

Evaluering av norsk forskning og høyere utdanning

Rapport med forslag til et nasjonalt rammeverk

Utarbeidet av en arbeidsgruppe nedsatt av Kunnskapsdepartementet

Levert 1. februar 2021

Innhold

1. Forord	2
2. Sammendrag.....	3
3. Mandat, sammensetning og arbeidsform	4
4. Innledning.....	7
5. Brukere, behov og formål.....	10
6. Nærmere om evaluering av forskningskvalitet	15
7. Nærmere om evaluering av kvalitet i høyere utdanning	24
8. Koblinger og koordinering	28
9. Mot et nasjonalt rammeverk i Norge.....	32
10. Arbeidsgruppens forslag til nasjonalt rammeverk	43

1. Forord

Kunnskapsdepartementet nedsatte sommeren 2019 en arbeidsgruppe som fikk mandat til å utarbeide et forslag til nasjonalt rammeverk for evaluering av kvalitet i norsk forskning og høyere utdanning.

Arbeidsgruppen takker for et interessant oppdrag, og leverer med dette sin rapport. Anbefalingene i rapporten er enstemmige. Gruppens arbeid er nærmere beskrevet i kapittel 3, og forslaget til rammeverk er å finne i kapittel 10.

Medlemmene i arbeidsgruppen vil ved avslutningen av arbeidet benytte anledningen til å minnes Steinar Johannesen, seniorrådgiver i Kunnskapsdepartementet, som var medlem i gruppen og gjorde en stor innsats i oppstarten av dette arbeidet. Det var et stort tap for oss alle da Steinar døde brått og uventet i april 2020, bare 53 år gammel. Mange i arbeidsgruppen kjente Steinar godt, og enkelte hadde ham som nær kollega og venn gjennom mange år. Vi minnes Steinar som en uvanlig klok, vennlig og kunnskapsrik mann.

2. Sammendrag

Denne rapporten analyserer behovene for evalueringer av kvaliteten i norsk forskning og høyere utdanning, drøfter status og utfordringer for dagens evalueringsarbeid, og kommer med forslag til et nasjonalt rammeverk for dette arbeidet. Rapporten er utarbeidet av en bredt sammensatt arbeidsgruppe som ble oppnevnt av KD sommeren 2019. Arbeidsgruppen har blant annet bygget på en kartlegging av bruken av evalueringer utført av Proba samfunnsanalyse og en seminarserie i regi av prosjektet R-QUEST ved NIFU. I kapittel 3 gis en nærmere omtale av arbeidsgruppen, dens mandat og arbeidsform.

I kapittel 4 gis en innledende beskrivelse av hvordan evalueringer av kvalitet har vært gjennomført i Norge, og hva som er status i dag. Systemene i andre land beskrives, med vekt på Storbritannia og Nederland.

I kapittel 5 drøftes ulike aktørers behov for informasjon om kvaliteten på norsk forskning og høyere utdanning, med utgangspunkt i en EU-rapport om forskningsevalueringer og rapporten fra Proba om norske aktørers informasjonsbehov. Konklusjonen er at behovene varierer, men at det også er betydelig overlapp. Institusjonene som utfører forskning og høyere utdanning har de mest komplekse behovene. Med deres behov som utgangspunkt, kan det gjennomføres evalueringer som også svarer på andre aktørers informasjonsbehov.

I kapittel 6 og 7 gis nærmere beskrivelser av hvordan evaluering av forskning og høyere utdanning har vært utført i Norge, og hvilke utfordringer disse evalueringene står overfor. I kapittel 8 beskrives koordineringsbehovene, med særlig vekt på koblingen mellom forskning og høyere utdanning og aktørenes behov for forutsigbarhet. Probas kartlegging av brukernes behov peker særlig på bedre forutsigbarhet som et viktig mål.

I kapittel 9 drøfter arbeidsgruppen viktige momenter og veivalg for et nasjonalt rammeverk på bakgrunn av gjennomgangen i de foregående kapitlene. Blant spørsmålene som drøftes er: ulike aktørers informasjonsbehov og formålene med evalueringer; behovene for tilpasninger til institusjonenes ulikheter; summative vs. formative evalueringer; behovene for kobling mellom evaluering av forskning og utdanning; nasjonalt vs. desentralisert ansvar; legitimitet og troverdighet; samt koordineringsbehov og koordineringsmekanismer.

I kapittel 10 presenterer arbeidsgruppen sitt forslag til rammeverk. Rammeverket inneholder et målbilde som skal ivareta alle relevante aktørers informasjonsbehov om kvaliteten i norsk forskning og høyere utdanning. Forslaget viderefører i stor grad dagens ansvarsforhold, i den forstand at Forskningsrådet fremdeles har hovedansvaret for forskningsevalueringer og NOKUT et tilsvarende hovedansvar for utdanningsevalueringer. Det legges imidlertid opp til langt bedre løpende koordinering, blant annet gjennom et eget koordineringsforum. Tilpasning til institusjonenes ulikhet foreslås ivaretatt gjennom bruk av nasjonale protokoller med rom for spesifisering av lokale forhold av betydning for evalueringen. Videre foreslås det at NOKUTs utdanningsevalueringer utvides fra noen utvalgte programtyper til å dekke bredden i norsk høyere utdanning. For å unngå dobbeltrapportering foreslås det å endre bestemmelsen i Studiekvalitetsforskriften om at institusjonene må gjennomføre periodiske evalueringer av sine studieprogrammer. I tillegg inneholder rammeverket noen sentrale prinsipper for evalueringen av forskning og høyere utdanning i Norge.

3. Arbeidsgruppens sammensetning, mandat og arbeidsform

KD nedsatte sommeren 2019 en arbeidsgruppe som fikk i oppdrag å komme med forslag til et nasjonalt rammeverk for evaluering av norsk forskning og høyere utdanning.

Arbeidsgruppen har vært ledet av Kunnskapsdepartementet (KD), med medlemmer utpekt av Universitets- og høyskolerådet (UHR), Forskningsinstituttene fellesarena (FFA), Nasjonalt organ for kvalitet i utdanningen (NOKUT), Forskningsrådet, Norsk studentorganisasjon (NSO) og Olje- og energidepartementet (OED), og har bestått av følgende personer:

Vidar Grøtta, seniorrådgiver, KD (leder)
Steinar Johannesen, seniorrådgiver, KD (til april 2020)
Åsa Sandnes, seniorrådgiver, KD (fra august 2020)
Stine Grønvold, prorektor ved HINN, UHR
Marit Reitan, dekan ved NTNU, UHR (til september 2020)
Åse Gornitzka, viserektor ved UiO, UHR (fra oktober 2020)
Sveinung Skule, direktør i NIFU, FFA (til oktober 2019)
Espen Solberg, forskningsleder i NIFU, FFA (fra november 2019)
Stephan Hamberg, assisterende avdelingsdirektør, NOKUT
Jon Holm, spesialrådgiver, Forskningsrådet
Terje-André Kvinlaug, student ved UiB, NSO
Espen Bernhard Kjærgård, seniorrådgiver, OED

Arbeidsgruppen har forankret sitt arbeid hos ledelsen i HFI/EIE i KD.

Opprinnelig frist for å levere rapport var 1. november 2020. Fristen ble utsatt til 1. februar 2021.

Gruppen fikk følgende mandat:

Bakgrunn

Det er helt avgjørende for Norge som kunnskapssamfunn at utdannings- og forskningsaktiviteten har høy kvalitet. Det er derfor viktig at vi som samfunn kan vite hvor god den norske forskningen og de norske utdanningene er.

I Norge har det vært gjennomført evalueringer av forskningskvalitet siden 1980-tallet, og av kvaliteten i høyere utdanning siden 2000-tallet. Forskningsrådet har i dag et ansvar for å "sørge for evaluering av norsk forskning" (Forskningsrådets vedtekter, § 2), mens NOKUT har ansvar for å "gjennomføre evalueringer av betydning for å kunne bedømme kvaliteten i høyere utdanning" (UH-loven, § 2-1 (4)). Samtidig er det et grunnleggende prinsipp at institusjonene selv har ansvar for at forsknings- og utdanningsaktiviteten har høy kvalitet, og flere institusjoner har gjennomført evalueringer av egen virksomhet.

Vurdering av kunnskapsinstitusjonenes faglige aktivitet gjøres på mange måter og på ulike nivåer, og er dessuten fulgt av en løpende diskusjon om hvordan best måle hva som er god kvalitet.

I senere tid har det vært reist spørsmål om veien videre for evaluering av norsk forskning og høyere utdanning. I kjølvannet av SAMEVAL/KOMBEVAL har Forskningsrådet signalisert at de ønsker å tenke nytt om sin evalueringspraksis, og NOKUT har også evalueringer på sin dagsorden. UHR har på vegne av universitetene og høyskolene etablert en arbeidsgruppe som skal se nærmere på et beslektet tema, nemlig systemene for vurdering av forskning og forskere i lys av pågående endringsprosesser, bl.a. når det gjelder åpen forskning.

Begrepet evaluering brukes om mange ulike vurderingsprosesser på ulike nivåer. I denne sammenhengen menes større evalueringsprosesser forankret på nasjonalt nivå med det formål å vurdere kvaliteten på et avgrenset område (faglig, tematisk eller grupper av institusjoner) innenfor forskning og/eller høyere utdanning. Med forskning menes hele bredden av norsk forskning, dvs. både grunnforskning og utfordringsdrevet forskning, disiplinforskning og tverrfaglig forskning. I dag gjennomføres evalueringer av flere organer, ofte samtidig og ukoordinert. Det er derfor behov for å vurdere å utvikle et nasjonalt rammeverk for denne typen evaluering av forskning og høyere utdanning, slik en del andre land har. Dette vil gjøre evalueringsprosessene mer forutsigbare, de vil få større effekt og være mer effektive.

For å bidra til at arbeidet med evaluering av forskning og utdanning ses i sammenheng, har Kunnskapsdepartementet etablert dette utviklingsprosjektet, som skal kartlegge situasjonen og komme med anbefalinger for videre arbeid.

Mål og rammer for KDs utviklingsprosjekt

Målet med prosjektet er å sikre effektiv og koordinert evaluering av norsk forskning og høyere utdanning på en slik måte at ulike aktørers og samfunnets behov for kunnskap om kvaliteten i forskning og høyere utdanning imøtekommes.

Som ledd i prosjektet tar KD, i dialog med prosjektgruppen, sikte på å lyse ut et eksternt oppdrag som skal kartlegge de ulike aktørenes behov for og bruk av evaluering av norsk forskning og høyere utdanning. Med aktørene menes universiteter og høyskoler, institutter, departementer, direktorater, Forskningsrådet og NOKUT. På bakgrunn av kartleggingen skal prosjektet komme med forslag til et nasjonalt rammeverk for den samlede evalueringsvirksomheten av forskning og høyere utdanning.

Prosjektet skal se evaluering i sammenheng med monitorering og lovpålagte evalueringsaktiviteter ved institusjonene og de nasjonale organene. Arbeidet må også ses i sammenheng med andre relevante pågående prosesser.

Prosjektet skal komme med forslag til tiltak som skal sikre:

- ivaretagelse av behovene for evaluering*
- avklaring av ansvar, roller og oppgaver*
- ønsket grad av koordinering, herunder at den samlede respondentbelastningen og belastningen på miljøene som evalueres, holdes så lav som mulig*

Prosjektet skal ha et perspektiv som omfatter forskning og høyere utdanning i institutt- og UH-sektorene, og skal involvere representanter for disse, men prosjektet kan, om det viser seg mest hensiktsmessig, konsentrere sine forslag til tiltak om UH-sektoren.

Arbeidsgruppen har siden oppstarten hatt møter om lag annenhver måned, mot slutten noe hyppigere. Møtene har siden koronavirusutbruddet i mars 2020 i hovedsak vært digitale.

I tillegg til arbeidsmøtene har forskningsprosjektet R-QUEST ved NIFU arrangert fire seminarer med temaer av relevans for arbeidet.

Som bidrag til kunnskapsgrunnlaget for gruppens arbeid har Proba samfunnsanalyse på oppdrag fra KD utarbeidet rapporten *Kartlegging av evalueringer av kvalitet i forskning og høyere utdanning*.¹

¹ Proba Samfunnsanalyse (2020) *Kartlegging av evalueringer av kvalitet i forskning og høyere utdanning. Sluttrapport*.

4. Innledning

Kort historikk

Kvalitetskontroll av vitenskap gjennom fagfelleevaluering går tilbake til 1600-tallet, men systematiske evalueringer av hele fagmiljøer er av ganske ny dato. Storbritannia var først ute. Siden 1986 er det gjennomført fagfellebasert evaluering av alle disipliner i hele Storbritannia hvert 5. år. Etter hvert har mange land etablert nasjonale systemer for evaluering (se mer om dette nedenfor).

I Norge begynte de ulike forskningsrådene med spredte evalueringer av forskning i siste halvdel av 1980-årene. Da det fusjonerte Forskningsrådet ble etablert i 1993, ble evalueringen satt mer i system. Siden slutten av 1990-tallet har Forskningsrådet gjennomført en lang rekke evalueringer av forskningskvalitet, særlig fagevalueringer, som har vært gjennomført i to runder i de fleste fag. Den første runden med evalueringer ble gjennomført over mer eller mindre samme lest – som enkeltfagsevalueringer basert på omfattende lesning av vitenskapelige arbeider – mens utviklingen i andre runde er gått i retning av stadig bredere evalueringer som inkluderer hele fagområder og flere evalueringsdimensjoner, slik som samspill mellom forskning og utdanning, og samfunnseffekter av forskning.

På oppdrag fra Forskningsrådet har NIFU gjort to såkalte metaevalueringer der ulike sider ved evalueringsvirksomheten ble vurdert.² Enkelte av de tidlige evalueringene av norsk forskning på 1980- og 1990-tallet omfattet også utdanning, men det var først da Norgesnettrådet ble opprettet i 1997 at evaluering av høyere utdanning ble mer systematisert. Særlig lærerutdanning ble gjenstand for viktige evalueringer rundt årtusenskiftet. Etter at NOKUT overtok evalueringsansvaret, har lærerutdanninger fortsatt vært viktig, sammen med andre profesjonsutdanninger som for eksempel ingeniørutdanning. I en periode har evalueringsvirksomheten for det meste vært rettet mot tematiske problemstillinger, før KOMBEVAL ble iverksatt i 2014, med nasjonal evaluering av fire samfunnsvitenskapelige programmer.

Evalueringsarbeidet i Norge i dag

Høy kvalitet i norsk forskning og høyere utdanning har vært hovedmålet i en lang rekke politiske reformer og styringsdokumenter de siste tiårene, fra Kvalitetsreformen via Strukturreformen til Kvalitetsmeldingen og den nåværende Langtidsplanen for forskning og høyere utdanning.³ Evalueringer har stått sentralt i mye av dette arbeidet, både som et kvalitetstiltak i seg selv, og som kunnskapsgrunnlag for andre tiltak.

Organiseringen av evalueringsarbeidet og ansvarsforholdene på feltet har imidlertid ligget fast i lang tid. Forskningsrådet har siden 1993 hatt det samme ansvaret for å “sørge for evaluering av norsk forskning” (Forskningsrådets vedtekter § 2) mens NOKUT siden opprettelsen har hatt ansvar for å “gjennomføre evalueringer av betydning for å kunne bedømme kvaliteten i

² Se K. E. Brofoss (1997) *Metaevalueringen. En gjennomgang av Norges forskningsråds evalueringspraksis*. Norges forskningsråd, og K. E. Brofoss (2004) *En gjennomgang av Forskningsrådets fagevalueringer*. Arbeidsnotat 7. NIFU. Se også Liv Langfeldt (1999) *Evaluering av forskningskvalitet Et kritisk blikk på fagfelleevaluering*. NIFU skriftserie nr. 9/99.

³ Se Meld. St. 18 (2014-2015) *Konsentrasjon for kvalitet — Strukturreform i universitets- og høyskolesektoren*, Meld. St. 16 (2016-2017) *Kultur for kvalitet i høyere utdanning*, og Meld. St. 4 (2018-2019) *Langtidsplan for forskning og høyere utdanning 2019-2028*.

høyere utdanning” (UH-loven, § 2-1 (4)). Hvordan Forskningsrådet og NOKUT håndterer dette ansvaret i dag er nærmere omtalt i kapittel 6 og 7.

KD har, som overordnet myndighet for sektoren og etatsstyringsansvarlig for Forskningsrådet og NOKUT, et systemansvar også når det gjelder evaluering. Når det gjelder høyere utdanning er det dessuten tatt inn en hjemmel i UH-loven for at departementet skal kunne pålegge NOKUT å gjennomføre bestemte evalueringer ved behov, jf. UHL § 2-1 (4). Denne hjemmelen har vært brukt ved enkelte anledninger, og KD har også gitt Forskningsrådet og NOKUT i oppdrag å samarbeide om en bestemt evaluering (SAMEVAL/KOMBEVAL i 2014). Utover slike enkeltstående initiativer har departementet fram til nå ikke gitt bestemte rammer for evalueringsvirksomheten, og heller ikke sett de ulike formene for evaluering som deler av et helhetlig system, slik man har gjort i en del andre land (se mer om dette i neste kapittel).

I tråd med det grunnleggende prinsippet om at institusjonene selv har ansvar for at forsknings- og utdanningsaktiviteten har høy kvalitet, gjennomfører institusjonene selv også ulike former for evalueringer av egen virksomhet. Siden 2017 har Studiekvalitetsforskriften pålagt institusjonene regelmessige evalueringer av utdanningskvaliteten. I senere år har mange av institusjonene også gjennomført større evalueringer av ulike sider ved virksomheten med utgangspunkt i målene i sine strategiske langtidspaner. I tillegg kommer ulike andre prosesser ved institusjonene med betydning for kvalitetsarbeidet, kanskje særlig når det gjelder utdanning, hvor det stilles krav til institusjonenes systematiske kvalitetsarbeid.

