrivende element i høyere utdanning. *Økonomi* er dermed også en faktor som har hatt betydning for utviklingen av NOKUT og dets kvalitetssikringsrutiner.

Gjennom å etablere tydelige regler og enkle kriterier kan man dermed hevde at man sparer både utdanningsinstitusjonene og NOKUT for ressurser i søknadsprosessen.

En siste faktor som kan nevnes er Norges involvering i Bologna-prosessen, og de krav som der ligger knyttet til kvalitetssikring av studietilbud for studenter som beveger seg over landegrensene. Uten at det skal påstås at det eksisterer internasjonale standarder når det gjelder kriterier for godkjenning av studieprogram, er det likevel grunn til å hevde at internasjonale kontakter ofte kan bidra til større grad av sammenlignbarhet og kopiering. Den økte populariteten til akkreditering som sådan, og den økte utbredelsen av såkalte "student satisfaction surveys" i høyere utdanning er bare noen eksempler på effektene av den økende internasjonaliseringen. *Internasjonaliseringen* kan dermed hevdes å bidra til fokus på bestemte kriterier og standarder når et studieprogram skal akkrediteres.

Oppsummert kan man derfor hevde at akkrediteringssystemet i Norge er ment å ivareta mange hensyn, og at "faglighet" ikke er enerådende som begrunnelse for det eksisterende systemet. At man har et kriterium som betoner den regionale relevansen for studieprogram på PhD-nivå, er f.eks. et tydelig tegn på at den politiske dimensjonen fremdeles kan gjøre seg gjeldende i Norge. Regional relevans kan vanskelig hevdes å være et kriterium som er definert ut fra faglige hensyn. At NOKUT har etablert skriftlige kriterier som gjelder uansett om søker er offentlig eller privat, tilsier også at forhold som uavhengighet og rettsikkerhet er tilgodesett i det eksisterende systemet.

I dagens akkrediteringssystem skal alle de ulike hensynene ivaretas i en og samme prosedyre og av et felles sett av kriterier. Dette tilsier at en viktig utfordring fremover vil være å balansere de ulike hensynene på en måte som sikrer legitimiteten til prosedyrene hos de involverte parter. Vår gjennomgang av de eksisterende kriterier og prosedyrer som eksisterer på akkrediteringsfeltet viser at dagens kriterier og prosedyrer, som i prinsippet skal være faglig begrunnet, ikke er så entydige som man skulle anta. Samtidig er det viktig å understreke at det finnes mange mulige svar på den utfordring det er å balansere faglighet, rettsikkerhet, uavhengighet, økonomi og internasjonalisering. Under gis det noen eksempler på det mulighetsrom som eksisterer.

4.2 Å balansere ulike hensyn i faglige vurderinger

Konklusjonen på foregående kapittel er at det er en iboende usikkerhet i enhver faglig vurdering. Generelt er det to hovedveier for å redusere usikkerhet i faglige bedømmelser: å involvere flere eksperter (altså bruke et bredere sett av fagekspertise) eller å basere vurderingene på klarere og mindre skjønnsbaserte kriterier og standarder. Det første alternativet fremmer faglighet, det andre fremmer likebehandling, rettsikkerhet og økonomi.

Eksplisitte og entydige kriterier reduserer usikkerheten gjennom økt reliabilitet og forutsigbarhet. Jo enklere og jo mer operasjonelle kriterier (som eksempelvis kvantitative indikatorer), jo mindre behov er det for fagekspertise til å foreta vurderingene, og jo enklere for de evaluerte å forutsi resultatet av vurderingen. Samtidig blir det vanskeligere å sikre *valide* vurderinger – vurderinger som samsvarer med hva fagmiljøene forstår med faglig nivå – og vanskeligere å sikre at vurderingene er tilpasset ulike typer studier og fag.

