

Kartlegging av evalueringer av kvalitet i forskning og høyere utdanning

Sluttrapport

Rapport 2020 - 12

Proba-rapport nr. 12, Prosjekt nr. 19105
ISSN: 1891-8093
HSS/KMV/OSL/AG, 25.9.2020

--

Kartlegging av evalueringer av kvalitet i forskning og høyere utdanning

Sluttrapport

Utarbeidet for Kunnskapsdepartementet

Forord

Denne sluttrapporten avslutter prosjektet Kartlegging av evalueringer. Oppdragsgiver for prosjektet er Kunnskapsdepartementet, og vi vil takke for at vi har fått muligheten til å gjennomføre denne kartleggingen.

Kjernen i kartleggingen er interessentanalysen som utgjør grunnlaget for denne sluttrapporten. I arbeidet med denne har vi vært i kontakt med et stort antall personer som har vært fullstendig uvurderlige for sluttresultatet. Vi vil rette en stor takk til oppdragsgivere, deltakere i prosjektgruppen, og alle interessentene vi har intervjuet. Vi har møtt stor velvilje, stort engasjement og særdeles mye kompetanse. Prosjektet kunne ikke vært gjennomført uten deres hjelp, og vi har både lært mye og satt stor pris på møtene med dere.

Henrik Skaug Sætra har vært prosjektleder for utredningen, og prosjektmedarbeidere på prosjektet har vært Kaja Meeg Valvatne og Otto S. Lillebø. Audun Gleinsvik var kvalitetssikrer.

Oslo 25. september, 2020

Proba Samfunnsanalyse

Innhold:

SAMMENDRAG OG KONKLUSJONER	1
1 INNLEDNING	5
1.1 Problemstillinger	5
1.2 Metode og utvalg	6
1.3 Oppbygging og struktur	7
2 FORMÅLET MED EVALUERINGER	8
2.1 «Accountability» og kunnskapsinnhenting	8
2.2 Benchmarking og sammenligning nasjonalt og internasjonalt	9
2.3 Kvalitetsutvikling/læring	9
2.4 Forskningspolitikk, prioritering og strategiske grep	10
2.5 Formålet med utdanningsevaluering	10
3 VURDERING AV TIDLIGERE EVALUERINGER	12
3.1 Frekvens og periodisering	12
3.2 Omfang og belastning	13
3.3 Innretning og metodikk	14
3.4 Hvem eller hva evalueres?	15
3.5 Vurdering av utdanningsevalueringer	17
3.6 Samordning av evaluering av forskning og utdanning	19
3.6.1 Behov, fordeler og mulige utfordringer	19
3.6.2 Mulige måter å gjøre det på/løsninger	20
4 BRUK OG EFFEKTER AV EVALUERINGER	22
4.1 Bruk av evalueringer	22
4.2 Effekter av evalueringer	27
5 BEHOV FOR EVALUERINGER	34
5.1 Behov for evalueringer av forskning	34
5.2 Behov for evalueringer av utdanning	38
5.3 Forskjellige formål og forutsetninger	42
5.4 Løpende indikatorer og foreliggende data	44
5.5 Behov for koordinering og forutsigbarhet	45
6 UTVIKLINGSTREKK	48
6.1 Åpen forskning	48
6.2 Tverrfaglighet	48
6.3 Impact og samfunnsnytte	49
6.4 «Fake news», utdanning og tillit til forskning	49
6.5 «Learning analytics» og økende tilgang til data	50
6.6 Øvrige utviklingstrekk	51
7 OPPSUMMERING	52

Vedlegg 1: Internasjonal sammenligning

Vedlegg 2: Typologisering

Sammendrag og konklusjoner

Resymé

Evalueringer er et sentralt virkemiddel for å sikre høy kvalitet i forskning og høyere utdanning. I dag gjennomføres evalueringer på mange ulike måter av flere aktører, og dette oppleves som både uoversiktlig og uforutsigbart av mange i sektoren. Det er derfor et behov for en helhetlig gjennomgang av evalueringssystemet for å undersøke hvilken rolle evalueringer kan og bør spille når det kommer til kvalitetssikring og kvalitetsutvikling blant forsknings- og undervisningsinstitusjoner i Norge.