Evalueringsarbeidet i enkelte andre land

I den rapporten Proba har utarbeidet som del av kunnskapsgrunnlaget for dette prosjektet er det en gjennomgang av hvordan evalueringene av forskning og høyere utdanning er innrettet i Nederland, Sverige, Finland og Storbritannia.⁴ Gjennomgangen konsentrerer seg om universitets- og høyskolesektoren. Den går ikke detaljert inn i de enkelte landenes systemer, men legger vekt på å tydeliggjøre de avveiningene som ligger bak valg av ulike evalueringssystemer for utdanning og forskning i de enkelte landene. Når det gjelder utdanningsevalueringer baserer Proba seg på en bred tilnærming til kvalitetsarbeidet og inkluderer tilsyn.

Systemene i de fire landene er så kompliserte og sammensatte at det ikke alltid er full enighet om hvordan de best kan beskrives og kategoriseres. I gjennomgangen framkommer det likevel tydelige forskjeller mellom dem. Forskjellene er blant annet knyttet til organisering, innsamling og bruk av data, hvilket nivå som blir evaluert, hvem som utfører evalueringene og hvordan de er koblet til finansering og politikkutvikling.

Ett viktig skille går mellom evalueringssystemer som er innrettet mot å avdekke særegenheter ved institusjoner/fagmiljøer og systemer som skal gjøre det mulig å sammenligne dem. De britiske evalueringssystemene er i den siste kategorien, noe som har medført at det er etablert klare nasjonale retningslinjer for både forsknings- og utdanningsevalueringene sine. Utdanningsevalueringene utføres av evalueringsorganet The Quality Assurance Agency for Higher Education (QAA), og bygger i stor grad på løpende indikatorer. Storbritannia har

⁴ Proba (2020), vedlegg 2.

imidlertid gjort noen endringer de siste årene blant annet ved at QAA har samarbeidet med sektoren og institusjonene om å utvikle den kvalitetsstandarden (Quality Code) som alle studieprogrammene skal følge.

Research Excellence Framework (REF) er det britiske systemet for å evaluere forskningskvalitet, og er store standardiserte og regelmessige revisjonslignende gjennomganger av forskningen. Resultatene i REF brukes i fordeling av forskningsmidler. Til tross for at britene legger mye arbeid i å lage felles målesystemer for forsknings- og utdanningskvalitet er sammenligning vanskelig også der. Det skyldes blant annet at undersøkelsesenheter ikke er helt konsistente over tid.

Probas sammenligning av ulike land trekker også opp et skille mellom formative og summative evalueringer. Igjen stikker Storbritannia seg ut, med et evalueringssystem som ligger nært et summativt ideal. Det knyttes til at evalueringene baserer seg på oppnådde resultater, men også til at det er lite skreddersøm og lite kontakt mellom evaluator og de som evalueres før og under evalueringen.

Nederland trekkes fram som et land med sterk vekt på det formative elementet. Der er institusjonene og enhetene med på å sette målene for evalueringene, oppnevne evalueringskomité, gjennomføre selvevalueringer, m.m.. Akkrediteringssystemet i Nederland er også blitt endret til å fokusere mer på gjensidig tillit og ansvarliggjøring av institusjonene. For å legge til rette for oppfølging retter evalueringene i Nederland seg mot mange målgrupper, både forskerne og ledelsen ved de ulike forskningsinstitusjonene, samt myndigheter, privat sektor og samfunnet generelt.

På utdanningssiden framheves det svenske systemet som en kontrast til det summative britiske. Svenskene har ifølge Proba lite fokus på direkte målbare resultater og mye på prosesser som forbedrer studentenes læring. Alle landene som er undersøkt har likevel en form for evaluering som ser på institusjonens kvalitetssikringssystemer som helhet. Alle landene baserer seg også på fagfellellevurderinger, både når de evaluerer og akkrediterer.

Storbritannia har en fast syklus for sine utdannings- og forskningsevalueringer. Det samme har Nederland, men der kan institusjonene i større grad bestemme hvilke enheter som skal inngå i evalueringen. Også i Sverige og Finland gjennomføres regelmessige summative oppsummeringer av forskningen, inkludert bibliometriske analyser.

I Probas rapport trekkes det også fram at det er viktig å gjøre kost-nytte avveininger i samband med evalueringsarbeidet. Evalueringer er ett av flere kvalitetsvirkemidler, og det må vurderes om ressursene som brukes på evalueringer fører til større kvalitetsforbedringer enn om de brukes på andre tiltak. Hvor mye arbeid skal legges i innsamling av data? Hvilke vurderinger kan basere seg på grovere indikatorer? Spesielt de britiske evalueringene av utdanning og forskning er så omfattende og kostbare at de tidvis har blitt betraktet som en byrde for sektoren.

5. Brukere, behov og formål

I en vurdering av hvordan Norge best kan organisere og innrette arbeidet med evalueringer av kvalitet i forskning og høyere utdanning, er det naturlig å ta utgangspunkt i de overordnede målene for den virksomheten som skal evalueres. Universitets- og høyskolelovens angir som formål at UH-institusjonene "tilbyr høyere utdanning på høyt internasjonalt nivå" og "utfører forskning og faglig og kunstnerisk utviklingsarbeid på høyt internasjonalt nivå" (§1-1). For instituttsektoren er det overordnede målet gitt gjennom regjeringens nylig framlagte instituttstrategi, hvor det heter at "instituttsektoren skal utvikle kunnskapsgrunnlag for politikkutforming og bidra til bærekraftig utvikling og omstilling, gjennom forskning av høy kvalitet og relevans".⁵

Det er imidlertid en lang rekke aktiviteter og tiltak som skal bidra til at disse overordnede målene nås. Det primære er institusjonenes faglige virksomhet og deres strategiske arbeid for å utvikle den faglige kvaliteten, som ofte omtales som "kvalitetsarbeidet". I tillegg kommer det statlige virkemiddelapparatet. For å komme fram til et retningsgivende målbilde for evalueringsvirksomheten, må vi spørre hva som skal være det spesifikke bidraget til den formen for evalueringer som er tema for denne rapporten. I en slik konkretisering av evalueringsvirksomhetens særlige funksjon og formål, er det naturlig å ta utgangspunkt i hvilke aktører som har behov for evalueringer av denne typen, og spørre hvilken bruk de kan tenkes å gjøre av dem.

Det er gjort en del forskning og utredningsarbeid internasjonalt på grunnlag av erfaringer fra ulike land gjennom de tiårene som er gått siden denne typen evalueringer av kvalitet i forskning og høyere utdanning først ble iverksatt. I en synteserapport om forskningsevaluering som en bredt sammensatt ekspertgruppe utarbeidet for EU-kommisjonen i 2010, gis det en oppsummering i form av en tabell med stikkordsmessig oversikt over aktørene og deres behov.⁶ Tabell 1 nedenfor er utarbeidet på grunnlag av tabellen i denne synteserapporten, men tilpasset den norske konteksten og det norske organisasjonslandskapet.

Tabell 1 Brukere av evalueringer og deres behov

Brukere:	Kan bruke evalueringer til:
<i>Den enkelte forsknings- og/eller høyere utdanningsinstitusjon</i> • Styret • Ledelse og administrasjon • Forskergrupper • Studieprogrammer	• Strategi og planlegging • Posisjonering • Kvalitetsforbedring • Kvalitetssikring • Synliggjøring av kvalitet for å skape tillit og legitimitet • Benchmarking • Finne samarbeidspartnere • Vurdere konkurransesituasjonen
<i>Nasjonale myndigheter og myndighetsorganer innen forskning og høyere utdanning</i> • Regjeringen • Kunnskapsdepartementet • Forskningsrådet • NOKUT • Diku	• Policyarbeid på nasjonalt systemnivå • Etatsstyring av statlige institusjoner • Vurdering av kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Utvikle tiltak for forbedring av kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter

⁵ Kunnskapsdepartementet (2020) *Strategi for helhetlig instituttpolitikk*, s. 4.

⁶ European Commission (2010) *Assessing Europe's University-Based Research. Expert Group on Assessment of University-Based Research*

• Unit • Kompetanse Norge	• Vurdere nasjonal konkurransesituasjon i internasjonal kontekst • Synliggjøring av kvalitet for å skape tillit og legitimitet
<i>Akademiske organisasjoner</i> • UHR • Vitenskapsakademiene • Fagorganisasjoner og profesjonsorganisasjoner • Studentorganisasjoner	• Policyarbeid • Vurdering av kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Forbedring av kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Synliggjøring av kvalitet for å skape tillit og legitimitet
<i>Individuelle forskere og UH-lærere</i>	• Vurdere karrierevalg • Vurdere samarbeidspartnere • Vurdere akademisk infrastruktur og miljø
<i>Individuelle studenter</i>	• Vurdere studiested • Vurdere studieprogram og annen deltakelse • Vurdere karriere i academia
<i>Offentlige myndigheter utenfor forskning/høyere utd</i> • Øvrige departementer • Øvrige direktorater • Øvrige statlige tjenester • Kommunal sektor	• Vurdere kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Identifisere potensielle samarbeidspartnere • Identifisere leverandører av forskning og ekspertrådgivning • Identifisere potensielle arbeidstakere
<i>Privat næringsliv</i> • Arbeidsgiver/takerorganisasjoner • Enkeltbedrifter	• Vurdere kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Identifisere potensielle samarbeidspartnere • Identifisere leverandører av forskning og ekspertrådgivning • Identifisere potensielle arbeidstakere
<i>Tredje sektor (NGO/frivillige org)</i>	• Vurdere kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Identifisere potensielle samarbeidspartnere • Identifisere leverandører av forskning og ekspertrådgivning • Identifisere potensielle arbeidstakere
<i>Media/offentlig debatt</i>	• Sammenligne kvalitet på tvers av institusjoner, nasjonalt og internasjonalt • Vurdere kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Informere om studievalg og karrierevalg • Belyse grunnlaget for tillit og legitimitet
<i>Internasjonale organisasjoner (FN, OECD, EU, etc.)</i>	• Policyarbeid • Vurdere nasjonal konkurransesituasjon i internasjonal kontekst • Vurdere kvalitet, produktivitet, bærekraft, relevans og samfunnseffekter • Identifisere potensielle samarbeidspartnere

Brukerne i tabellen kan grupperes langs ulike dimensjoner. De fleste er opptatt av både forskning og høyere utdanning, men enkelte er hovedsakelig opptatt av én av delene. Noen er produsenter av forskning og/eller høyere utdanning, noen har oppgaver på systemnivå og

noen er brukere. Noen har sitt ståsted innenfor forsknings- og høyere utdanningssektoren; noen kommer fra andre sektorer.

Det aktørene i tabellen har til felles, er at de har legitime behov for informasjon om kvaliteten i norsk forskning og høyere utdanning. Når det skal utarbeides et rammeverk for evaluering av forskning og utdanning i Norge, må det derfor gjøres en vurdering av hvordan og i hvilken utstrekning de ulike brukernes behov for informasjon kan dekkes gjennom evalueringer. Det kan være aktuelt å skille mellom behov hos 1) aktører som utfører forskning og høyere utdanning, 2) aktører i kunnskapssektoren som driver myndighetsutøvelse og/eller politikkutforming 3) brukere av og deltakere i forskning og høyere utdanning for øvrig. Ser man nærmere på behovene, er noen viktige forskjeller mellom disse kategoriene av brukere, selv om det også er stor grad av overlapp på andre områder, som også Probas analyse viser.

Aktører som utfører forskning og høyere utdanning har antakelig de mest sammensatte behovene. Som det framgår av tabellen ovenfor, har disse aktørene flere bruksområder av summativ karakter. Kvalitetstilstanden ved fagmiljøene kan måles mot sammenlignbare fagmiljø ved andre institusjoner og eventuelt også i en internasjonal kvalitetsmålestokk. Evalueringene vil utgjøre et viktig beslutningsgrunnlag, eksempelvis knyttet til strategiarbeid, synliggjøring av kvalitet, posisjonering og vurdering av konkurransesituasjonen. Summative vurderinger er imidlertid ikke tilstrekkelig som grunnlag for å få til kvalitetsforbedring, som i nesten alle tilfeller vil være en avgjørende målsetting. Til bruk i det løpende kvalitetsarbeidet er det nødvendig med formative vurderinger, råd og anbefalinger. Slike vurderinger må gis med utgangspunkt i konteksten ved den enkelte institusjon for å ha tilstrekkelig relevans.

Aktører i kunnskapssektoren som driver myndighetsutøvelse og/eller politikkutforming (herunder nasjonale myndigheter og akademiske organisasjoner) har behov for summative vurderinger av kvalitetstilstanden. Særlig for nasjonale myndigheter (jf. tabell 1) er det avgjørende å få kunnskap om hvilke fagmiljøer som er ledende, hvilke som holder jevnt god kvalitet og hvilke som har kvalitetsutfordringer innen ulike fagområder. Både etatsstyring av universiteter og høyskoler, utvikling av instituttpolitikken, finansiering, dimensjonering og innretting av virkemiddelapparatet for øvrig, må bygge på solid kunnskap om kvalitetstilstanden i ulike deler av sektoren. Samtidig er det også av betydelig interesse for både nasjonale myndigheter og akademiske organisasjoner å følge med på hvilke formative vurderinger som gjøres for de enkelte institusjoner i slike evalueringer, i tillegg til eventuelle formative vurderinger og råd som gis til det nasjonale nivået.

Brukere av og deltakere i forskning og høyere utdanning for øvrig vil ofte være opptatt av å finne ut av kvalitetstilstanden på et bestemt område. Det kan være en virksomhet som vurderer å inngå et samarbeid om forskning eller en student som skal søke opptak til et masterprogram. For slike aktører vil summative vurderinger av kvalitetstilstanden være det viktigste kunnskapsgrunnlaget. Både sammenligning av kvaliteten ved ulike institusjoner og nærmere beskrivelser av kjennetegn ved den faglige virksomheten på ulike institusjoner vil være av avgjørende betydning, mens mer formative vurderinger som handler om institusjonenes kvalitetsarbeid på lengre sikt ofte vil ha mindre interesse.

I tillegg til disse tre brukerkategoriene kan man se på mediene og internasjonale organisasjoner som en fjerde kategori med *observatører*. Ettersom deres inntrykk og beskrivelse av kvalitetstilstanden i norsk forskning og høyere utdanning ofte får stor betydning, er det viktig at de har et godt kunnskapsgrunnlag, og evalueringer av kvalitet vil kunne være en viktig informasjonskilde. Hvilket informasjonsbehov denne kategorien brukere

har, vil variere og kan være vanskelig å definere. I de fleste tilfeller vil lett tilgjengelige summative vurderinger utgjøre et viktig kunnskapsgrunnlag, men i enkelte situasjoner kan det også være behov for mer utfyllende informasjon. Det er grunn til å tro at grundige og fagfellebaserte evalueringer av kvalitet vil gi et bedre kunnskapsgrunnlag for både medier og andre observatører enn de omdiskuterte rangeringene det ofte vises til i medieoppslag og i den offentlige debatt.

Kartleggingen som er gjennomført av Proba samfunnsanalyse har gitt utdypende informasjon om brukernes behov. Proba oppsummerer sine funn slik:

Evalueringer utgjør et viktig kunnskapsgrunnlag for mange interessenter. Hva de bruker dette kunnskapsgrunnlaget til avhenger av hva slags aktør det er snakk om – de som jobber med forskningspolitikk bruker det i politikktutforming, mens de som arbeider på UH-institusjoner bruker det til utviklingsarbeid internt på institusjonen. Andre – som for eksempel fagforeninger og andre interesseorganisasjoner – kan også bruke evalueringene til politisk påvirkningsarbeid eller i strategisk arbeid. Hvordan evalueringene er innrettet – og hvordan de måler kvalitet – får også effekter internt på institusjonene, både på organisasjons- og individnivå. (...)

De fleste er enige om at evalueringer virker skjerpende på institusjonene og «tvinger» dem til å fokusere på kvalitetsarbeid og -utvikling. Her kommer det fram at hele prosessen rundt evalueringene oppleves som nyttig, ikke bare resultatet. I påvente av en evaluering, i forberedelsesfasen og gjennom å samle inn datamateriale, dokumentasjon og egnevalueringer, fører evalueringene til en stor grad av læring internt på institusjonene. (...)

Til sist ser vi at evalueringer i liten grad når ut til andre enn KD, evaluatorene (Forskningsrådet og NOKUT) og forskningsprodusentene selv. Enkelte andre departementer og direktorater følger med som «interesserte lesere» på evalueringer som dekker deres egne sektorer, men ellers når resultatene av evalueringer i liten grad ut til samfunnet for øvrig.⁷

På grunnlag av intervjuer og øvrig kartlegging, har Proba kategorisert ulike formål eller bruksområder for evalueringsvirksomheten. Siden de fleste respondentene var mer kjent med evaluering av forskning enn av utdanning, tar følgende liste først og framst utgangspunkt i bruken av forskningsevalueringer, selv om bruksformene analytisk sett har like stor relevans for utdanningsevaluering:

- «Accountability» og kunnskapsinnhenting
- Benchmarking og sammenligning nasjonalt og internasjonalt
- Kvalitetsutvikling/læring
- Forskningspolitikk [og utdanningspolitikk], prioritering og strategiske grep⁸

I tillegg gir Proba en egen drøfting av utdanningsevalueringer, med fokus på NOKUTs rolle, og framhever særlig to informasjonsbehov: "et ønske om å vite om studentene får det de har krav på, eller blir lovet", og "institusjonenes evne og forutsetninger for å drive egenutvikling".⁹

Samlet sett kan det se ut til at det på tross av viktige forskjeller mellom brukerne også er vesentlig overlapp i informasjonsbehovene. Ved å ta utgangspunkt i de aktørene som har mest komplekse behov, nemlig institusjonene som utfører forskning og høyere utdanning, samtidig

⁷ Proba 2020, ss. 2-3.

⁸ Proba 2020, ss. 8-10.