Alternativet er å bruke kriterier som krever fagekspertise og gir rom for skjønn. Slike kriterier vil kunne tolkes i forhold til ulike kontekster og fag, og vil slik kunne gi høyere validitet. I slike vurderinger reduseres usikkerheten, og reliabiliteten økes, ved å involvere flere fagekspertiser og bredere fagekspertise i vurderingene. Om en kun benytter et lite antall eksperter er det forholdsvis høy sannsynlighet for at en annen ekspertgruppe ville komme fram til en annen konklusjon (lav reliabilitet).

Under drøfter vi ulike alternativer langs begge disse hovedveiene – mer standardisering eller mer vekt på fagfellevurdering – og måter å kombinere dem på. Spørsmålet som skal besvares er hvordan man kan håndtere usikkerhet i vurderinger av faglig nivå og dokumenterte resultater, og balanserer faglighet, likebehandling, uavhengighet, økonomi og internasjonalisering i en akkrediteringsprosess.

Alternativer for å redusere usikkerhet og balansere ulike hensyn

NOKUT har ett kriteriesett som gjelder alle fag og benytter fagekspertise i vurderingene. Noen av kriteriene er entydige/ikke skjønnsbaserte og kan avgjøres uten fagekspertise (eksempelvis vurderingskriterier ved tilsetning og opprykk; andel med gitt type formalkompetanse), mens andre krever fagfellekompetanse (eksempelvis nivået på studiets innhold og om forskningsmiljøet er på internasjonalt nivå). NOKUTs evalueringsutvalg har forholdsvis begrenset fagfellekompetanse (2-4 vitenskapelig ansatte ved tilsvarende utdanninger, samt én student per utvalg). Under diskuterer vi ulike alternativer til dette og hvordan disse kan bidra til mer valide og reliable vurderinger.

To-trinns-prosess

En alternativ fremgangsmåte er en to-trinns-prosess som skiller skjønnsbaserte vurderinger og vurderinger som ikke er skjønnsbaserte. I en slik prosess avgjør en først (trinn én) om utdanningene fyller de formelle og entydige kravene for akkreditering. De utdanningene som fyller de formelle kravene, går så videre til trinn to for en vurdering av om de også fyller de mer skjønnsbaserte kravene (fagfellevurdering). Ved en slik to-trinns-prosess kan en unngå kostnadskrevenne fagfellebaserte vurderinger i saker der konklusjonen kan avgjøres av klare entydige kriterier. Av større betydning i vår sammenheng er imidlertid at en slik to-delning kan redusere usikkerhet i vurderingene. De ikke-skjønnsbaserte kriteriene tydeliggjøres og deres utvetydighet fremheves. Dertil kan fagfellekomponenten rendyrkes og designes bedre for å sikre reliabilitet – gjennom å legge mer vekt på fagkompetanse ved sammensetningen av ekspertisen som benyttes i trinn to. Med andre ord sikres

rettsikkerhet, likebehandling og økonomi ved å skille ut entydige krav i trinn én, mens faglighet sikres i trinn to.

Spørsmålet er om en slik prosess ville oppfattes som like legitim som dagens fremgangsmåte. Svaret henger sammen med hvilke funksjoner man tillegger fagfellekomponenten i vurderingene. Om grunnen til at man bruker fagfeller til å trekke konklusjoner basert på entydige kriterier som ikke gir rom for skjønn, er at kriteriene er kontroversielle og man derfor trenger fagfeller for å sikre legitimitet til konklusjonene, ville en slik løsning kunne skape tvil rundt hele grunnlaget for vurderingsmodellen. Hvis dette er tilfelle ville man måtte sikre legitimiteten til de entydige kriteriene før man skiller dem ut til et eget trinn i vurderingsprosessen. En sikring av legitimiteten kan eksempelvis innebære en gjennomgang med ulike fagmiljøer for å sikre at kriteriene passer deres fag og ikke er kontroversielle.