Vi har gjennomført en interessentanalyse der vi har undersøkt sentrale aktørers vurderinger av gjennomførte evalueringer, hvordan de bruker evalueringene og hvilke behov de har for evalueringer av kvalitet i forskning og høyere utdanning. Det finnes svært mange, og til dels meget ulike, oppfatninger og meninger blant interessentgruppene. Det er allikevel relativt bred enighet om at evalueringer er nyttige, nødvendige og lærerike, og at de spiller en viktig rolle i kvalitetsutvikling. Likevel etterspør mange av interessentene endringer i evalueringenes innretning og metodikk, samt større grad av koordinering og forutsigbarhet i evalueringssystemet som helhet.

Bakgrunn

Norge som kunnskapssamfunn er avhengig av at både utdanning og forskning holder høy kvalitet, og evalueringer er både et viktig virkemiddel og en sentral informasjonskilde i denne sammenheng. I dag gjennomføres evalueringer og andre vurderinger av forskning og utdanning av flere aktører, på mange ulike måter og nivåer. De siste årene har det derfor blitt diskutert hvordan man best kan måle kvalitet, og hvordan man kan sikre en god og effektiv evalueringspraksis.

Kunnskapsdepartementet har derfor satt i gang et utviklingsprosjekt som skal undersøke muligheten for å utvikle et nasjonalt rammeverk for evalueringer av forskning og høyere utdanning. Formålet 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.

Problemstilling

Som en del av kunnskapsgrunnlaget for utviklingsprosjektet har vi kartlagt de mest sentrale aktørenes behov for og bruk av evaluering av norsk forskning og høyere utdanning. Med de mest sentrale aktørene menes Kunnskapsdepartementet, Forskningsrådet og NOKUT, samt et utvalg av universiteter og høyskoler, forskningsinstitutter, departementer, direktorater og helseforetak. Vi har intervjuet totalt 53 representanter for 34 organisasjoner.

Vi har undersøkt hvordan sentrale aktører vurderer allerede gjennomførte evalueringer, og hvilke styrker, svakheter og forbedringspunkter de ser i dagens evalueringspraksis. Vi har også undersøkt hvordan aktørene bruker evalueringer, og hvilke effekter evalueringer har hatt – både innad i kunnskapssektoren og i samfunnet for øvrig. Til slutt har vi kartlagt hvilke behov de ulike aktørene har for evalueringer og hvordan de ønsker at et eventuelt nasjonalt rammeverk for evalueringer skal utformes.

I intervjuene kommer det klart frem at de ulike aktørenes oppfatninger, perspektiver og behov i stor grad påvirkes av hva slags organisasjon de representerer. Vi ser for eksempel klare forskjeller mellom små og store institusjoner og mellom UH-sektoren og instituttsektoren. Vi har avdekket at de forskjellige interessentene har til dels ulike behov,

oppfatninger og ønsker, men at det likevel eksisterer en viss enighet om noen av de sentrale temaene i interessentanalysen.

Vurderinger av gjennomførte evalueringer

Vi har spurt interessentene om deres oppfatninger av evalueringenes frekvens, omfang, innretning og metodikk. Generelt virker det som om aktørene er fornøyde med både frekvens og omfang av evalueringene. Det er tydelig at evalueringer innebærer en stor arbeidsbelastning for både faglig og administrativt personale på institusjonene, men at dette oppfattes som rimelig og nødvendig. Så lenge institusjonene ser en klar effekt og nytte av evalueringene, anses ikke belastningen som urimelig.

Når det kommer til innretning og metodikk finner vi delte meninger blant interessentene og de ulike aktørgruppene. Selv om flere setter pris på sammenligninger og mer summative evalueringer, er det likevel mange som etterspør et større fokus på å bidra til institusjonenes utvikling og kvalitetsarbeid. Mange mener at evalueringene svarer godt på *hvordan* ting står til, men ikke *hvorfor*.