⁹ Proba 2020, s. 10.

som man også har de øvrige aktørenes behov for øye, har det i praksis vist seg mulig å finne fram til evalueringsdesign og evalueringsprosedyrer som kan imøtekomme mange av brukernes behov. Probas rapport viser til at mange evalueringer i Norge har vært såkalte hybrider som nettopp har tatt sikte på dekke flere formål. Samtidig kan forskjellene mellom brukernes perspektiver og interesser gjøre det nødvendig å differensiere formidlingen av evalueringenes resultater, slik at de tilpasses de ulike målgruppene. Studiesøkere og beslutningstakere i ulike virksomheter kan neppe forventes å lete seg fram i evalueringsrapporter som ofte har institusjonene og andre aktører i kunnskapssektoren som sine primære målgrupper. Formidling av resultater fra evalueringene til ulike målgrupper bør antakelig vies større oppmerksomhet framover.

Oppsummering og målformuleringer

Gjennomgangen av brukerne og deres behov i avsnittene ovenfor tyder på at kvaliteten i norsk forskning og høyere utdanning er av interesse og betydning for mange aktører.

Informasjonsbehovene er noe forskjellige, men overlappet er allikevel så stort at evalueringsarbeidet bør kunne innrettes slik at det, med visse tilpasninger og særskilte tiltak, svarer på flere behov.

Arbeidsgruppen mener derfor at evalueringsvirksomheten som omfattes av rammeverket bør innrettes slik at de bidrar til følgende målsettinger:

- a. kvalitetsarbeidet ved institusjonene
- b. kunnskapsbasert policyutvikling
- c. et godt kunnskapsgrunnlag for oppdragsgivere, studiesøkere og samarbeidspartnere
- d. tillit til forskning og høyere utdanning i befolkningen
- e. åpenhet om bruk av offentlige midler

6. Nærmere om evaluering av forskning

I dette kapittelet ser vi på hvordan evaluering av forskning i Norge har vært gjennomført som en del av Forskningsrådets virksomhet. Evalueringer av norsk forskning har vært gjennomført i ulike former tilbake til 70-tallet i Norge. De første evalueringene hadde karakter av kartlegginger slik som for eksempel *Humaniora-utredningen* og *Humanistisk forskning i Norge*.¹⁰ Etter at Forskningsrådet ble etablert i 1993, som en sammenslåing av fem tidligere råd for ulike fagområder og tema, kom evalueringene av norsk forskning inn i en fastere form med bruk av internasjonale fagfeller som har fått i oppdrag å vurdere kvaliteten i norsk forskning sett i et internasjonalt perspektiv.¹¹

Kontekst og historikk

Ansvar for "å sørge for evaluering av norsk forskning" ble tatt inn i vedtektene for Forskningsrådet (§2) i 1993. Den første fagevalueringen ble ferdigstilt i 1997, i kjemi. Forskningsrådet har de siste 20 årene gjennomført to runder med fagevalueringer for de fleste fag. Den første runden med evalueringer ble gjennomført over mer eller mindre samme lest med relativt smalt definerte evalueringer av enkeltfag hvor hovedvekten var på å beskrive den faglige utviklingen og vurdere forskningens kvalitet og bærekraft i de ulike deler av faget og ved den enkelte institusjon.¹²

For å belyse forhold knyttet til ledelse, organisering og balanse mellom UH-institusjonenes ulike formål, har Forskningsrådet i andre runde av fagevalueringene gått i retning av bredere evalueringer som inkluderer hele fagområder og flere evalueringsdimensjoner, slik som samspill mellom forskning og utdanning, institusjonenes organisering og strategier, og samfunnseffekter av forskning. Det sentrale oppgaven til evalueringskomiteen har hele tiden vært å vurdere om norsk forskning holder mål sett i et internasjonalt perspektiv. I de senere evalueringene er imidlertid perspektivet utvidet for å gi en bedre forståelse av hvordan organisering og strategier ved institusjonene bidrar til kvalitet i forskning, og betydningen av kvalitet i forskning for utdanningsoppdraget og institusjonenes samfunnsrolle mer generelt. Det har vært en ambisjon hos Forskningsrådet at fagevalueringene dermed skulle få en større relevans for faglige prioriteringer og strategisk ledelse ved institusjonene og nasjonalt.

En viktig oppgave for de internasjonale evalueringskomiteene har vært å gi råd til forskningsmiljøene, til Forskningsrådet og til departementene om tiltak som kan bidra til å styrke kvalitet og relevans av fagene. Evalueringene har altså hatt et *formativt* siktemål om å bidra til læring i tillegg til å gjøre en *summativ* vurdering av kvaliteten i forskning i seg selv.

Forskningsrådets ulike evalueringer

Forskningsrådets evalueringer av norsk forskning omfatter mer enn fagevalueringer. Utvikling i metoder og avgrensning av fagevalueringene må derfor ses i sammenheng med andre evalueringsaktiviteter. Utover evalueringer av egne virkemidler, evaluerer Forskningsrådet også norske forskningsinstitutter. Dette ansvaret er knyttet til forvaltningen av grunnbevilgningen til fire arenaer av forskningsinstitutter: miljø, primærnærings,

¹⁰ NAVF (1976) *Humaniora-utredningen*, 10 bind, og NAVF (1984) *Humanistisk forskning i Norge*, 3 bind.

¹¹ <https://www.forskningsradet.no/Statistikk-og-evalueringer/statistikk-og-evalueringer/evalueringer/vare-evalueringer/>

Se også Brofoss 2004 s. 23.

samfunnsvitenskapelige og teknisk-industrielle institutter. Forskningsinstituttene har hatt mulighet til å delta i relevante fagevalueringer, noe som har ført til en viss overlapp mellom evalueringsløp. Det kan også være overlapp mellom fagevalueringer og tematiske evalueringer – som brukes som underlag for oppfølging og utvikling av nasjonale strategiske prioriteringer (eksempelvis klimaforskning og utdanningsforskning). Forskningsrådet evaluerer dessuten jevnlig forskningssentre og andre institusjoner – både på oppdrag fra departementer og på eget initiativ til bruk i styringen (f.eks. for å vurdere videre finansiering).

Forskningsrådets mandat gir ingen føringer for hva som skal være evalueringsenheten som undersøkes. Formuleringen "sørge for evaluering av norsk forskning" må sies å være en meget vid formulering som gir Forskningsrådet stor frihet i å utforme de evalueringsløp og definere de evalueringsenheter som anses som nødvendige og formålstjenlige. Kunnskapsdepartementet kan imidlertid gi føringer gjennom sine årlige tildelingsbrev.

Fagevalueringer

I Forskningsrådets evalueringer har fag som evalueringsenhet blitt operasjonalisert gjennom vurdering av faglig tilhørighet for administrative enheter som universitetsinstitutter eller forskergrupper, eller gjennom en mer organisk forståelse av fagfellesskapet basert på enkeltforskeres selvforståelse eller publiseringspraksis. Ved universitetene har inndelingen i organisatoriske enheter historisk sett vært basert på disiplinfaglige skillelinjer. På nasjonalt nivå finner vi igjen referansen til de tradisjonelle akademiske disiplinene i organiseringen av det nasjonale fagfellesskapet i Norge, både i de nasjonale fagrådene under Universitets- og høyskolerådet (UHR) og i faginndelingen av tidsskrifter som benyttes i tellekantsystemet.¹³

Slikt samsvar har vært mindre åpenbart ved høyskolene og i profesjonsutdanningene som i større grad har funnet sin enhet ved en relasjon til fagets praktiske bruk utenfor akademien. I instituttsektoren har man på tilsvarende måte organisert virksomheten ut fra samfunnsoppdraget, heller enn basert på akademiske disipliner. Dette har medført diskusjoner om hvorvidt fagevalueringer – slik de er utformet av Forskningsrådet med inndeling i evalueringspaneler basert på fagdisipliner – er egnet til å vurdere kvaliteten ved den tverrfaglige og ofte problemorienterte forskningen ved instituttene, og i noen grad ved miljøer som springer ut av profesjonsutdanningene.

En annen måte å forstå 'fag' som en akademisk disiplinær enhet, er som kulturelt fellesskap av normer, regler og referanser av den typen som også kjennetegner profesjoner mer allment. Slike fagkulturelle normer vil også bidra til å skape sammenheng mellom de ulike delene av den akademiske praksis, og dermed gjøre skillet mellom forskning, utdanning og formidling mindre relevant.

Hvordan operasjonaliseres 'fag' som enhet i fagevalueringer?

Institusjoner

Da de første fagevalueringene ble gjennomført på slutten av 1990-tallet, var det i stor grad samsvar mellom den disiplinbaserte avgrensningen i fagevalueringene og en disiplinbasert organisering av UH-institusjonene. I løpet av de siste 20 årene har det skjedd en utvikling av

¹³ Norsk publiseringsindikator (<https://npi.nsd.no/>)

organiseringen i UH-sektoren i retning av større og mer tverrfaglige, ofte tematisk definerte, enheter. I den andre runden med fagevalueringer som startet rundt 2010, har Forskningsrådet forsøkt å ta hensyn til utviklingen i institusjonens organisering ved å evaluere bredere fagområder. Eksempler på dette er: Biologi, helse og medisin (2011), humaniora (2017) og samfunnsvitenskap (2018). Når slike brede områder skal vurderes, fordeles evalueringsenhetene gjerne på flere vurderingspaneler med eksperter som har spesialistkompetanse i fagene som skal evalueres. Panelene har gjerne vært disiplinbaserte, men med visse tilpasninger for å fange opp tverrfaglig forskning. Det kan likevel hevdes at disiplinen setter standard for hva som skal til for at en bestemt type kunnskap skal kunne anses som forskningsbasert, og at det dermed er krevende å vurdere kvalitet i forskning utenfor en disiplinær ramme.¹⁴

Forskergrupper

Den faktiske organiseringen av forskning skjer ofte på et nivå som ligger mellom den formelle administrative enheten (institutt, senter e.l.) og den enkelte forsker. Organisering i forskergrupper har lenge vært ansett som en effektiv måte å kombinere faglige, personalmessige og materielle ressurser innenfor naturvitenskap, medisin og teknologi. Banebrytende forskning gjøres ofte i grupper fordi de kan kombinere kunnskap og ferdigheter fra ett eller flere fag gjennom samlokalisering av relevante personer og ressurser. Forskergrupper er også en fleksibel måte å organisere aktiviteter som går på tvers av mer etablerte administrative enheter.

Forskergrupper har derfor vært den grunnleggende evalueringsenheten i evalueringer innen MNT-feltet. Her er gruppens prestasjoner blitt vurdert basert på medlemmenes CV'er og resultater dokumentert gjennom publikasjonslister. Komiteen har i lite grad vurdert publikasjonenes innhold.

Innen humaniora og samfunnsvitenskap er det mindre vanlig at forskning er organisert i grupper.¹⁵ I den første runden med evalueringer av humsam-fag ble forskningsmiljøer identifisert ved at institusjonene meldte inn enkeltforskere ved relevante enheter som institutter eller sentre. For å unngå at enkeltpersoner skulle kunne indentifiseres i vurderingene har det likevel vært krav om minimum fem aktive forskere i det aktuelle faget ved enheten. Forskergrupper ble introdusert som evalueringsenhet for humsam-fag i de siste evalueringene av humaniora og samfunnsvitenskap. Forskningsrådet la til rette for at både forskergrupper og individuelle forskere uten gruppetilknytning kunne meldes inn til evalueringen. Resultatet var at noe under halvparten av de innmeldte forskerne var knyttet til en forskergruppe.

¹⁴ Et eksempel på disiplinenes rolle som portvoktere for den akademiske kunnskapen kan vi finne i den seneste evalueringen av samfunnsvitenskap der komiteen gir følgende anbefaling: "Institutions should ensure an appropriate balance between interdisciplinary and disciplinary work to maintain high-quality scientific output. For example, strong and long-lasting links should be established between disciplinary and interdisciplinary units in a particular location." Forskningsrådet (2018) *Evaluation of the Social Sciences in Norway Report from the Principal Evaluation Committee*. s. 38.

¹⁵ Magnus Gulbrandsen (2016) *The humanities in Norway: research, research organisation and external engagement*, NIFU Report; 2016:36, s. 18.

Personer og publikasjoner

For å evaluere for eksempel humanistisk forskning må det lages oversikt over hvor det drives humanistisk forskning i Norge, og hvem som tilhører det humanistiske fagfellesskapet. I NIFUs forskerpersonaleregister knyttes personer til fag gjennom utdanningsbakgrunn, altså fra hvilke fag de har sine akademiske grader. Dette er informasjon som er brukt i analyse av personellressurser innen de ulike fag i Norge. Videre er det mulig å identifisere vitenskapelig personales fagtilknytning ved å se på hvor de publiserer. Kopling mellom publikasjon og fag gjøres enklest for tidsskriftsartikler fordi alle tidsskrifter som benyttes av forskere i Norge har fått definert fagtilknytning i den norske publiseringsindikatoren (NPI). Slik identifikasjon gjennom publiseringskanal er selvsagt ikke entydig, men kan likevel brukes for å identifisere relevante miljøer ut over tradisjonelle disiplinorganiserte enheter.

Organisering (roller og prosess)

Ansvar for planlegging og gjennomføring av fagevalueringer er i praksis blitt forvaltet av ulike fagstyrer i Forskningsrådet med representanter fra fagområdene som skal evalueres. Evalueringene er blitt initiert av det relevante fagstyret med sikte på at alle fag skal evalueres med om lag 10 års intervaller. Det er også Forskningsrådets fagstyrer som har fastsatt evalueringens mandat og oppnevnt eksperter til evalueringskomiteen. Forskningsrådet har hatt utstrakt dialog med fagmiljøene under planlegging og oppfølging av fagevalueringene. Institusjonen har kunnet spille inn forslag til eksperter, og opplegg for egevaluering er lagt fram for innspill, ofte i flere runder.

Når evalueringsmandatet er vedtatt, går stafettpinnen over til evalueringskomiteen bestående av internasjonale fageksperter. Innenfor mandatet gjør komiteen sine egne vurderinger og har ofte, og spesielt for de store evalueringene, blitt støttet av et sekretariat. Forskningsrådets administrasjon har bidratt i datainnsamling og har bestilt underlagsmateriale, for eksempel analyser av personale og publisering. Forskningsrådets administrasjon har hovedsakelig hatt en rolle som tilrettelegger for prosessen med armlengdes avstand til komiteens vurderingsarbeid. Dette for å sikre komiteens uavhengighet, særlig med tanke på at den også skal vurdere og gi råd om Forskningsrådets bidrag til fagutvikling.

Evalueringstema

Hva som skal være tema for evalueringene og hvilke spørsmål komiteen skal besvare er angitt i evalueringens mandat. Mens de første evalueringene brukte mest plass på forskningsinterne tema – slik som fagets egen historie og utviklingsmuligheter – har den siste runden av evalueringer lagt større vekt på den organisatoriske og samfunnsmessige kontekst hvor faget inngår.

Strategi, ressurser og organisering

I de siste evalueringene har spørsmål om institusjonens organisering, strategier og prioritering av ressurser opptatt en betydelig del av institusjonens egevaluering og komiteens rapport. Dermed har komiteen også i større grad kunnet gi råd om hvorvidt organisering og strategi ved den enkelte institusjon er formålstjenlig for å oppnå høy kvalitet i forskning. På bakgrunn av informasjon om institusjonenes ressurser og resultater har komiteene også kunnet vurdere om ambisjonene som uttrykkes i gjeldende strategier virker overambisiøse, eller kanskje ikke ambisiøse nok.

Forskningskvalitet

Selv om all forskning har noen fellestrekk og noen felles verdier, er det allikevel slik at kjennetegnene på forskning av høy kvalitet i praksis varierer fra fag til fag.¹⁶ Perspektivene på forskningskvalitet, og hvilke kjennetegn som blir vektlagt, kan også variere med vurderingens formål, for eksempel forskningens anvendelsesmuligheter versus fagintern teori- og metodeutvikling.¹⁷ Denne kompleksiteten har tradisjonelt vært håndtert ved at vurdering av forskning utføres av fagfeller med kjennskap til kvalitetsstandarder som gjelder for den enkelte forskningsdisiplin. Fordelen med slike ekspertvurderinger er at de ikke er avhengig av at vurderingskriteriene gjøres eksplisitte. Dermed kan ulike egenskaper ved forskningsvirksomheten, og dens resultater, vektlegges ulikt for forskjellige fag, innenfor et evalueringsregime med de samme overordnede evalueringskriteriene.

Fagevalueringene har støttet seg til dette prinsippet ved å benytte komiteer eller paneler sammensatt av fagfeller som arbeider innenfor de samme disiplinene som de evaluerte miljøene. Som hovedregel har derfor inndelingen av evalueringene vært disiplinbaserte. Dette har skapt noen utfordringer for vurdering av tverrfaglige miljøer som tidligere nevnt i rapporten.

Datagrunnlaget for vurdering av forskningskvalitet har i stor grad bestått av fagfellevurderte vitenskapelige publikasjoner. Som vi tidligere har omtalt, har måten publikasjonene er blitt vurdert på, variert fra fag til fag og over tid. Det er også en økende oppmerksomhet i dag om betydningen av andre typer av resultater av forskning, slik som datasett, programvare, audiovisuelle eller ulike materielle uttrykk som film, utstillinger eller prototyper. Slike resultater har i lite grad blitt inkludert i datamaterialet for fagevalueringene.

Forskningskvalitet er blitt vurdert basert på vitenskapelige publikasjoner, enten direkte ved lesning eller basert på ulike kjennetegn ved dem. Datamaterialet i de siste fagevalueringene har inkludert:

1. fulltekst publikasjoner (kun humsam)
2. bibliografiske data i form av publikasjonslister og lister over publikasjonskanaler
3. bibliometriske data slik som publiseringsnivå (NPI) og siteringer.