Mer fleksible kriterier

Et alternativ til å skille ut ikke-skjønnsbaserte vurderinger er å åpne for skjønn også i vurderingene av det som i dag fremstår som utvetydige krav. Eksempelvis kan krav om formalkompetanse formuleres som at en "tilstrekkelig" andel av undervisningspersonalet må ha førstestillingskompetanse, i stedet for at andelen tallfestes i forskriftene.

Mer fleksible kriterier kan imidlertid være uønsket i akkrediteringssammenheng. Ut fra hensyn som likebehandling og rettsikkerhet er det ønskelig med klare standarder når spørsmålet dreier seg om et gitt studium oppfyller minstekrav. Med mer "fleksible" kriterier reduseres reliabiliteten. Gevinsten er økt validitet gjennom kriterier som bedre kan tilpasses de enkelte fag og studium, og rom for vurderinger som bedre samsvarer med hvordan fagmiljøene forstår faglig nivå. Dette alternativet innebærer med andre ord økt faglighet på bekostning av redusert likebehandling, rettsikkerhet og reliabilitet.

Fagspesifikke kriterier/standarder

Et alternativet til et mer fleksibelt sett av kriterier er å ha ulike kriterier for ulike vurderinger. Det britiske *Quality Assurance Agency for Higher Education* (QAA) har eksempelvis et omfattende sett av fagspesifikke standarder. For hvert som fag blir vurdert er det på forhånd definert standarder som de som gjennomfører vurderingene skal forholde seg til. I mange fag er det eksempelvis presentert lange lister over hvilke kunnskaper studentene skal komme ut med etter et bachelor- eller masterstudium i faget. For eksempel sies det i forskriften for kjemifaget at hver institusjon selv kan utforme sin grad, samtidig som det listes en rekke aspekter som skal være dekket i undervisningen¹¹.

Å utarbeide slike standarder er både tid- og kostnadskreven, og kan også være konserverende om ikke kriteriene kontinuerlig tilpasses den faglige utviklingen. Slike standarder gir likevel mindre fare for ensretting og press mot å fylle fellesfaglige

¹¹ <http://www.qaa.ac.uk/academicinfrastructure/benchmark/honours/chemistry.pdf>

resultatmål, enn når man opererer med ett sett av kriterier felles for alle fag. I forhold til et felles sett av kriterier og standarder for alle fag, gir fagspesifikke kriterier og standarder økt faglighet og validitet uten at det slakkes på likebehandling, rettsikkerhet og reliabilitet.

Forundersøkelser

Et alternativ til fagspesifikke standarder er å ha egne ”forundersøkelser” forut for hver vurderingsprosess der en tilpasser og tolker kriteriesettet og de generelle standardene i forhold til det enkelte fag som skal vurderes. Slike prosesser kan legge grunnlag for akkrediteringsbeslutninger bedre tilpasset det enkelte fag. Forundersøkelser inngår i evalueringprosedyrene til Danmarks Evalueringsinstitutt. Formålet er å tilpasse evalueringen til det enkelte evalueringobjekt (Askling et al. 1998:13; Tvernö et al. 2005:40). Om slike forundersøkelser gir rom for eksempelvis å utdype og tilpasse vurderingskriteriene, kan de være et alternativ til å ha ferdig utarbeidede kriterier og standarder for alle fag. På samme måte som fagspesifikke standarder og kriterier gir en slik metode økt faglighet og validitet uten at det slakkes på likebehandling, rettsikkerhet og reliabilitet.

4.3 Hvordan ivareta fagligheten i vurderingsprosessen?

I forrige delkapittel skisserte vi muligheter for å balansere ulike hensyn og håndtere usikkerhet i faglige vurderinger. En annen tilnærming til denne problematikken er å se på hvilke roller og funksjoner som tillegges de ulike aktørene i en vurdering av faglig nivå og dokumenterte resultater. Her ser vi spesielt på hvilket handlingsrom og hvilken myndighet fagfeller gis i vurderingene.