Det er også brukt mål på produktivitet slik som antall publiseringspoeng per forskerårsverk, med forbehold om at kvantitet ikke må forveksles med kvalitet.

Vurderingene av forskningskvalitet som utføres av evalueringenes paneler og komiteer er i stor grad basert på fagfellenes kunnskap om hva som kjennetegner kvalitet innenfor den enkelte disiplin. Innenfor hvert panel har fagfellene også diskutert seg fram til en felles vurdering av hvert enkelt evalueringsobjekt (gruppe, institutt e.l.) basert på en karakterskala definert av Forskningsrådet. Bruk av en felles karakterskala kan gjøre det mulig å sammenlikne kvalitet på tvers av miljøer og institusjoner, med visse viktige forbehold:

1. Karakterskalaen vil med nødvendighet legge vekt på et begrenset sett med egenskaper som kan være mer eller mindre relevant for den enkelte disiplin.

¹⁶ Se for eksempel R. K. Merton (1973) *The Sociology of Science – Theoretical and Empirical Investigations*. (Chicago & London: The University of Chicago Press).

¹⁷ Liv Langfeldt, Maria Nedeva, Sverker Sörlin og Dyla. A. Thomas (2019) "Co-existing Notions of Research Quality: A Framework to Study Context-specific Understandings of Good Research", *Minerva* 58.

2. Bruk av karakterskalaer må harmoniseres for å sikre at den brukes likt. Dette er mulig å gjøre mellom paneler som deltar i den samme evalueringsprosessen, men er vanskeligere å se for seg på tvers av separate evalueringsprosesser.
3. En karakter sier ikke alt. Det er en generell fare for at det som gis en karakter tillegges for mye vekt i forhold til deler av virksomheten som ikke gis en karakter.
4. Karakterskalaen kan misforstås som et signal om at miljøer som ikke får topp karakter har en eller annen form for defekt. Selv om alle miljøer gjerne kan ha ambisjoner om å bli bedre, gir det ikke mening å forvente at alle skal være internasjonalt ledende. Eksempelvis kan det å være basis for god utdanning være det viktigste formålet for forskning ved mange læresteder og kontakt med praksisfeltet regionalt og nasjonalt kan være like viktig som kontakt med det internasjonale forskerfellesskapet for å oppnå god utdanningskvalitet.

Relevans for utdanning

Det har vært en vanlig kritikk mot fagevalueringene at de gir et skjevt blikk på institusjonens virksomhet ved bare å vurdere forskning. Forskningen ved UH-institusjonene er tett knyttet til utdanningsoppdraget. Når institusjonene skal følge opp en fagevaluering må det derfor gjøres en vurdering av hvordan alternative oppfølgingstiltak og prioriteringer innen forskning vil påvirke utdanningsvirksomheten.

UH-institusjonenes utdanningsoppdrag er tatt inn som en kontekst for forskning i de seneste fagevalueringene. Konkret er dette gjort ved å inkludere spørsmål om forskningens relevans for utdanning i institusjonenes egne evalueringer, sammen med en vurdering av hvordan utdanningsoppdraget påvirker dimensjonering av forskningsaktiviteter. For evalueringen av samfunnsvitenskap (SAMEVAL) ble det utarbeidet en egen rapport som så nærmere på spillet mellom forskning og utdanning (KOMBEVAL). Denne rapporten var et samarbeid mellom NOKUT og Forskningsrådet og kom i tillegg til Forskningsrådets ordinære fagevaluering av samfunnsvitenskapelig forskning. utfordringene med å kombinere forskning og utdanning i en felles evaluering vil bli drøftet nærmere i kapitlene 8 og 9

Samfunnsbidrag

Både UH-sektoren og instituttsektoren har som en sentral oppgave å formidle kunnskap til samfunnet. Denne formidlingen skjer ofte i andre kanaler enn gjennom de fagfellelvurderte publikasjonene som er grunnlaget for vurdering av forskning i fagevalueringene. Forskningens relevans for samfunnet er kommet inn som et nytt element i de seneste evalueringene. Det er imidlertid flere utfordringer med å dokumentere og vurdere forskningens samfunns effekter: Forskningsprosessen er grunnleggende uforutsigbar, og veien fra forskningsresultater til effekter på samfunnsnivå kan være lang og kronglete. Videre er det krevende å etablere årsakssammenhenger mellom en bestemt forskningsinnsats og bestemte effekter, fordi komplementære forhold som ligger utenfor forskningsinstitusjonenes kontroll har stor betydning. Når forskning fører til endringer på samfunnsnivå slik som nytt lovverk, nye næringsvirksomhet, bedre helse, eller mer generelt at folks holdninger og handlinger påvirkes, er det ofte et resultat av mangfoldige prosesser med involvering av aktører innenfor og utenfor forskningssystemet.

For å favne et slikt mangfold er det behov for fleksible metoder for datainnhenting og vurdering. Forskningsrådet har høstet erfaring med en case-basert metode som ble utviklet i den britiske universitetsevalueringen Research Excellence Framework (REF 2014). Fordelen med denne metoden er at den kan favne et mangfold av ulike samfunnseffekter fra ulike typer av forskning. Metoden er likevel kritisert for implisitt å forutsette en lineær direkte sammenheng mellom forskningsresultater og endringer på samfunnsnivå, og for å legge vekt på de ekstraordinære effektene istedenfor forskningens 'normale' interaksjon med samfunnet.¹⁸ Fordi alle metoder har svakheter, vil det framover være viktig å utvikle flere supplerende metoder for å dokumentere og vurdere samfunnseffekter. En grunnleggende forutsetning for å kunne evaluere samfunnseffekter er at de forskningsutførende institusjonene forholder seg aktivt til interessenter utenfor academia og selv jevnlig innhenter dokumentasjon om hvordan deres forskning har påvirket samfunnet. På dette området framstår systemer og rutiner for dokumentasjon langt mindre utviklet for forskning enn for utdanning.

Formidling og oppfølging av fagevalueringene

KD har tidligere gitt Forskningsrådet i oppgave å organisere oppfølgingen av evalueringenes anbefalinger, og dette har etter hvert etablert seg som en fast praksis. Selv om institusjonen selv forventes å ta ansvar for oppfølging i eget hus, har Forskningsrådet etter de fleste av evalueringene tatt initiativ til en oppfølging i nasjonal regi. Dette er delvis gjort gjennom å etablere nasjonale oppfølgingsutvalg – som utarbeider fagplaner på nasjonalt nivå og gir innspill til oppfølging i Forskningsrådet – og delvis ved at Forskningsrådet har satt av egne strategiske midler til nasjonale oppfølgings tiltak.

Fagevalueringene har vært benyttet i Forskningsrådets rådgivning overfor departementene og har videre bidratt til kunnskapsgrunnlaget for utvikling av Forskningsrådets virkemidler. Et eksempel på dette er etableringen av SFF-virkemiddelet som svarte på utfordringen med små miljøer og fragmentering dokumentert i den første runden med fagevalueringer (se Brofoss 2004). Utvikling av søknadstypen Unge forskertalenter i FRIPRO, som svarte på karriereutfordringer som ble påpekt i flere evalueringer i andre runde, er et annet eksempel.

I 2012-14 gjennomførte Forskningsrådet og UHR et felles prosjekt med sikte på å styrke arbeidet med fagutvikling på nasjonalt nivå, og ved UH-institusjonene.¹⁹ Prosjektgruppa fant at Forskningsrådets fagevalueringer jevnt over har høy legitimitet i fagmiljøene og at de gir et godt og nyttig kunnskapsgrunnlag for institusjonenes strategiske arbeid. Fagevalueringene skaper generelt en større bevissthet i fagmiljøene om deres relative plassering i en nasjonal og internasjonal kontekst med hensyn til publiseringsprofiler, publiseringspoeng, forskningsmidler og forskningsresultater generelt. Fagevalueringene gir viktige bidrag til institusjonenes arbeid med å styrke kvalitet og relevans av deres forskningsvirksomhet. Fagevalueringene kan for eksempel identifisere drivere for forskningskvalitet som er relevante for alle institusjoner, slik som kritisk masse i forskningsgrupper, bedre infrastruktur, tverrfaglighet, internasjonalisering, mobilitet og tydelige karriereløp.

¹⁸ G. Sivertsen & I. Meijer (2019) "Normal versus extraordinary societal impact: how to understand, evaluate, and improve research activities in their relations to society?" *Research Evaluation*, Vol. 29/1, s. 66–70.

¹⁹ UHR/Forskningsrådet (2014) *Gode råd – Hva så?* (internt dokument).

Rapporten påpeker videre noen utfordringer knyttet til oppfølging: Når fagevalueringer skal følges opp ved institusjonene må anbefalinger og tiltak som er utformet på et nasjonalt nivå innordnes i institusjonenes strategiske planer. En særlig utfordring er å finne balansen mellom top-down og bottom-up styring når det skal settes strategiske mål for virksomheten. Fagmiljøene vil ofte bruke gode vurderinger og anbefalinger i fagevalueringene som utgangspunkt for forslag om konkrete satsninger og prioriteringer til ledelsesnivået ved institusjonene. Samtidig må institusjonsledelsen prioritere tiltak innenfor fakultetets budsjettramme og ta hensyn til de strategiske mål som er satt for den aktuelle enheten og institusjonen som helhet. Også når det gjelder Forskningsrådets utlysninger av nasjonale midler til oppfølging av fagevalueringer, er det institusjonens ansvar å sørge for at de prosjektene som søkes finansiert er de som best støtter opp om de strategiske målene som er satt for faget ved institusjonen.

Tilbakemeldinger tyder på at flere av forskningsinstitusjonene har benyttet evalueringene aktivt i sitt strategiske arbeid. Nyten for institusjonene kan imidlertid avhenge av timingen av evalueringene i forhold til institusjonenes strategiarbeid. Tilsvarende peker Probas kartlegging av erfaringer med evalueringer i sektoren at lengre planleggingshorisont og bedre forutsigbarhet for de nasjonale evalueringene vil gjøre det enklere for institusjonene å planlegge egne evalueringer og undersøkelser slikat de unngår dobbeltarbeid.

Utfordringer med nåværende modell

Norske UH-institusjoner har stor autonomi i hvordan de benytter sin grunnbevilgning. Samtidig er det en politisk forventning om at hver institusjon utvikler en faglige profil basert på komparative fortrinn og gjør tydelige strategiske prioriteringer. Derfor er det interessant å se på hvordan evalueringer av forskning kan bidra til en sterkere strategisk styring i sektoren. Liknende forventninger retter seg mot forskningsinstitutter som mottar statlig basisbevilgning gjennom Forskningsrådet.

Å styrke den vitenskapelige kvaliteten, er et sentralt mål for offentlige bevilgninger til forskning, men forskning kan også ses som et middel for å oppnå andre mål. Derfor blir det i tillegg viktig å spørre hvordan den vitenskapelige kvaliteten bidrar til måloppnåelse innen utdanning, formidling og innovasjon for UH-sektoren på den ene siden, og instituttsektorens evne til å levere oppdatert relevant kompetanse til sine oppdragsgivere i offentlig og privat sektor på en andre. Å vurdere forskningens kvalitet utelukkende ut fra rene forskningsinterne standarder blir for snevert for å kunne si noe om hvorvidt mottakerne av offentlige grunnbevilgninger til forskning gjør god bruk av disse midlene. Det kan hevdes at et evalueringsregime som skal bidra til bedre strategisk styring derfor må vurdere forskningsaktiviteten, og resultatene av denne, i lys av mål som er satt for denne aktiviteten ved institusjonene og nasjonalt.

I den siste runden med fagevalueringer har Forskningsrådet ønsket å gi institusjonene mulighet til å bruke evalueringsresultatene mer strategisk gjennom å evaluere hele fagområder som samsvarer bedre med et strategisk nivå ved institusjonen, eksempelvis et fakultet eller en avdeling. Inkluderingen av utdannings- og samfunnsoppdraget har videre gitt ledere innen UH-institusjonene mulighet til å se flere deler av sin virksomhet i sammenheng. I tillegg til utfordringen med å gjøre evalueringene mer relevant for den strategiske styringen ved den

enkelte institusjon, er det en stadig større utfordring å evaluere forskning på nasjonalt nivå etter en inndeling i fagdisipliner. Så lenge den totale forskningsvirksomheten i landet ikke evalueres samlet og samtidig (slik man for eksempel gjør i Storbritannia) vil en inndeling i fag eller fagområder ikke gi et komplett bilde av alle relevante institusjoner. Særlig vil tverrfaglige miljøer og tematisk organiserte institusjoner – slik som forskningsinstituttene – kunne oppleve at disiplinorienterte evalueringer ikke er godt tilpasset deres virksomhet. Grad av tverrfaglighet varierer mellom fagene, men vi observerer i dag en generell trend mot mer tverrfaglig eller flerfaglig forskningsorganisering.

For å styrke den strategiske betydningen av fagevalueringene for den enkelte institusjon, har Forskningsrådet startet utprøving av en ny institusjonsbasert evalueringsmodell hvor vurderinger av forskningsvirksomheten i større rad knyttes opp mot institusjonenes strategiske mål og faglige innretning. Den nye evalueringsmodellen er definert gjennom en evalueringsprotokoll som gir føringer for evaluering av den enkelte institusjon når det gjelder evalueringstema (forskningens kvalitet og relevans for utdanning og samfunn) og datainnsamling. Disse føringene er i stor grad samsvarende med tidligere fagevalueringer. Tilpasningen til strategiske mål og faglige innretning skjer gjennom to grep: 1) Forskningsrådet inviterer de relevante institusjonene til å melde inn relevante evalueringsenheter. Disse enhetene skal være reelle organisatorisk enhet (grupper, senter, institutter eller liknende). 2) Institusjonene definerer et referansepunkt (benchmark) som hver enhet skal måles opp mot og kan beskrive strategiske mål som er relevante for enheten. Deretter gjennomfører institusjonene en egevaluering etter felles mal, men med mulighet for å supplere med data som kan brukes til å vurdere måloppnåelse for evalueringsenheter. En slik institusjonstilpasset modell gir blant annet mulighet til å vurdere tverrfaglige miljøer samlet og basert på formålet med den spesifikke organiseringen av forskningen.

I en pilot for rettsvitenskapelige fag (JUREVAL) som nå er under gjennomføring, har Forskningsrådet oppnevnt evalueringskomiteen og sørget for at det utarbeides felles kvantitative analyser av personale og publikasjoner. Formålet med dette er å sikre mulighet for sammenlikning mellom institusjoner og vurdering av den samlede forskningsvirksomheten nasjonalt.

7. Nærmere om evaluering av kvalitet i høyere utdanning

Formålet med NOKUTs evalueringer

I mange organisasjoner gjennomføres evalueringer med formål om etterrettelighet, kontroll og ansvarliggjøring. I NOKUT ivaretas kontrollfunksjonen gjennom tilsyn, og NOKUTs evalueringer er derfor primært et «mykt» virkemiddel for å stimulere til kvalitetsutvikling i sektoren, gjennom å vurdere den faktiske kvaliteten på utdanningene. Evalueringene gir institusjonene, Kunnskapsdepartementet og andre interessenter kunnskap om kvalitetstilstanden og gir samtidig institusjonene anbefalinger om hva disse kan eller bør gjøre for å videreføre eller heve kvaliteten på utdanningene.

At NOKUT skal gjennomføre evalueringer av høyere utdanning reguleres både i Lov om universiteter og høyskoler og i studiekvalitetsforskriften. I Lov om universitets- og høyskoleloven § 2-1 (4) står det:

NOKUT skal gjennomføre evalueringer av betydning for å kunne bedømme kvaliteten i høyere utdanning. Departementet kan pålegge NOKUT å foreta slike evalueringer. Alle evalueringer som foretas av NOKUT, er offentlige, og NOKUT skal bidra til at disse gjøres kjent.

I studiekvalitetsforskriften § 1-5 (2), om NOKUTs oppgaver innenfor høyere utdanning, er det også oppført som eget punkt at NOKUT skal «[e]valuere for å bedømme kvaliteten i høyere utdanning».

Evaluering vs. tilsyn

Når NOKUT gjennomfører tilsyn og evalueringer er det en rekke likhetstrekk i måten disse blir utført på. Både tilsyn og evalueringer baseres på bruk av eksterne sakkyndige i vurderingene, tydelige evalueringskriterier eller forskriftskrav, og framgangsmåter som skal sørge for pålitelige vurderinger og sikre likebehandling av evaluerings- og tilsynsobjektene. Det er likevel noen sentrale forskjeller, i all hovedsak knyttet til konsekvenser og oppfølging i etterkant. Som nevnt er NOKUTs kontrollfunksjon ivaretatt gjennom ulike tilsyn rettet mot studietilbud, institusjoner eller andre objekter. Kravene som ligger til grunn for disse tilsynene er regulert i UH-loven, studiekvalitetsforskriften og studietilsynsforskriften. Her har NOKUT sanksjonsmuligheter dersom et tilsyn avdekker feil eller mangler. NOKUT kan kreve at tilsynsobjektet retter opp mangler innenfor gitte frister, og i ytterste konsekvens kan akkrediteringsrettigheter trekkes tilbake.

NOKUTs evalueringer tar ikke utgangspunkt i lovpålagte krav til institusjonene, men vurderer grader av kvalitet og/eller måloppnåelse ut fra evalueringskriterier som blir utviklet og tilpasset den enkelte evaluering. På denne måten beveger NOKUTs evalueringer seg ut over minstekrav i lover og forskrifter, og et viktig formål med evalueringene er å bidra til videre kvalitetsutvikling. Dette gjøres gjennom å skaffe til veie kunnskap og legge til rette for læring, som igjen kan lede til aktiviteter, programmer og tiltak. En konsekvens av at evalueringene beveger seg ut over minstekrav, er at NOKUT ikke kan kreve at svakheter forbedres. For at evalueringen skal nå det ønskede formålet, er NOKUT derfor avhengig av at evalueringsobjektet, eller andre interessenter, bruker evalueringen i sitt utviklingsarbeid. Dette fordrer at interessenter oppfatter evalueringen som relevant og av høy kvalitet.

Kontekst og historikk

NOKUT ble etablert i 2002, i forbindelse med Kvalitetsreformen. I de første årene var det et fokus på akkreditering av nye studieprogrammer og etterhvert tilsyn med etablerte studieprogrammer. NOKUTs første evaluering kom i 2006, og var en evaluering av alle allmennlærerutdanningene på oppdrag fra Kunnskapsdepartementet. Dette var en omfattende evaluering der man vurderte mange ulike temaer og brukte store ressurser på datainnsamling. Det ble også gjennomført institusjonsbesøk ved alle de 20 institusjonene som deltok i evalueringen. I 2008 gjennomførte NOKUT en lignende evaluering av alle to- og treårige ingeniørutdanninger, og i 2010 en evaluering av alle førskolelærerutdanninger. Alle tre evalueringene ble gjennomført på oppdrag fra og finansiert av Kunnskapsdepartementet. Hovedformålet med disse evalueringene var å skaffe til veie et kunnskapsgrunnlag om kvaliteten i utdanningene, samt bidra til at institusjonene kunne bruke resultatene av evalueringene videre i sitt eget systematiske kvalitetsarbeid.

Mellom 2010 og 2014 fikk ikke NOKUT flere evalueringsoppdrag fra departementet og gjennomførte heller ingen evalueringer på eget initiativ. I 2014 fikk NOKUT to nye evalueringsoppdrag fra Kunnskapsdepartementet. Det ene var å samarbeide med Forskningsrådet om en fagevaluering hvor både utdanning og forskning inngikk (Kombinerte fagevalueringer), og den andre var en evaluering hvor NOKUT skulle vurdere kvaliteten på noen norske studieprogram og sammenligne disse med kvaliteten på lignende utenlandske studieprogram (EUROMA).²⁰

I tillegg til de evalueringene som er nevnt over, har NOKUT gjennomført en evaluering av arbeidslivsrelevans i disiplinlag i 2020, og startet opp en evaluering av lektorutdanningene som skal avsluttes i 2020. I begge disse evalueringen er fokuset på læring og utvikling, mer enn på summative vurderinger og ansvarliggjøring.

Også andre aktører har gjennomført evalueringer av utdanning i Norge det siste tiåret. Et eksempel på dette er følgegruppen for lærerutdanningsreformen.²¹ Denne gruppen ble nedsatt av Kunnskapsdepartementet i 2010 og hadde som formål å følge innføringen av de nye grunnskolelærerutdanningene og bidra til å sikre at reformen ble gjennomført etter intensjonene. Et annet eksempel er Nordisk institutt for studier, innovasjon, forskning og utdanning (NIFU) sin evaluering av ph.d. utdanningene i Norge fra 2012.²² Evalueringen ble gjennomført på oppdrag av Kunnskapsdepartementet og Forskningsrådet.

Evalueringsstema og enheter

Som nevnt innledningsvis er NOKUTs evalueringsvirksomhet hjemlet i UH-loven og studietilsynsforskriften, der det står at evalueringene skal bedømme kvaliteten i høyere utdanning. Disse evalueringene, og vurderingene som ligger til grunn, har som primærformål å skaffe til veie kunnskap og legge til rette for læring og utvikling av utdanningskvaliteten i norsk høyere utdanning.

²⁰ Se <https://www.nokut.no/prosjekter-i-nokut/kombinerte-fagevalueringar/> for mer om de kombinerte fagevalueringene og <https://www.nokut.no/prosjekter-i-nokut/euroma--benchmarking-of-master-programmes-across-european-countries/> for mer om den europeiske sammenligningen.

²¹ Se <https://www.regjeringen.no/no/dokumenter/Rapporter-fra-folegruppen-for-larerutdanningsreformen/id753381/> for mer om følgegruppen for lærerutdanningsreformen.

²² Taran Thune m. fl. (2012) "PhD education in a Knowledge society: An evaluation of PhD education in Norway", NIFU Report 25/12.

Det er studieprogrammet som er den sentrale enheten studenten møter på universitetet, og det er også den sentrale enheten for NOKUTs evalueringer. NOKUT er ikke bundet til å bruke studieprogrammer som enhet, men det er den mest naturlige evalueringsenheten for å evaluere utdanningskvalitet. Et alternativ kan være å bruke fakulteter eller institusjoner som evalueringsenhet, men det er få indikatorer, eller kvalitetsdimensjoner, på fakultets- eller institusjonsnivå som kan benyttes for å gi et svar på om utdanningskvaliteten er god. NOKUT har gode erfaringer med institusjonsovergripende tilsyn som vurderer institusjonenes systematiske kvalitetsarbeid. Disse tilsynene vurderer institusjonenes systemer for internkontroll, og gir en pekepinn på hvorvidt institusjonen forvalter sine fullmakter på en god måte. Tilsynene sier likevel ingenting om den faktiske utdanningskvaliteten ved institusjonen, kun hvorvidt institusjonen har systemer på plass for å fange opp og forbedre sviktende kvalitet. Det kan videre være relevant å undersøke utdanningskvaliteten i samtlige studietilbud innenfor et fakultet, men resultatene av en slik vurdering vil ikke ha nytteverdi ut over dette fakultetet. Ved å konsentrere utdanningsevalueringene mot studieprogram, sikrer man at man kommer så tett på den faktiske utdanningskvaliteten som mulig.

Data og metoder

Utdanningskvalitet er et vidt og mangefasettert begrep og det er ikke etablert noe omforent sett av indikatorer som fullt ut kan sies å være dekkende for begrepet. I forskning og policy er det normalt å skille mellom innsatsfaktorer, prosessfaktorer og resultatfaktorer. Typiske innsatsfaktorer er startkompetansen til studentene, kompetansen og størrelsen på fagmiljøet som er ansvarlige for utdanningen. Typiske resultatfaktorer er gjennomstrømming og karakterer. Prosessfaktorer er det studentene møter gjennom studiet sitt. Alt fra programdesign, undervisning – og vurderingsformer, pensum etc. Prosessfaktorene er det mest sentrale når det kommer til NOKUTs definisjon av kvalitet.²³ Det er få, om noen kvantitative indikatorer som kan si noe om læringsprosessene. Evalueringer av utdanningskvalitet krever derfor innsamling av mye data som ikke er lett tilgjengelig. Dette kan være spørreundersøkelser, intervjuer/institusjonsbesøk og ikke minst selvevalueringer med tilhørende dokumentasjon.

Evalueringer av utdanningskvalitet krever derfor data fra mange kilder, både kvalitative og kvantitative. Dette kan gjøre evalueringene arbeidskrevende både for institusjonene og for de sakkyndige som gjennomfører vurderingene. Ofte er vurderingene basert på informasjon og data som framstilles utelukkende for evalueringen og/eller utilgjengelig informasjon og data som institusjonene tilgjengeliggjør for NOKUT og evalueringspanelene.

Oppfølging

NOKUTs evalueringer av utdanningskvalitet har som hensikt å gi utdanningsprogrammene, NOKUT, Kunnskapsdepartementet og andre relevante interessenter kunnskap om kvaliteten i utdanningene, og å gi de samme aktørene anbefalinger om hva man bør gjøre for å heve kvaliteten ytterligere. Mer indirekte bruker NOKUT evalueringene som et kunnskapsgrunnlag i sine vurderinger om framtidige tilsyn. Til forskjell fra Forskningsrådet, sitter ikke NOKUT med økonomiske virkemidler som kan brukes i oppfølgingen av evalueringer, f.eks., til påskjønnelse av studieprogrammer med særlig god kvalitet eller som stimuli til studieprogrammer som sakker akterut innenfor ulike kvalitetsdimensjoner. Per i dag ligger

²³ Se

https://www.nokut.no/contentassets/5979c996834c47f4a296269de44436b0/kvalitetsomrader_for_studieprogram.pdf

denne type virkemidler hos Direktoratet for internasjonalisering og kvalitetsutvikling i høyere utdanning (DIKU). Det er derfor sentralt at det er god kommunikasjon mellom NOKUT og DIKU, slik at DIKU, i tillegg til annet kunnskapsgrunnlag, kan bruke kunnskapsgrunnlaget fra evalueringene i sine kvalitetsprogrammer.

NOKUT har til nå ikke hatt noen systematisk oppfølging av evalueringene. NOKUT følger selv opp anbefalinger som kommer til NOKUT, mens det forventes at institusjonene tar ansvar for de råd og anbefalinger som evalueringene gir. Der evalueringene frambringer anbefalinger til Kunnskapsdepartementet formidler NOKUT dette videre. Det er altså de aktørene som er involvert i evalueringene, i all hovedsak evalueringsobjektene selv, som er ansvarlige for oppfølgingen.

Utfordringer med nåværende modell

I Norge har vi tilgang til mye data knyttet til høyere utdanning. Vi har en nasjonal studentundersøkelse, og institusjonene rapporterer på en rekke indikatorer til Database for høyere utdanning. Utfordringen med disse dataene er at de i hovedsak knytter seg til innsats og resultatfaktorer og ikke prosessfaktorer, som best framskaffes gjennom evalueringer. Institusjonenes systematiske kvalitetsarbeid bidrar til at institusjonene selv får et bilde av den faktiske kvalitetstilstanden på sine studieprogrammer. Det er allikevel slik at institusjonene i sitt kvalitetsarbeid vurderer forskjellige kvalitetsdimensjoner og fokuserer til en viss grad på forskjellige indikatorer. Dette er naturlig og ønskelig. Samtidig betyr dette at vi mangler et nasjonalt komparativt bilde på den faktiske utdanningskvaliteten studentene mottar, og som institusjonene kan bruke for å vurdere kvaliteten på sine utdanninger opp imot andre lignende utdanninger.

Den største utfordringen med måten utdanningsevalueringer gjennomføres på, er at det ikke finnes noen helhetlig modell. NOKUT har gjennomført en håndfull evalueringer, men det har ikke vært noen systematisk, helhetlig tenkning som ligger bak hvem eller hva som er blitt evaluert, eller hvem som har tatt initiativet til de ulike evalueringene. Kunnskapsdepartementet har gitt både NOKUT og andre aktører forskjellige typer oppdrag knyttet til evaluering av utdanningsvirksomheten. Andre evalueringer har NOKUT satt i gang selv med utgangspunkt i NOKUTs strategi og tematiske satsninger.

Konsekvensen av dette er at institusjonene selv, myndighetene og samfunnet generelt mangler kunnskap om hvor høy kvaliteten på norsk høyere utdanning faktisk er. Et godt kunnskapsgrunnlag om kvalitetstilstanden vil kunne utgjøre et viktig styringsverktøy for både departement og underliggende etater som NOKUT og DIKU. For institusjonene vil sammenlignbare vurderinger av kvalitetstilstanden muliggjøre diskusjoner om spesialisering og differensiering innenfor fagområder. For studentene vil mer kunnskap om hvor kvaliteten på utdanningen er god, kunne være med på å vekte valg av studiested, selvsagt sammen med andre faktorer.

Erfaringene tyder altså på at det er behov for et rammeverk som kan sikre periodiske evalueringer for å bedømme den faktiske kvaliteten på studieprogrammer i norsk høyere utdanning. Et slikt rammeverk må sikre at alle utdanningstilbud blir evaluert av fagfeller med kompetanse på det spesifikke fagfeltet. Det er også sentralt at slike evalueringer er komparative og bygger på et datagrunnlag som også tar for seg viktige prosessfaktorer som eksempelvis hvordan studieprogrammenes oppbygging og sammenheng sikrer studentene et godt læringsutbytte.

8. Koblinger og koordinering

Sammenlignet med en del andre land har det vært relativt lite systematisk koordinering av evalueringsvirksomheten i Norge. Som beskrevet i kapittel 6 har Forskningsrådet gjort sine egne løpende vurderinger av hvilke fagområder som skulle prioriteres, og i hvilken rekkefølge. Over tid har det etablert seg et syklisk mønster der de fleste fagområder er evaluert i to omganger de siste tiårene. For utdanningsevaluering er historikken betydelig kortere, og det har foreløpig bare vært kapasitet til å prioritere et utvalg studieprogrammer, hovedsakelig profesjonsutdanninger, som beskrevet i kapittel 7. Flere av evalueringene er blitt initiert av KD, noe som også har hatt konsekvenser for NOKUTs egne planer. Både Forskningsrådet og NOKUT har varslet berørte parter om planlagte evalueringer, og for forskningsevalueringene har det også i en periode vært utarbeidet en kalender på Forskningsrådets nettsider. Men planleggingshorisonten har vært kort.

Fram til 2018 var det relativt liten grad av kobling mellom evaluering av forskning og høyere utdanning. I en del av forskningsevalueringene ble det gjort visse vurderinger av utdanningssiden, men generelt har det vært omtalt som en utfordring, særlig for forskning som er koblet til profesjonsstudier, at dette ble gjort i relativt liten grad. NOKUTs utdanningsevalueringer har i noen grad omtalt forskningen som kunnskapsgrunnlag for studieprogrammene, og også gitt vurderinger av lærekreftenes kompetanse (herunder forskerkompetanse), men ikke gitt noen selvstendig vurdering av forskningskvaliteten. Forskningsrådet har i sine seneste evalueringer inkludert utdanningsvirksomheten som en viktig kontekst for forskning, men uten vurdering av utdanningskvaliteten i seg selv. I enkelte tilfeller har det vært en viss sammenheng mellom Forskningsrådets evalueringer av forskning og NOKUTs utdanningsevalueringer, slik det i noen grad var mellom evalueringen av pedagogisk forskning i 2004 og evalueringen av allmennlærerutdanningen i 2006, men stort sett har Forskningsrådet og NOKUT planlagt og gjennomført sine evalueringer hver for seg.

I 2014 fikk imidlertid Forskningsrådet og NOKUT i oppdrag av KD å samarbeide om en felles evaluering av samfunnsfagene. Bakgrunnen var et økt fokus på samspillet mellom utdanning og forskning og forskningsbasert utdanning, som senere også kom til uttrykk i Meld. St. 18 (2014-2015) *Konsentrasjon for kvalitet – Strukturreformen i universitets- og høyskolesektoren* og Meld. St. 16 (2016-2017) *Kultur for kvalitet i høyere utdanning*. Som blant annet Probas analyse viser, var det utfordrende å lage et felles design for evaluering av forskning og utdanning. Dette skyldtes flere forhold, men kanskje særlig at kvalitetsbegrepet, relevante indikatorer og datakilder og organiseringen er ulik for henholdsvis forskning og utdanning. Løsningen ble å evaluere forskning og utdanning hver for seg, og så gjennomføre en separat studie av sammenhenger.

Koblingen mellom forskning og høyere utdanning

Siden det er et vesenstrekk ved høyere utdanning at den skal være forskningsbasert (med noen unntak), må evalueringsvirksomheten på en eller annen måte fange opp denne koblingen. Som nevnt ovenfor har både NOKUTs utdanningsevalueringer og Forskningsrådets forskningsevalueringer i noen utstrekning gjort dette. Etter erfaringene fra SAMEVAL/KOMBEVAL (se ovenfor) er spørsmålet om man burde gå enda lengre i den retningen, og koordinere (eller integrere) evalueringsvirksomheten i NOKUT og

Forskningsrådet med det formål å gjøre koblingen av forskning og utdanning til et sentralt spørsmål i alle (eller de fleste) evalueringsløp.

Internasjonal forskning har riktignok hatt vanskelig for å påvise noen entydige generelle effekter av at høyere utdanning er forskningsbasert.²⁴ På den annen side kan det hevdes at nettopp derfor ville det være av interesse å få belyst hvordan denne koblingen arter seg i fagene og studieprogrammene på ulike norske institusjoner, og få analysert hvordan dette påvirker kvaliteten. Videre er det ikke bare forskningens påvirkning på utdanningen som er interessant; utdanningsvirksomheten virker også inn på forskningen. Undervisningen og studentenes læringsarbeid er med på å forme forskningen, og i tillegg formes forskningen på en grunnleggende måte ved at studentene utgjør forskningens rekrutteringsgrunnlag. Hvordan ulike fagmiljøer håndterer dette vil det være interessant å evaluere.

Et annet moment som taler for evalueringer med fokus på koblingen mellom forskning og utdanning, er at det i mange fag er avgjørende med en analyse av denne sammenhengen for å kunne følge opp en evaluering på en god måte. I noen profesjonsfag, for eksempel, kan forbedring av forskningskvaliteten i ulike deler av faget være så tett forbundet med utviklingen av det tilknyttede studieprogrammet at hver av delene vanskelig lar seg behandle isolert. Et siste moment er at begge typene evalueringer er avhengige av å innhente betydelig med informasjon om fagmiljøer, noe som kunne gjøres under ett for å spare tid og arbeid hos de som evaluerer og de som blir evaluert.

Allikevel er det mange ting som skiller forskning og høyere utdanning. For det første er organiseringen er forskjellig. Forskningen foregår ofte i forskergrupper, mens utdanningen er organisert i studieprogrammer, og den organisatoriske koblingen mellom disse kan variere svært mye. Kvalitetsbegrepet er også til dels svært ulikt mellom forskning og utdanning, som nærmere beskrevet i kapittel 6 og 7. Videre er ekspertkompetansen som evalueringspanelene må ha til dels ulik for henholdsvis forskning og utdanning. Disse utfordringene, og det faktum at mange av aktørene Proba har intervjuet ikke synes å prioritere denne typen analyser, kan tale imot å basere et nasjonalt rammeverk på at høyere utdanning og forskning alltid skal evalueres i samme prosess. Dessuten vil koblede evalueringer av forskning og utdanning være lite relevant for instituttsektoren.