At en ”evaluator” utøver faglig skjønn, forutsetter som regel at han selv mener å ha en viss faglig autoritet på området. Slik autoritet er selvsagt også avgjørende for evalueringens faglige legitimitet. Dette krever en balanse mellom hvilken type kriterier og standarder som skal ligge til grunn for akkrediteringen og hva slags evalueringsutvalg som benyttes. Om kriteriene er en blanding av klare, entydige krav og spørsmål som krever faglig skjønn, vil sammensetningen av evalueringskomiteen kunne ha stor betydning for hvilke kriterier som blir viktigst for komiteens konklusjoner. En faglig sterk komité vil kunne legge mye større vekt på kriteriene som krever faglig skjønn, enn det en komité med begrenset fagkompetanse vil kunne. Små komiteer med begrenset fagkompetanse vil eksempelvis ikke føle seg kompetente til å foreta faglige vurderinger som går på tvers av det som fremstår som klare formalkrav. En faglig sterk komité vil derimot kunne veie faglige vurderinger opp mot formalkrav og eksempelvis konkludere med at en institusjon scorer så sterkt på dokumenterte resultater av den faglige virksomheten at dette oppveier at den ikke oppfyller alle formalkrav til innsatsfaktorer i utdanningen.¹²

¹² Eksempelvis at lærekraftene har høy realkompetanse på tross av de ikke oppfyller alle krav til formalkompetanse.

Poenget her er at om en har ulike typer kriterier uten å klare regler for hvordan de skal veies mot hverandre, vil sammensetningen av evalueringsutvalget kunne være den beslutningen som avgjør hvordan kriteriene ”vektes”. Om kriteriene som skal ligge til grunn for vurderingene er ”fleksible” og fellesfaglige, blir det viktig å sikre at komiteene har kompetanse til å tilpasse kriteriene til det enkelte fag og ikke minst ressursrammer som tillater en tilstrekkelig grundig faglig vurdering. Et alternativ til store og fagtunge komiteer vil her være en komité som får assistanse fra individuelle eksperter. Ved å innhente uttalelser fra flere uavhengige fagfeller til de vurderingene som krever faglig skjønn, kan en både øke validiteten og reliabiliteten i de faglige vurderingene. Her kan en også lettere bruke internasjonale fagfeller og slik øke mulighetene til å sammenlikne faglig nivå på tvers av landegrensler.

Arbeidsdelingen mellom sekretariat og evalueringskomiteen har også betydning. Om sekretariatet har en klar rolle i å sikre at mandatet og kriteriene tolkes i tråd med intensjonene kan dette sikre likebehandling og reliabilitet. Det kan imidlertid også redusere fagligheten og validiteten om sekretariatet forhindrer at kriteriene og standarder tilpasses ulike utdanninger. Alternativt kan sekretariatet ha ansvar for en prosess som munner ut i mer fagspesifikke kriterier (en ”forundersøkelse”), eksempelvis ved å involvere fagmiljøene og/eller å trekke veksler på fagspesifikke standarder utarbeidet i andre land.

Et annet spørsmål er hvordan evalueringsobjektet involveres i evalueringen og hvilken kontakt det er mellom evaluatorene og evalueringsobjektet. Det er normalt to kontaktpunkter mellom evaluatorene og evalueringsobjektet: egevalueringene og miljøbesøkene. I forhold til en ekstern vurdering av faglig nivå er funksjonen til institusjonenes egevalueringer å fremskaffe dokumentasjon og relevante data for de faglige vurderingene. Miljøbesøkene har en mer uklar funksjon i forhold til faglig vurdering. Miljøbesøkene er en slags kombinasjon av informasjonsmøte og høringsprosess, og har en tillitskapende funksjon. Besøkene gir de evaluerte innblikk i vurderingsprosessen og mulighet til direkte innspill til evaluatorene. Ved behov kan besøkene også bidra til utdypning av informasjon fra andre kilder, men vil sjelden ha betydning for faglige vurderinger. Et alternativ eller supplement til ordinære miljøbesøk er ”inspeksjoner”. Uanmeldte inspektører er eksempelvis del av prosedyrene ved akkrediteringer i britisk høyere utdanning. Kontroller utført av faste inspektører er også vanlig prosedyre for tilsynsmyndigheter på en rekke andre områder. I forhold til å kontrollere at det faglige nivået på en utdanning oppfyller visse minstekrav er trolig inspeksjoner et bedre redskap enn ordinære miljøbesøk. Uanmeldte inspektører er på den andre siden mindre egnet til å gi de evaluerte innblikk i vurderingsprosessen og mulighet til direkte innspill til evaluatorene.