Det kan imidlertid reises spørsmål ved om det finnes visse fag og studieprogrammer hvor koblede evalueringer av forskning og utdanning kan være den mest hensiktsmessige modellen. Det kan for eksempel være profesjonsfag, eller studiefag hvor det er indikasjoner på at forholdet til forskningsbasen byr på utfordringer. Det er imidlertid vanskelig å finne noe grunnlag for å definere på generelt grunnlag akkurat hvilke fag/programmer dette vil gjelde for. Dette må avklares gjennom mer konkret behandling av de som skal ha ansvar for evalueringsvirksomheten.

Koordinering

Spørsmålet om koordinering handler ikke bare om muligheten for å evaluere koblingen mellom forskning og utdanning. Som Probas rapport viser er det mange av de aktørene som har

²⁴ John Hattie og H. W. March (1996) "The Relationship Between Research and Teaching. A Meta-Analysis", *Review of Educational Research* 66.4.

behov for evalueringer som vil få vesentlig større nytte av dem hvis de kan gjøres mer forutsigbare, og hvis det er en gjennomtenkt sammenheng mellom evalueringene av forskning og utdanning i de ulike fagene, også om man ikke integrerer dem. I tillegg vil økt forutsigbarhet redusere arbeidsinnsatsen institusjonene må legge ned i hver evaluering. Ved å vite i god tid når evalueringer skal gjennomføres, kan institusjonene planlegge sine egne prosesser slik at arbeid med evalueringene ikke oppleves som ekstraarbeid.

Oppfølgingen av evalueringene ved institusjonene vil forenkles hvis det er kjent på forhånd når evalueringene kommer, og hvis man kan velge hensiktsmessige tidspunkter for evaluering av henholdsvis forskning og utdanning i samme fag. På den måten vil evalueringene lettere kunne inngå i institusjonenes løpende kvalitetsutviklingsarbeid, og institusjonene vil også lettere kunne vurdere hvilke øvrige former for evalueringer de selv vil ta initiativ til, slik at man unngår dobbeltarbeid.

Også for myndighetsorganer og andre aktører i kunnskapssektoren vil forutsigbarhet gi store fordeler. For KD vil både politikikutvikling, sektorstyring og etatsstyring få et bedre kunnskapsgrunnlag hvis man har forutsigbar og jevnlig evaluering av institusjonenes kjerneaktivitet. For NOKUT og Forskningsrådet vil forutsigbarhet gi begge organisasjoner mulighet til å planlegge bedre og mer langsiktig, noe som vil øke kvaliteten og effektiviteten i arbeidet med evalueringer. For øvrige aktører og observatører vil forutsigbarhet også være fordelaktig, ved at det sikrer jevnlig oppdatert kunnskapsgrunnlag.

Det er med andre ord store gevinster å hente for mange aktører hvis man kan få til bedre koordinering og større forutsigbarhet. Den eneste ulempen man kanskje kunne peke på, ved siden av selve koordineringsarbeidet som må gjøres, er at evalueringer som er kjent i svært god tid på forhånd kan tenkes å bli mer formalistiske, med et preg av "teaching to the test" enn evalueringer som kommer med relativt kort varsel. Risikoen for dette er vanskelig å vurdere, men trolig vil faren for slike strategiske tilpasninger være størst for gjentakende statiske evalueringsløp med faste indikatorer, noe som er mindre aktuelt i Norge. Forutsatt at evalueringsdesignet er gjennomtenkt og tilpasset formålene er det neppe skadelig om institusjonene forbereder seg og ønsker å vise seg fra sin beste side.

Oppsummering

Kartleggingen som Proba har gjennomført viser at mange aktører mener mangel på koordinering og forutsigbarhet er blant de største utfordringene for evalueringsarbeidet i Norge i dag. Imidlertid ser det ikke ut til at det er analyser av sammenhenger mellom forskning og utdanning som savnes mest. Selv om mange nok ville finne slike analyser interessante, er det lite erfaring med slike evalueringer i sektoren, og i Probas oppsummering av respondentenes vurderinger omtales tett kobling eller integrerte evalueringsløp som noe mange mener det vil være svært vanskelig å gjennomføre i praksis. Internasjonalt finnes det heller ikke noen etablert beste praksis for integrerte evalueringen av forskning og utdanning.

Det kan derfor se ut til at det er mest hensiktsmessig å legge vekten på bedre koordinering mellom forsknings- og utdanningsevalueringer, og lengre planleggingshorisont for dem begge. Både selve gjennomføringen, oppfølgingen ved institusjonene, og bruksverdien for andre aktører vil bli vesentlig bedre dersom det kunne etableres prosedyrer som sikret forutsigbare og syklisk gjentakende evalueringer av både forskning og utdanning innen ulike

fagområder. På den måten kunne man også legge til rette for å analysere sammenhenger mellom forskning og utdanning der dette oppleves som relevant.

I kapittel 9 gir arbeidsgruppen en nærmere omtale av hvordan dette kan gjøres i praksis.

9. Mot et nasjonalt rammeverk i Norge

I dette kapitlet vil arbeidsgruppen først sette evalueringene av kvalitet i forskning og høyere utdanning inn i en større kontekst. Deretter vil vi drøfte de viktigste momentene som bør inngå i et nasjonalt rammeverk for evaluering, basert på målsettingene og prinsippene som er behandlet i kapittel 5 og utfordringsbildet som er presentert i kapittel 6, 7 og 8. Avslutningsvis vil ulike modeller for organisering av evalueringsvirksomheten bli drøftet, som bakgrunn for arbeidsgruppens forslag til rammeverk i kapittel 10.

Forholdet til styringssystemet

Som del av den statlige forsknings- og høyere utdanningspolitikken inngår evalueringer av kvalitet i en større kontekst av styringsvirkemidler som samlet sett skal bidra til å nå de politiske målene for norsk forskning og høyere utdanning. For instituttsektoren er politikken og virkemiddelbruken fastlagt i KDs *Strategi for helhetlig instituttpolitikk*. Når det gjelder universitets- og høyskolesektoren skal styringssystemet behandles i en egen stortingsmelding ("styringsmeldingen") som er varslet våren 2021. Arbeidsgruppen har ikke hatt kjennskap til detaljene i styringsmeldingen under utarbeidelsen av denne rapporten. Basert på det som har vært formidlet av departementets politiske ledelse i innspillrunder og i offentligheten, samt enkelte utdypninger av departementets medlemmer i arbeidsgruppen, legger arbeidsgruppen imidlertid følgende premisser til grunn for sine drøftelser og forslag:

- Jevnlige evalueringer av kvaliteten i norsk forskning og høyere utdanning med bred legitimitet blant relevante aktører oppfattes som en viktig forutsetning for en ny styringspolitikk basert på utstrakt institusjonelt selvstyre.
- Et norsk rammeverk for evalueringer av kvalitet er ikke ment å kobles direkte til finansiering, slik for eksempel Storbritannia har gjort med sitt Research Excellence Framework.
- I tråd med mandatet vil evalueringsarbeidet som skal omfattes av rammeverket fokusere på kvaliteten i de faglige kjerneoppgavene forskning og høyere utdanning. Institusjonenes rammebetingelser og deres strategiske mål kan inngå som kontekst for vurdering av kvalitet, men evalueringene tar ikke sikte på å vurdere institusjonene som sådanne.
- Det er et mål for evalueringene som omfattes av rammeverket at de skal utgjøre et viktig kunnskapsgrunnlag for KDs arbeid med instituttpolitikken og etats- og sektorstyringen av UH-sektoren når det gjelder kvaliteten i primæroppgavene forskning og høyere utdanning. Det forutsettes imidlertid at KDs kunnskapsgrunnlag om andre viktige forhold som strategisk måloppnåelse, dimensjonering, organisering, kostnadseffektivitet, osv., i tillegg baserer seg på informasjon og indikatorer som primært kommer fra andre kilder.

Vi vil også her vise noen sentrale synspunkter fra UH-sektorens høringsinnspill til Aune-utvalgets forslag til ny UH-lov og innspillene til Kunnskapsdepartementets arbeid med styringsmeldingen som tematiserer erfaringer fra universitet og høyskoler med relevans for utformingen av nasjonalt system for evaluering:

- Strukturendringene i norsk høyere utdanning har gitt flere institusjoner som har størrelse og kapasitet til å håndtere institusjonell autonomi og ansvar.
- Styringssystemet må legge til rette for at institusjonene kan utvikle ulike profiler i tråd med politiske forventninger.

- Summen av styringsvirkemidler, retningslinjer og rapporteringskrav bør ikke øke, men snarere reduseres.
- Ansvars- og oppgavefordeling i styring av sektoren bør gjennomgås både vertikalt, mellom departement direktorat og institusjoner, og horisontalt på direktoratsnivå. Som en konsekvens av betydelige endringer i institusjonslandskapet, med færre og større institusjoner, mener UHR det kan det være rimelig å vurdere på nytt hvordan vi kan få til en god balanse mellom styring og institusjonenes autonomi, eller mellom behovet for autonomi opp mot behovet for nasjonal samordning.²⁵

Arbeidsgruppen har hatt en løpende diskusjon om både signalene fra nasjonale myndigheter og innspillene fra aktørene i sektoren i arbeidet med dette forslaget til nasjonalt rammeverk for evaluering av forskning og høyere utdanning. Vi vil komme tilbake til hvordan vi kan ta hensyn til institusjonenes erfaringer og behov i drøftingen av alternative evalueringsmodeller mot slutten av kapitlet.

Forholdet til finansieringssystemet

En viktig del av konteksten for evalueringsarbeidet er den indikatorbaserte delen av finansieringssystemet, som er basert på informasjon som til en viss grad tangerer datagrunnlaget for evalueringer. Tabellen nedenfor gir en oversikt over de indikatorene som inngår i hhv. UoH-sektorens, instituttsektorens og helseforetakenes resultatbaserte finansieringssystemer per 2020:

Tabell 2: Indikatorer som inngår i resultatbasert finansiering per sektor. 2020

Indikator	Sektor			Kommentar
	UH-sektoren	Institutt-sektoren	Helse-foretakene	
Publiserte artikler/publiseringspoeng	X	X	X	
Forskningsrådet-midler (+ RFF)	X		X	
EU-finansiering/Internasjonal finansiering	X	X	X	
Avlagte doktorgrader	X	X	X	
Oppdragsinntekter	X	X		For UoH også bidragsinntekter ekskl. Forskningsrådet/RFF
Studiepoeng	X			
Kandidater	X			
Utvekslingsstudenter	X			
Utreisende studenter på ERASMUS +	X			
Kliniske behandlingsstudier			X	Fra 2020

Kilder: Kunnskapsdepartementet, Helse- og omsorgsdepartementet, Norges forskningsråd

²⁵ https://www.regjeringen.no/contentassets/57997e33f4e74b2dbf917fae83ec4f04/universitets--og-hogskoleradet.pdf?uid=Universitets-_og_h%C3%B8gskoler%C3%A5det

Et sentralt aspekt ved indikatorene ovenfor er at framgang/tilbakegang på de områdene som måles, gir direkte utslag i finansiering. Det betyr for det første at indikatorene reflekterer sentrale politiske mål for sektorene. For det andre gjør de direkte økonomiske konsekvensene at indikatorene må tilfredsstille krav til enhetlig rapportering, transparens og kvalitetssikring. Evalueringene som vil omfattes av det nasjonale rammeverket har imidlertid en annen karakter, og går langt utover formelbasert bruk av indikatorer. Allikevel må disse delene av styringssystemet ses i sammenheng, slik at ikke evalueringene av kvalitet bare gjentar den vurderingen som institusjonene allerede får gjennom finansieringssystemet.

Forholdet til andre former for evaluering

I tillegg til evalueringer av kvaliteten vil det også være behov for andre former for evaluering i de sektorene hvor det drives forskning og høyere utdanning. Av etablerte evalueringstyper bør følgende nevnes særskilt:

- Forskningsrådets instituttevalueringer
- Forskningsrådets evalueringer av programmer og andre virkemidler, noen ganger initiert av finansierende departement
- KDs og andre departementers evalueringer av ulike reformer, tiltak og enkeltinstitusjoner
- Institusjonenes egne evalueringer, ofte knyttet til utvikling og oppfølging av egne strategier²⁶

Arbeidsgruppen ser det ikke som sitt mandat å foreslå rammer for disse andre typene evalueringer. Siden "evaluering" ofte brukes som en generisk betegnelse på mange ulike former for systematiske vurderinger, initiert og utført av mange ulike aktører, er det også vanskelig å se for seg et felles nasjonalt rammeverk for evalueringer av alle typer.

For å redusere rapporteringstrykket mest mulig bør det imidlertid vurderes hvor det eventuelt er overlapp. Arbeidsgruppen har ikke hatt anledning til å gå i detaljer for de evalueringstypene som ikke er omfattet av mandatet, men anbefaler at Forskningsrådet vurderer om, og eventuelt hvordan, data og resultater fra fagevalueringene kan gjenbrukes i øvrige evalueringsvirksomhet, eller vice versa, slik at dobbeltevaluering unngås. Tilsvarende bør NOKUT vurdere om etableringen av et nytt rammeverk gir mulighet for å organisere NOKUTs øvrige prosesser slik at rapporteringstrykket går ned. KD oppfordres også til å tilrettelegge sine evalueringer av reformer og tiltak slik at de i størst mulig grad kan bygge på de evalueringene som omfattes av det nasjonale rammeverket. Det er viktig at kontroll- og rapporteringstrykket ikke øker og at evalueringssystemet ikke trekker i en annen retning enn det som blir lagt til grunn i styringssystemet for øvrig..

Om kvalitetsbegrepet

Arbeidsgruppens mandat er avgrenset til evalueringer som fokuserer på kvalitet i forskning og høyere utdanning, til forskjell fra evalueringer som for eksempel er iverksatt for å vurdere måloppnåelse for tiltak i forskningspolitikken eller politiske reformer innen høyere utdanning. "Kvalitet" er imidlertid et komplekst begrep, og det finnes omfattende forskningslitteratur som diskuterer ulike sider ved kvalitetsbegrepet i henholdsvis forskning og høyere

²⁶ For en mer uttømmende liste, se typologien over evalueringer i Proba 2020.

utdanning.²⁷ Arbeidsgruppen ser det ikke som sin oppgave å gi spesifikke føringer for hvilket kvalitetsbegrep som skal anvendes i de evalueringene som skal omfattes av rammeverket for disse evalueringene. Dette mener gruppen er et faglig spørsmål som for en stor del bør overlates til de organene som skal ha ansvar for å gjennomføre evalueringene, i samråd med de panelene som skal gjøre selve de faglige vurderingene.

Noen overordnede merknader kan det likevel være naturlig å gi:

- Evalueringene bør ta utgangspunkt i veletablerte og anerkjente akademiske kvalitetskriterier, for å sikre bred legitimitet og tillit til forskning og høyere utdanning i samfunnet²⁸
- Evalueringene bør ikke bli for endimensjonale, men inkludere flere ulike kvalitetsdimensjoner
- Kvalitetskriteriene bør formuleres og operasjonaliseres slik at de fanger opp og tar høyde for særtrekkene ved det eller de fagområdene som evalueres, og den konteksten som den faglige virksomheten foregår i (eksempelvis profesjonsutdanning, oppdragsforskning eller kunstnerisk utviklingsarbeid)
- Det er grunn til å minne om at de evalueringene som omfattes av rammeverket primært skal vurdere kvaliteten på selve forskningen og utdanningen, og kan ikke omfatte alle sider av det løpende kvalitetsarbeidet som foregår på institusjonene.

Hva er god evaluering?

Det ligger heller ikke til arbeidsgruppens mandat å diskutere eller gi råd om evalueringsmetoder.²⁹ Spørsmål om valg av evalueringsdesign, kvalitative og/eller kvantitative metoder, bruk av indikatorer, o.l. overlates til aktørene som har ansvar for å gjennomføre evalueringer. Uten å gå inn på metodevalg lar det seg allikevel gjøre å si noe overordnet om hva som kjennetegner god evalueringspraksis, som et utgangspunkt for valg av organisering.

En viktig forutsetning for evalueringene skal oppfylle sine formål er at de oppfattes som legitime. Tilnærmingen til spørsmålet om legitimitet kan være litt ulik for aktørene på feltet. For institusjonene og deres fagmiljøer vil noe av det viktigste for legitimiteten være at evalueringsarbeidet gjennomføres i tråd med vitenskapens egne tradisjoner for fagfellevurdering. Dette innebærer at metodene i evalueringen må være forskningsbaserte, og at det må være eksperter med vitenskapelig kompetanse i relevante fag som gjør selve kvalitetsvurderingene. I tillegg til relevant faglig kompetanse, er det viktig at evalueringene bygger på solid kompetanse i evalueringsmetodikk. Selv om organer som Forskningsrådet og NOKUT har ansvar for prosessene og legger rammer for arbeidet, må kjernen i evalueringsarbeidet ivaretas av vitenskapelig personale som arbeider på faglig grunnlag.

For et lite land som Norge kan det være utfordrende å sette sammen paneler av fagfeller der det ikke forekommer bindinger av noe slag. Både av den grunn, og for å få et internasjonalt

²⁷ Se for eksempel R-QUESTs Policy-brief: <https://www.r-quest.no/wp-content/uploads/2017/12/R-QUEST-Policy-Brief1-2016-final.pdf>

²⁸ Hva som kjennetegner kvalitet i akademisk er en stor debatt, fra Mertons definisjon av akademiske normer på 40-tallet til mer resultatorienterte indikatorer som brukes mye i dag. R-QUEST har summert opp noen grunnleggende krav for at et arbeid skal kunne regnes som vitenskapelig: the work should be (1) well-founded in scientific methods, (2) give new knowledge and (3a) be relevant to the research community and/or (3b).

²⁹ Se for eksempel Science Europe *Position paper*, <https://www.scienceeurope.org/media/3twjxim0/se-position-statement-research-assessment-processes.pdf>

perspektiv på kvaliteten i norsk forskning og høyere utdanning, er det helt nødvendig å trekke inn internasjonale fagfeller i panelene, slik det har vært lang tradisjon for både i Forskningsrådet og NOKUT.