4.4 Avsluttende kommentar

De overordnede konklusjonene fra dette notatet er at ”faglig nivå” og ”dokumenterte resultater” kan tolkes på mange ulike måter og at det er en rekke ulike kriterier som kan benyttes i vurderingene, men at det er en iboende usikkerhet i enhver faglig vurdering. Det er skissert ulike måter å balansere de ulike hensyn som akkrediteringsprosessen skal ivareta. Hvilke av disse alternativene som vil være best egnet for NOKUT vil avhenge av forhold som ligger utenfor denne studien.

Det er likevel flere uavklarte spørsmål hvor videre studier kunne gi et bedre kriterie- og metodegrunnlag for å vurdere faglig nivå og dokumenterte resultater i akkrediterings-sammenheng. En studie av hvordan NOKUTs komiteer har forholdt seg til kriterier som forutsetter faglig skjønn, og hvordan de har forholdt seg til mer entydige kriterier og standarder, ville eksempelvis gi et bedre grunnlag for å avgjøre hvilken grad av skjønn som kriterier og standarder bør gi rom for, og hvordan sakkyndige utvalg bør settes sammen. En undersøkelse av hvilke muligheter som finnes for å dra veksler på mer fagspesifikke kriterier og standarder som benyttes i andre land, ville også bidra til å styrke beslutningsgrunnlaget.

Referanser

- Askling, Berit, Lars Nordskov Nielsen & Bjørn Stensaker (1998): *Mellom fag og politikk. En granskning av Evalueringscenteret og det danske evalueringssystemet for høyere utdanning*. Undervisningsministeriet juni 1998.
- Arnesen, Clara Åse & Sverre Try (2001): *Karakterers betydning for overgangen fra høyere utdanning til arbeidsmarkedet* Oslo: NIFU rapport 6/2001.
- Berg, Lisbeth (1995): *Examen Philosophicum: studietilknytning, innsats og resultat for ulike grupper av begynnerstudenter ved Universitetet i Oslo*. Oslo: Utredningsinstituttet for forskning og høyere utdanning, Rapport 1995:2.
- Chubin, Daryl E. & Edward J. Hackett (1990): *Peerless Science*, New York: State University of New York Press
- Cole, Stephen, Jonathan R. Cole & Gary A. Simon (1981): "Chance and Consensus in Peer Review" *Science* 214:881–886.
- Ewell, Peter (2004): *The National Survey of Student Engagement (NSSE)*. The University of North Carolina at Chapel Hill, PPAQ Research Program, <http://www.unc.edu/ppaq/docs/NSSE.pdf>
- Grøgaard, Jens B. (1995): *Skolekontroversen: belyst ved to norske utvalgsundersøkelser* Oslo: FAFO, Doktorgradsavhandling.
- Hansen, Hanne Foss & Birte Holst Jørgensen (1995): *Styring af forskning: Kan forskningsindikatorer anvendes?* Frederiksberg: Samfundslitteratur.
- Helland, H. *Sosial ulikhet blant sosial- og siviløkonomer*. Oslo: Universitetet i Oslo Doktorgradsavhandling, Rapport 1:2004 Institutt for sosiologi og samfunnsgeografi.
- Hovdhaugen, Elisabeth (2004): *Tidsbruk og ambisjon. Resultater fra stud.mag.-undersøkelsene 2001, 2002 og 2003*. Oslo: NIFU skriftserie 16/2004.
- Hovdhaugen, Elisabeth (2005): *Karaktersetting i etterkant av Kvalitetsreformen - endringer i strykprosent*. Oslo: NIFU STEP Arbeidsnotat 36/2005.
- KUF (1995): *Felles stillingsstruktur for undervisnings- og forskningsstillinger ved høyskoler og universiteter. Reglementer for opprykk til førsteamanuensis, førstelektor og førstebibliotekar*. Oslo: Kirke-, utdannings- og forskningsdepartementet, Rundskriv F-14-95, 01.02.95.
- Kuh, George & Ernest T. Pascarella (2004): "What does institutional selectivity tell us about educational quality?" *Change* 36(5): 52-58.
- Kyvik, Svein, Terje Bruen Olsen & Elisabeth Hovdhaugen (2003): *Opprykk til professor. Kompetanse eller konkurranse?* Oslo: NIFU Rapport 4/2003.
- Lindsey, Duncan (1988): "Assessing precision in the manuscript review process: A little better than a dice roll" *Scientometrics* 14(1-2): 75-82.
- Langfeldt, Liv (2002): *Decision-making in expert panels evaluating research. Constraints, processes and bias*. Oslo: NIFU Rapport 6/2002.
- Merton, Robert K. (1968): "The Matthew Effect in Science" *Science* 159: 56-63.
- Merton, Robert K. (1988): "The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property" *ISIS* 79(299): 606-623.
- Pascarella, Ernest T. (2001): "Identifying excellence in undergraduate education: Are we even close?" *Change* 33(3): 18-23.
- Pike, Gary R. (2003): *Measuring Quality: A Comparison of U.S. News Rankings and NSSE Benchmarks*. Paper presentert på AIE 2003, Tampa (FL), Mississippi State University.