I tillegg til den faglige siden av legitimiteten er det også et hensyn til mer generell habilitet som må ivaretas. Institusjonene er i konkurranse med hverandre langs mange dimensjoner, og det er naturlig at hver institusjon og hvert fagmiljø vil ønske å framstå best mulig i en evalueringsprosess. Ansvarsforholdene og organiseringen av evalueringsarbeidet må derfor være slik at det ikke kan stilles spørsmål ved om evalueringene er uavhengige av de som blir evaluert, og at de er upartiske i sine vurderinger. Dette er ikke til hinder for at evalueringsprosessene kan – og bør – legge opp til bred involvering av institusjonene som skal evalueres (inkludert berørte fagmiljøer og studenter), samt andre relevante aktører, så lenge involveringen ikke innebærer et uheldig press på innholdet i de faglige vurderingene. Tvert imot vil god involvering styrke både legitimiteten og relevansen av evalueringene for de som blir evaluert. Prosessene må være åpne for innspill slik at ulike forventninger og ønsker kan avstemmes. Det som er avgjørende er at ansvars plasseringen sikrer prosessens uavhengighet, og at rollefordelingen sikrer den faglige uavhengigheten i vurderingene.

Selv om det kan være ulike innganger til spørsmålet om legitimitet vil alle aktørene til syvende og sist være avhengig av at evalueringsarbeidet har bred tilslutning og legitimitet. Dette er noe av det viktigste rammeverket må sikre.

Evalueringer er omfattende prosesser som krever mye arbeid både av de som evaluerer og de som blir evaluert. Det er derfor avgjørende at evalueringsprosessene legges opp slik at rapporteringstrykket og øvrig administrativt arbeid holdes på et minimum. Det innebærer god koordinering, langsiktighet i planleggingen og god informasjon til berørte. Som drøftet ovenfor innebærer det også gjennomtenkt og koordinert innhenting av informasjon.

I gjennomgangen av de ulike brukernes behov i kapittel 5 ble det klart at det er relativt stort overlapp i informasjonsbehovene om kvalitet i forskning og høyere utdanning. Dette taler for at evalueringer av kvalitet kan gjennomføres med flere formål, selv om det vil være nødvendig med enkelte tilpasninger underveis i evalueringsprosessene for å sikre best mulig relevans for noen av aktørene. Av brukerne er det institusjonene selv og andre aktører i kunnskapssektoren som har de mest komplekse behovene for informasjon om kvaliteten på forskning og høyere utdanning. Det er derfor naturlig at evalueringsrapportene utarbeides med disse som primær målgruppe. For andre brukere kan det være nødvendig med egne formidlingstiltak i etterkant av evalueringene.

Det er en viss spenning mellom summativ evaluering, som preges av uavhengighet og en viss avstand, og formativ evaluering, hvor nærhet er en forutsetning, og hvor fortrolighet kan være en fordel. Allikevel er gevinstene ved å kombinere summative og formative elementer i samme evalueringsprosesser så store at arbeidsgruppen anbefaler et rammeverk basert på dette. Dette er i tråd med det som har vært praksis i Norge tidligere. I Probas typologi over evalueringer er både Forskningsrådet og NOKUT sine evalueringer kategorisert som sammensatte, selv om Proba mener Forskningsrådet har hatt vekten mer på det summative, mens NOKUT har hatt et mer formativt fokus. Å gjennomføre summative og formative evalueringer separat ville etter arbeidsgruppens vurdering medført betydelig merarbeid både

for institusjonene som skal evalueres og for de som skulle gjennomføre dem. I tillegg ville det være vanskelig å sikre forutsigbarhet og jevnt høy kvalitet i evalueringsarbeidet.

I kapittel 8 ble utfordringene ved å skulle evaluere forskning og høyere utdanning under ett, drøftet. Arbeidsgruppens vurdering er at dette vil resultere i så komplekse evalueringsdesign, og dessuten vil ha så liten relevans i mange tilfeller, at det ikke er hensiktsmessig med et rammeverk som baserer seg på dette som hovedmodell. Arbeidsgruppens forslag til rammeverk tar derfor utgangspunkt i at forskning og utdanning i hovedsak evalueres hver for seg, men med langt høyere ambisjoner for koordinering enn tidligere.

For å oppsummere drøftelsene ovenfor mener arbeidsgruppen at evalueringsvirksomheten bør organiseres og drives etter følgende overordnede prinsipper:

- a. evalueringsvirksomheten må være – og oppfattes som – legitim
- b. evalueringene skal bruke forskningsbaserte metoder, kvalitetsvurderingene skal gjøres av fagfeller, og internasjonale perspektiver skal ivaretas i panelene
- c. organiseringen og gjennomføringen skal være ressurseffektiv for de involverte, og skal så langt som mulig bruke data som allerede er tilgjengelige
- d. evalueringene skal planlegges og organiseres slik at aktørene enklest mulig kan gjøre bruk av dem
- e. aktørene skal involveres i utformingen, gjennomføringen og oppfølgingen
- f. evalueringsrapportene orienteres primært mot behovene til institusjonene og policyaktører i kunnskapssektoren
- g. formidling til brukere av og deltakere i høyere utdanning og forskning (herunder studenter og virksomheter) må tilrettelegges særskilt

To hovedmodeller: Storbritannia og Nederland

Internasjonalt kan vi skille mellom to hovedmodeller for evaluering av forskning og høyere utdanning. Skillet mellom disse handler primært om det er institusjonene selv eller et nasjonalt organ som har ansvar for utforming og gjennomføring av evalueringene. Den tydeligste nasjonale modellen finner vi Storbritannia, som har utviklet et heldekkende evalueringssystem for forskning og utdanning hvor evalueringstema og -kriterier besluttet av de sentrale forsknings- og utdanningsmyndighetene som også koordinerer en samlet evaluering av alle fag med nasjonale paneler.

Den tydeligste desentraliserte modellen finner vi i Nederland, hvor myndighetene har overlatt til institusjonene selv å gjennomføre evalueringer av sin virksomhet basert på en felles nasjonal protokoll. Protokollen er utviklet i et samarbeid mellom universitetenes organisasjoner, vitenskapsakademiet og det nederlandske forskningsrådet.³⁰ Det er også noen andre grunnleggende forskjeller mellom disse modellene, særlig når det gjelder kopling til finansiering. Det britiske systemet blir brukt til å omfordele ressurser mellom lærestedene basert på evalueringsresultatet, mens det nederlandske systemet ikke har noen finansielle konsekvenser. Fokuset i Storbritannia er utpreget summativt mens det er legitimering og læring som er hovedformålene for den nederlandske protokollen.

³⁰ Strategic Evaluation Protocol 2021-2027, https://www.vsnul.nl/en_GB/sep-eng.html

Når omfordeling av økonomiske midler til forsknings- og utdanningsinstitusjoner i Norge håndteres gjennom andre systemer enn evaluering, gir det oss et relativt stort handlingsrom for å tilpasse evalueringsregimet til andre formål. Mens omfordeling av midler normalt vil kreve summative vurderinger av resultater, har vi muligheten til å vektlegge vurderinger av forutsetninger for kvalitet og læring i tillegg som vesentlige elementer i et nasjonalt rammeverk for evaluering.

Valg av modell for forskningsevaluering i Norge

Organiseringen av forskningsevaluering i Norge har en viss likhet med den britiske modellen ved at evalueringene har vært styrt av et nasjonalt organ, og at ulike institusjoner er vurdert utfra felles kriterier etter en nasjonal faginnndeling. Når det gjelder formål er det imidlertid en vesentlig forskjell at Forskningsrådets evalueringer ikke har hatt noen direkte økonomiske konsekvenser.³¹

Arbeidsgruppen har, blant annet på bakgrunn av den internasjonale sammenligningen som Proba gjennomført, drøftet om det ville være hensiktsmessig å legge om forskningsevalueringen i Norge i tråd med den nederlandske modellen, i den forstand at ansvaret for evalueringer ble overført til institusjonene. En målsetting med dette kunne være å legge enda bedre til rette for formative evalueringsprosesser der hver enkelt institusjon fikk legge premissene for evalueringene.

Det er imidlertid flere argumenter som taler imot en slik løsning. For det første ville det være en stor utfordring å sikre legitimiteten til en slik modell. Egenevalueringer er ikke i seg selv uvanlig som ledd i en større evalueringsprosess, men mange aktører ville trolig stille spørsmål ved upartiskheten i et rammeverk hvor hele eller svært mye av ansvaret ble overført til den som skal evalueres. For det andre ville det være utfordrende å ivareta det behovet for sammenlignbarhet som gjennomgangen av brukernes behov har vist at er viktig for mange aktører, inklusive institusjonene selv. Et tredje argument er hensynet til ressurseffektivitet. Evalueringsarbeid er resurskrevende, både med tanke på administrative prosesser (herunder rapportering av underlagsdata) og når det gjelder tilgang på faglig ekspertise til evalueringspanelene. Det virker svært sannsynlig at det ville innebære betydelig merarbeid for norsk forskning og høyere utdanning samlet sett dersom hver av institusjonene skulle utvikle egne løsninger for evalueringer av kvalitet. I tillegg er det varierende hvor stor kapasitet de ulike institusjonene ville ha til å ivareta et slikt ansvar.

På bakgrunn av disse argumentene finner arbeidsgruppen at den modellen vi hittil har hatt i Norge, og som når det gjelder ansvarsforhold ligner den britiske, er den som best ivaretar de hensynene som er beskrevet ovenfor. Dette tilsier en videreføring av Forskningsrådets hovedansvar for forskningsevalueringene. Samtidig mener arbeidsgruppen at det er flere trekk ved den nederlandske modellen som vi kan trekke veksler på. Ett av disse er bruken av protokoller.

³¹ Ett unntak fra dette prinsippet ble gjort i oppfølging av evalueringen av humaniora (HUMEVAL 2017) hvor vurdering av forskergrupper ble lagt til grunn for fordeling av oppfølgingsmidler til institusjonene. En slik praksis ble tydelig frarådet av komiteen for evaluering av samfunnsvitenskap (SAMEVAL 2018).

Når det gjelder hva slags typer forskningsevaluering det er mest hensiktsmessig å satse på, viser Probas analyse at Forskningsrådets fagevalueringer imøtekommer mange av aktørenes behov på en god måte, selv om de er bedre tilpasset de tradisjonelle breddeuniversitetene enn nye universiteter og forskningsinstitutter. Derfor mener arbeidsgruppen at det er viktig i det videre arbeidet å se på hvordan de nasjonale evalueringsprosessene kan bli mest mulig relevante for den enkelte institusjon. Forskningsrådet har under utprøving en evalueringsprotokoll som gir mulighet for at evalueringsmandatet inkluderer faglige mål som er satt ved den enkelte lærested innenfor en felles nasjonal ramme. Arbeidsgruppen anbefaler at Forskningsrådet bygger videre på erfaringene fra piloten i rettsvitenskap som nå er under gjennomføring.

Samtidig viser gjennomgangen i kapittel 5 at mange av aktørene har bruk for sammenlikninger av kvalitetstilstanden innenfor et forskningsfag eller studieprogram ved ulike institusjoner. Dette taler for et rammeverket som sikrer at vi fortsatt får en viss enhetlighet i evalueringsprosessene på tvers av institusjonene som kan gi sammenliknbare resultater på institusjonsnivå og grunnlag for å tegne et nasjonalt bilde av situasjonen innenfor det enkelte fag.

Valg av modell for evaluering av høyere utdanning i Norge

Når det gjelder evaluering av høyere utdanning er det store variasjoner i hvordan dette blir gjort internasjonalt. Flere land, som for eksempel Nederland og Danmark, har tidligere gjennomført re-akkrediteringer av alle studietilbud hvert femte til sjette år. De siste årene har flere land (inkludert Danmark) gått vekk fra krav om re-akkreditering og flyttet fokus til å gjennomføre tilsyn med institusjoner kvalitetssikringssystemer og kvalitetsarbeid. Som med forskningsevalueringer skiller Storbritannia seg ut. Storbritannia har utviklet et, per i dag, frivillig system der utdanningsoppdraget i stor grad blir summativt evaluert og institusjonene blir rangert i tre kategorier, det såkalte Teaching Excellence Framework. Systemet er ikke direkte knyttet til finansiering, men rangeringen til en institusjon påvirker hvor mye institusjonen kan kreve av studenter i studieavgift. Selv om det er mange forskjellige modeller for vurdering av utdanningskvalitet internasjonalt, er det ingen av modellene som gir aktørene en heldekkende vurdering av den faktiske kvaliteten på utdanningene. Et nesten rent kvantitativt opplegg som i Storbritannia gir ikke et uttømmende bilde av kvalitetene studentene møter i sine studieprogram.

Institusjonene i Norge har et tydelig lovpålagt ansvar for å sikre utdanningskvaliteten og NOKUTs ansvar er å sikre at institusjonene jobber systematisk med dette arbeidet gjennom tilsyn med kvalitetsarbeid, samt å føre tilsyn med studieprogrammer når man har grunn til å tro at disse ikke tilfredsstiller lovpålagte krav. I tillegg rapporterer institusjonene på en rekke kvantitative indikatorer som for eksempel kandidater, gjennomstrømming, karakterer og kjønnsbalanse i studieprogrammene. Det samles også inn data om studentenes tilfredshet med sine studieprogram (Studiebarometeret).

På tross av dette er det mange indikasjoner på at både institusjonene og andre aktører mangler et helhetlig bilde av den faktiske kvaliteten studentene møter i sine studieprogrammer. For å bøte på dette ble det i 2017 gjennomført en endring i Studiekvalitetsforskriften der institusjonene selv ble pålagt å gjennomføre periodiske evalueringer av studieprogrammene.

Representanter fra arbeids- eller samfunnsliv, studenter og eksterne sakkyndige, som er relevante for studietilbudet, skal ifølge forskriften bidra i evalueringene. En fordel med en så desentralisert løsning er at den gir den enkelte institusjon eierskap til evalueringsprosessen. Samtidig innebærer det at det er, og vil være, store forskjeller på hvordan evalueringene gjennomføres, hva som blir vurdert og når de gjennomføres. Det betyr at ordningen slik den er regulert nå ikke vil gi et nasjonalt bilde av utdanningskvaliteten.

Arbeidsgruppen har drøftet to alternative modeller for evalueringer av utdanning: en nasjonal modell som gir NOKUT ansvaret for gjennomføringen, og en desentralisert modell som bygger videre på institusjonenes ansvar for periodiske evaluering av sine studieprogrammer ved å utforme en felles nasjonal protokoll for disse evalueringene, noe som vil gi mulighet for å utarbeide nasjonale synteser.

Den nasjonale modellen vil innebære en utvidelse av NOKUTs evalueringsarbeid til å dekke bredden av norsk høyere utdanning. Ambisjonen bør være at man arbeider seg systematisk gjennom de ulike fagområdene innenfor ett gitt tidsintervall, og at man har flere fagområder under evaluering til enhver tid. Evalueringene koordineres av NOKUT i tett dialog med institusjonene og Kunnskapsdepartementet. NOKUT samler inn og tilrettelegger data som vurderes opp mot etablerte kriterier av uavhengige fageksperter. Hvis en slik modell velges, bør det vurderes om denne rollen bør medføre regelendringer, eller om den dekkes av dagens regelverk. NOKUTs rolle må vurderes i sammenheng med en større gjennomgang av ansvars- og rollefordeling mellom departement, direktorat og UH-institusjonene, slik innspill og høringsuttalelser i sektoren har etterlyst.

De største fordelene med en nasjonal modell er at vi vil få et nasjonalt kunnskapsgrunnlag om kvalitetstilstanden på norsk høyere utdanning, og samtidig lette arbeidsbyrden for institusjonene. Dette oppnås delvis ved å gjennomføre målrettede og effektive evalueringsprosesser og delvis ved å endre bestemmelsen om periodiske evalueringer i studiekvalitetsforskriften slik at man kan sikre koordinering og ivareta behovet for et nasjonalt bilde av kvalitetstilstanden. For å sikre at evalueringene innrettes slik at de tar hensyn til lokale betingelser og behov, samt det institusjonelle mangfoldet i UH-sektoren, kan man ta i bruk protokoller der noe av evalueringen er felles og noe er tilpasset særskilte behov ved institusjonene.

Alternativet vil være en desentralisert modell der hovedansvaret for evalueringer av utdanningskvalitet ligger hos institusjonene (men da med endret regelfesting, se nedenfor). For å sikre at vi får et godt kunnskapsgrunnlag på nasjonalt nivå vil også dette alternativet kreve at NOKUT i samråd med institusjonene blir enige om en felles evalueringsprotokoll. En slik evalueringsprotokoll må tydeliggjøre hvilke kvalitetsdimensjoner som skal evalueres, hvilke evalueringskriterier og data som skal ligge til grunn for vurderingene, og hvem som skal gjennomføre vurderingene. I tillegg må man enes om en evalueringsplan slik at alle studieprogrammer innenfor et fagområde blir evaluert innenfor et gitt tidsintervall. Institusjonene kan legge til kvalitetsdimensjoner for å oppnå en bedre lokal tilpasning, men det vil være viktig at evalueringsprotokollen sikrer en viss standardisering slik at man får fram et nasjonalt bilde på utdanningskvaliteten. For å sikre nasjonal sammenliknbarhet, vil det være hensiktsmessig at NOKUT gjennomfører sammenstillingen, mens institusjonene er ansvarlige for gjennomføringen, herunder innsamling og tilrettelegging av data, for hver enkelt studieprogramevaluering.

Den største fordelen med den desentraliserte modellen er at institusjonene selv får muligheten til å bygge ut evalueringsprotokollen slik at man i enda større grad kan bygge inn lokale hensyn og strategiske prioriteringer i evalueringene. En slik modell vil gi institusjonene enda større mulighet til å tilpasse evalueringenes mandat til deres behov. Et mulig negativt aspekt vil være at institusjonene likevel vil være bundne til fastlagte tidsintervaller, og at det vil være arbeidsbelastende for institusjonene å gjennomføre den praktiske koordineringen og gjennomføringen av evalueringene selv. Det vil også kreve at institusjonene rekrutterer og honorerer fagfellene. For NOKUT, Kunnskapsdepartementet og andre aktører vil det nasjonale bildet kun bli synlig etter at samtlige studietilbud innenfor et fagområde har blitt «lokalt» evaluert innenfor den fastsatte tidsrammen.

Arbeidsgruppens vurdering er at begge modellene vil imøtekomme behovene som er nevnt i mandatet, og som er nærmere drøftet i denne rapporten. Videre mener arbeidsgruppen at fordeler og ulemper ved de to modellene bør ses i lys av de veivalgene som skal gjøres i den styringspolitikken som er under utvikling, og som vil legges fram for Stortinget våren 2021.

De to modellene vil begge gjøre det nødvendig å endre formuleringen av bestemmelsen i Studiekvalitetsforkriften om periodiske evalueringer av studieprogrammene. Arbeidsgruppen anbefaler at departementet vurderer fordeler og ulemper ved de to modellene i lys av høringen i forbindelse med en slik forskriftsendring.

Arbeidsgruppen vil for øvrig understreke hvor viktig det er at studentenes stemme blir godt ivarettatt. I den forbindelse viser arbeidsgruppen til at alle norske institusjoner, samt NOKUT er pålagt å følge Standards and Guidelines for Quality Assurance in the European Higher Education Area, og gjennom dette sikres studentene en tydelig innflytelse i institusjonenes kvalitetsarbeid.

Oppsummering om valg av modeller

Oppsummert har arbeidsgruppen kommet fram til at det vil være formålstjenlig å utarbeide parallelle og koordinerte nasjonale evalueringsløp for henholdsvis forskning og høyere utdanning som med gitte tidsintervaller gir en nasjonal oversikt over kvalitetstilstanden i de ulike fagområdene. Dette innebærer i stor grad en videreføring av den nasjonale organiseringen av forskningsevalueringer, mens det innebærer en mer systematisk tilnærming når det gjelder utdanningsevaluering ved at innhenting, analyse og vurdering av underlagsdata koordineres av NOKUT, enten gjennom en nasjonal eller en desentralisert modell (se ovenfor).

Konsekvenser for koordinering

Det er behov for koordinering på flere plan for at et nasjonalt rammeverk for evaluering skal fungere godt. En tydelig funn i interessentanalysen til Proba er at institusjonene ønsker seg mer forutsigbarhet om hva som skal evalueres når. Det er også et behov for å koordinere evalueringene til NOKUT, Forskningsrådet, og eventuelt andre aktører som evaluerer på oppdrag fra Kunnskapsdepartementet for å redusere respondentbelastningen.

I tillegg er det som drøftet i kapittel 8 nødvendig med løpende koordinering for å vurdere i hvilke fagområder det er aktuelt at forskning og utdanning evalueres sammen. Arbeidsgruppen mener det er vanskelig å vurdere på generelt grunnlag hvilke fagområder

dette vil være mest relevant for, og hvordan evalueringene av henholdsvis forskning og utdanning skal koordineres i det enkelte fag. Det kan argumenteres for at koblingen mellom de to er spesielt relevant for profesjonsfag, men samtidig er det noen ganger slik at dette temaet er underfokuset i disiplin-fag, og nettopp derfor bør gjøres til gjenstand for evaluering. Behovet for evalueringer som ser på koblingen mellom forskning og høyere utdanning vil også variere over tid.

For å sikre en god koordinering og kopling mellom forskning og utdanning, og for å sikre best mulig forutsigbarhet for berørte aktører, foreslår arbeidsgruppen at det etableres et koordineringsforum for evalueringsarbeidet som ledes av Kunnskapsdepartementet. Koordineringsforumets hovedoppgave bør være å etablere en overordnet rullerende plan for forsknings – og utdanningsevalueringer. Dette vil forde at Forskningsrådet og NOKUT koordinerer sine evalueringer av forskning og utdanning på et overordnet fagområde-nivå, slik at evalueringene skjer innenfor det samme tidsrommet for de samme fagområdene, mens evalueringsenheter defineres av institusjonene selv.

Arbeidsgruppen mener Kunnskapsdepartementet bør invitere relevante aktører som deltagere og formulere et overordnet mandat for koordineringsforumet. Møtefrekvensen for forumet vil sannsynligvis kunne variere over tid. I implementeringsfasen kan det være behov for hyppigere møter, men når et rammeverk er på plass og en rullerende kalender for virksomheten er etablert, vil det sannsynligvis være tilstrekkelig med årlige møter. Kunnskapsdepartementet, NOKUT, Forskningsrådet, institusjonene, studenter, forskere og undervisere, samt representanter fra berørte sektorer vil være aktuelle deltakere. Arbeidsgruppen foreslår at Kunnskapsdepartementet vurderer nærmere hvilke aktører som bør representeres, hvilket nivå deltakelsen på være på, og hvilke prosedyrer for saksforberedelse og beslutninger som er nødvendig for at forumet skal kunne ivareta sine formål.

10. Arbeidsgruppens forslag til nasjonalt rammeverk

1. Virkeområde og formål med rammeverket

Rammeverket omfatter primært nasjonale evalueringer av kvalitet i forskning og høyere utdanning. Med dette menes forsknings- og fagfellebaserte evalueringer gjennomført etter mandat fra et nasjonalt myndighetsorgan i kunnskapssektoren, og som består av:

- a. beskrivelse av den aktiviteten som evalueres, inkludert konteksten
- b. vurdering av kvaliteten på aktiviteten, opp mot internasjonal standard (der dette er relevant)
- c. sammenligning av kvalitetsnivå mellom nasjonale og evt internasjonale institusjoner¹
- d. råd og anbefalinger

Formålet med dette rammeverket er å sikre effektiv og koordinert evaluering av norsk forskning og høyere utdanning på en slik måte at ulike aktørers og samfunnets behov for kunnskap om kvaliteten i forskning og høyere utdanning imøtekommes.

De nasjonale evalueringene inngår i en kontekst av andre virkemidler for kvalitetssikring og -utvikling, både på nasjonalt nivå (for eksempel tilsyn, akkreditering og programevalueringer), og på institusjonsnivå (for eksempel kvalitetssikringssystemer og institusjonenes egne evalueringer). Rammeverket skal bidra til helhet og sammenheng i bruken av disse virkemidlene, og det er en forutsetning at rapporteringsomfanget samlet sett ikke øker.

2. Formål med nasjonale evalueringer av kvalitet

Evalueringsvirksomheten som omfattes av rammeverket skal bidra til:

- a. kvalitetsarbeidet ved institusjonene
- b. kunnskapsbasert policyutvikling
- c. et godt kunnskapsgrunnlag for oppdragsgivere, studiesøkere og samarbeidspartnere
- d. tillit til forskning og høyere utdanning i befolkningen
- e. åpenhet om bruk av offentlige midler

3. Overordnede prinsipper

Evalueringsvirksomheten organiseres og drives etter følgende overordnede prinsipper:

- a. evalueringsvirksomheten må oppfattes som legitim
- b. evalueringene skal bruke forskningsbaserte metoder, kvalitetsvurderingene skal gjøres av fagfeller, og internasjonale perspektiver skal ivaretas i panelene
- c. organiseringen og gjennomføringen skal være ressurseffektiv for de involverte, og skal så langt som mulig bruke data som allerede er tilgjengelige
- d. evalueringene skal planlegges og organiseres slik at aktørene enklest mulig kan gjøre bruk av dem
- e. aktørene skal involveres i utformingen, gjennomføringen og oppfølgingen
- f. evalueringsrapportene orienteres primært mot behovene til institusjonene og policyaktører i kunnskapssektoren
- g. formidling til brukere av og deltakere i høyere utdanning og forskning (herunder studenter og virksomheter) må tilrettelegges særskilt

4. Særskilt om forskningsevaluering

Forskningsrådet har ansvar for nasjonale evalueringer av norsk forskning. I den videre omtalen skilles det mellom *fagevalueringer* og *temaevalueringer*. Fagevalueringer har kvalitetsvurdering og -forbedring som sitt primære formål, og skal over tid omfatte den samlede forskningsaktiviteten i UH-sektoren og instituttsektoren. Temaevalueringer gjennomføres etter behov for å vurdere forskningskvalitet og -kapasitet opp mot nasjonale mål og prioriteringer, f.eks. i Langtidsplanen for forskning og høyere utdanning.

Fagevalueringene skal:

- legge til rette for læring og fagutvikling ved de evaluerte institusjonene
- legges opp slik at de gir grunnlag for sammenligning av kvalitetsnivå mellom institusjoner, og sett opp mot en internasjonal forskningsfront
- ta i betraktning forskningsinstitusjonens og det aktuelle fagmiljøets profil, strategiske mål og ressursituasjon
- ta hensyn til tverrfaglig forskningsaktivitet og -organisering, blant annet i instituttsektoren, når det skal gjøres faglige avgrensninger for den enkelte evaluering
- ta hensyn til evalueringsenhetenes spesifikke kontekst, herunder utdanningsoppdraget i UH-sektoren og oppdragsmarkedet for instituttsektoren
- dokumentere organisatoriske forhold som kan tenkes å framme eller hemme kvalitet, for eksempel gjennom en egevaluering ved institusjonene
- tilrettelegges slik at evalueringsdata skal gi grunnlag for en substansiell vurdering av forskningens innhold og dens relevans i en videre institusjonell og samfunnsmessig kontekst

Temaevalueringer skal:

- baseres på data som allerede er samlet inn gjennom fagevalueringer og øvrig nasjonal FoU-statistikk, rapportering til Forskningsrådet etc.
- gi vurderinger og anbefalinger om kapasitet, kvalitet og relevans for norsk forskning innenfor et definert tema/område sett opp mot definerte forskningspolitiske mål på nasjonalt nivå
- inkludere summative elementer som gir grunnlag for å vurdere nivået på innsatsen på tvers av institusjoner, og på aggregert nasjonalt nivå
- vurdere måloppnåelsen i sammenheng med de samlede nasjonale ressurser som er brukt innenfor et gitt forskningsområde/tema
- dokumentere organisatoriske forhold som kan tenkes å framme eller hemme kapasitets- og kvalitetsutvikling – med særlig oppmerksomhet om det samlede virkemiddelapparatet og dets effekt på forsknings- og innovasjonssystemet
- inkludere et systemperspektiv

Ambisjonen for evalueringer av kvaliteten i norsk forskning er å dekke den samlede forskningsaktiviteten i UH-sektoren og instituttsektoren i løpet av en 10-års periode. For å sikre forutsigbarhet og god koordinering med andre prosesser utarbeides det en rullerende langtidsplan for evalueringer av forskning. Rekkefølge og frekvens vil bli fastsatt av Forskningsrådet i samråd med berørte aktører i et eget koordineringsforum (se nedenfor, under koordinering og planlegging).

I tillegg til fagevalueringer og temaevalueringer vil det fremdeles være behov for andre former for evaluering i regi av Forskningsrådet, eksempelvis instituttevalueringer og

programevalueringer (jf nærmere omtale i rapporten). Der det benyttes ulike evalueringssløp for ulike formål skal disse utformes slik at det blir en tydelig arbeidsdeling mellom dem.

5. Særskilt om utdanningsevaluering

NOKUT har ansvar for nasjonale evalueringer av norsk høyere utdanning. Den mest sentrale evalueringsformen er *nasjonale utdanningsevalueringer*, som gjennomføres for å vurdere de mest sentrale kvalitetsområdene for samtlige studieprogrammer innenfor et fagområde. Eksempelvis kan slike evalueringer ta for seg alle sosionomutdanninger eller alle utdanninger i rettsvitenskap, enten på alle utdanningsnivå (bachelor, master, PhD) eller kun på utvalgte nivåer.

I tillegg til de overordnede målene for nasjonale evalueringer (se innledningen) skal nasjonale utdanningsevalueringer mer spesifikt bidra til:

- kvalitetsforbedring ved institusjonene ved at de får muligheten til å sammenligne sine studietilbud med tilsvarende studietilbud ved andre institusjoner, og med en internasjonal kvalitetsmålestokk. Dette gir bedre muligheter for både kvalitetsheving og for spissing/diversifisering av studietilbudene.
- en kunnskapsbasert politikk for høyere utdanning fra Kunnskapsdepartementet, et tydelig kunnskapsgrunnlag for NOKUTs regelverksforvaltning og risikobaserte tilsyn og for Dikus prioriteringer og innretning av kvalitetsstimulerende programmer
- god informasjon til studiesøkere, studenter og arbeidsgivere om kvalitetsstilstanden på studieprogrammer
- å muliggjøre en lavest mulig grad av statlig regulering av høyere utdanning ved å gi NOKUT, KD og andre relevante aktører regelmessig informasjon om utdanningenes innhold og kvalitet

De nasjonale utdanningsevalueringene omfatter innsatsområder som til sammen er avgjørende for kvaliteten ved et studieprogram. I følge NOKUTs dokument *Kvalitetsområder for studieprogrammer*², er følgende kvalitetsområder avgjørende for utdanningskvaliteten på et studieprogram:

- Startkompetanse
- Læringsmiljø
- Samfunn og arbeidsliv
- Programdesign
- Utdanningsfaglig kompetanse
- Læringsbane
- Kunnskapsbase
- Læringsutbytte

I tillegg til de behov som ivaretas gjennom nasjonale utdanningsevalueringer oppstår det til tider behov for å gjennomføre mer tematiske utdanningsevalueringer. Dette er evalueringer som kun tar for seg spesifikke kvalitetsområder og som ofte ikke vil dekke hele det nasjonale tilbudet av studieprogram innenfor et fagområde. Et eksempel på dette er evalueringen av arbeidslivsrelevans i disiplinlag.

Ambisjonen for nasjonale utdanningsevalueringer er å komme gjennom det samlede utdanningstilbudet i løpet av en tiårsperiode. Dette fordrer konsentrerte og effektive evalueringprosesser blant annet ved at evalueringene i stor grad baserer seg på syntetisering

av eksisterende data, herunder statistikk fra DBH, studiebarometertall, studieplaner og annen dokumentasjon fra studieprogrammene.

For å sikre forutsigbarhet og god koordinering med andre prosesser utarbeides det en rullerende langsiktig plan for utdanningsevalueringer. Rekkefølge og frekvens vil bli fastsatt av NOKUT i samråd med berørte aktører i et eget koordineringsforum (se nedenfor, under koordinering og planlegging).

I vurderingen av hvilke fagområder som skal evalueres når skal det tas hensyn til faktorer som tidsperiode siden siste evaluering, endringer i regelverk for utdanningene eller andre reformer, behov for kunnskap om spesielle utdanninger etc. Det er også viktig at rekkefølgen tar hensyn til andre prosesser og evalueringer i sektoren for å sikre god koordinering. Den langsiktige evalueringsplanen tar utgangspunkt i NOKUTs kategorisering av utdanningstyper. NOKUTs kategorisering bygger på norsk standard for utdanningsgruppering (NUS2000) fra Statistisk sentralbyrå, men er tilpasset for at man skal bedre skal kunne sammenligne lignende studieprogrammer. NOKUTs kategorisering inneholder ca. 45 utdanningstyper.

6. Særskilt om institusjonenes egne evalueringer

De nasjonale evalueringene av forskning og høyere utdanning som er omtalt ovenfor skal hensynta og koordineres med følgende evalueringer som institusjonene selv har ansvar for:

Institusjonenes evaluering av kvalitet:

Institusjonene har ansvar for kvaliteten i det de utfører av forskning og høyere utdanning, og de har etablert ulike virkemidler for å ivareta dette ansvaret. Siden 2017 har Studiekvalitetsforskriften pålagt universitetene og høyskolene å gjennomføre egne jevnlig evalueringer av studiekvaliteten. Arbeidsgruppen foreslår at denne bestemmelsen endres når det er etablert en langtidsplan for nasjonale utdanningsevalueringer, slik at innhenting av grunnlagsdata blir koordinert for å kunne bidra til å gi et nasjonalt bilde av kvalitetstilstanden i ulike fagområder, i tillegg til å benyttes ved den enkelte institusjon.

Institusjonenes evaluering av strategisk måloppnåelse:

Forsknings- og høyere utdanningsinstitusjoner har videre et ansvar for å fastsette mål for egen virksomhet, og for måloppnåelsen knyttet til disse.³ I tillegg velger de fleste institusjonene å utarbeide strategiske planer med en lengre tidshorisont. Dette tilsier at institusjonene selv bør ha tydelig eierskap og handlefrihet i hvordan de organiserer evalueringer av egen institusjon med utgangspunkt i sine egne strategiske mål.

Løpende rapportering og monitorering

Samtidig er det viktig at data som rapporteres løpende fra grunnenhetene ved institusjonene blir best mulig utnyttet, både i institusjonenes egne evalueringer og i de nasjonale evalueringene. Dette gjør det nødvendig med en viss koordinering av rapporteringen, blant annet for å unngå dobbeltarbeid, og for å sikre god sammenlignbarhet.

7. Koordinering og planlegging

Nasjonale evalueringer av kvalitet i forskning og høyere utdanning planlegges og koordineres for å oppnå:

1. Mulighet til å analysere sammenhenger:
 - a. mellom forskning og høyere utdanning i ulike fag/programmer
 - b. mellom forskning/utdanning og øvrige kjerneoppgaver ved institusjonene, herunder innovasjon
 - c. mellom enkeltinstitusjoners strategiske arbeid og kvalitetstilstanden i deres forskning, høyere utdanning og evt andre kjerneoppgaver
2. Forutsigbarhet, slik at aktørene kan vurdere hvordan de vil gjøre nytte av de nasjonale evalueringene i andre prosesser
3. Lavest mulig rapporteringstrykk

Koordinering mellom aktørene sikres gjennom:

1. Et koordineringsforum forankret i KD
2. Offentlig tilgjengelige, langsiktige planer for evalueringsvirksomheten