- Saraff, Shimon, George Kuh, John Hayek, Camille Kandiko, Ryan Padgett & Kim Harris (2005): *The Institutional Quality Debate: U.S. News Indicators and the National Survey of Student Engagement*. Paper presentert på ASHE 2005 i Philadelphia (PA), Indiana: Center for Postsecondary Research, Indiana University.
- Stensaker, Bjørn (1998): *Evaluering av studie- og læringsmiljø. En veiledning for faglig og administrativt personale ved universiteter og høyskoler* Oslo: Cappelen Akademisk Forlag.
- Stensaker, Bjørn (2004): "The blurring boundaries between accreditation and audit – the case of Norway". Side 342-365 i S. Schwarz & D. F. Westerheijden (red.): *Accreditation in the framework of evaluation activities. Current situation and dynamics in Europe*. Dordrecht: Kluwer Academic Press.
- Stensaker, Bjørn & Lee Harvey (2006): "Old wine in new bottles? A comparison of public and private accreditation schemes in higher education" Forthcoming in *Higher Education Policy*.
- Travis, G. D. L. & Harry M. Collins (1991): "New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System" *Science, Technology & Human Values* 16(3): 322-341.
- Tvernö, Henrik, Sigbrit Franke, Kirsi Lindroos, Rolf Sandahl & Eva Åström (2005): *Med andras ögon. Ekstern utvärdering av Danmarks Evalueringsinstitut*. Stockholm: Höskoleverket, Höskoleverkets rapportserie 2005:39R.
- UFD (2003): *Forskrift om akkreditering, evaluering og godkjenning etter lov om universiteter og høyskoler og lov om private høyskoler*. Oslo: Utdannings- og forskningsdepartementet, 2.1.2003.
- Wiers-Jenssen, Jannecke, Bjørn Stensaker & Jens Grøgaard (2002): "Student Satisfaction: towards an empirical deconstruction of the concept" *Quality in Higher Education* 8(2):183-195.
- Aamodt, Per Olaf (2001): *Studiegjennomføring og studiefrafall. En statistisk oversikt*. Oslo: NIFU Skriftserie 14/2001.