Interessentene mener derfor det ikke er tilstrekkelig å basere seg på kvantitative data, indikatorer og bibliometri i evalueringene. Flere mener at mer kvalitative metoder, som «site visits», bruk av internasjonale ekspertpaneler og egenevalueringer fra institusjonene selv, er viktige.

En type evaluering mange virker godt fornøyde med er NFRs fagevalueringer. Mange av interessentene ønsker en fortsettelse av disse, og at de blir gjennomført på en mer forutsigbar måte med for eksempel faste sykluser på mellom 5 og 10 år.

Vi spurte også de ulike interessentene om de mener det ville være mulig og/eller nyttig å evaluere forskning og utdanning *sammen*. De fleste mente dette ville være svært nyttig ettersom de to aktivitetene ofte påvirker hverandre. Likevel uttrykte mange tvil om det ville være mulig å gjøre dette på en god måte i praksis.

Bruk og effekter av evalueringer

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. Flere aktører peker på at institusjonene må tilpasse seg evalueringenes fokus, for eksempel ved å opprette forskergrupper dersom evalueringer bruker dette som evaluingsenhet. I tillegg vil evalueringenes tolkning av hva kvalitet er legge føringer for institusjonenes kvalitetsarbeid, samt de ansattes forsknings- og undervisningsaktiviteter. Her er særlig de fagansatte selv og fagforeningene opptatt av å utvikle en forutsigbar evalueringspraksis for å skape en viss stabilitet for de ansatte.

De fleste er enige om at evalueringer virker skjerpene på institusjonene og «tvinger» dem til å fokusere på kvalitetsarbeid og -utvikling. Her kommer det frem 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 egenevalueringer, fører evalueringene til en stor grad av læring internt på institusjonene. Likevel påpeker flere aktører at evalueringene bare er én av faktorene som påvirker kvalitetsarbeidet. Andre viktige faktorer er for eksempel lovverk

og forskrifter, internasjonal og nasjonal konkurranse om forskningsmidler og markedskrefter. Sistnevnte gjelder i størst grad for private institusjoner og instituttsektoren.

For at resultatene av evalueringer skal spille en større rolle, etterspør enkelte interessenter også tydeligere konsekvenser av evalueringer – enten i form av å premiere god innsats, muligheter til å søke oppfølgingsmidler eller ved å identifisere områder eller institusjoner med et ekstra behov for hjelp og støtte.

Til sist ser vi at evalueringer i liten grad når ut til andre enn KD, evaluatorene (NFR 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.

Behov for evalueringer

Aktørene vi har intervjuet ser positive aspekter ved flere ulike typer evalueringer, og det blir uttrykt et behov for både fagevalueringer, institusjonsevalueringer og mer summative sammenligninger. Svært mange av interessentene sier de ønsker flere fagevalueringer, men at det også er positivt med ulike evalueringer som fokuserer på ulike sider ved forskningsproduksjon og undervisning. Flere virker skeptiske til at evaluatorene kun skal begrense seg til én type evaluering.

Mange av aktørene har behov for at evalueringene skal ta hensyn til de ulike institusjonenes samfunnsoppdrag og roller. Flere peker på at man ikke kan evaluere et stort breddeuniversitet og en mindre, profesjonsrettet høyskole etter de samme kvalitetskriteriene. Det samme gjelder for eksempel et disiplinfokusert universitet og et mer tverrfaglig institutt som driver mye oppdragsforskning. Flere mener derfor at man i større grad bør evalueres ut ifra sine egne forutsetninger.

Til sist påpeker tilnærmet alle interessentene at evalueringsregimet, slik det er i dag, er uoversiktlig og uforutsigbart. En grunn til dette er at det har kommet flere aktører i sektoren som har utdannings- eller forskningskvalitet som en del av sitt mandat. I tillegg har rapporteringen og tilgangen på bibliometriske og andre kvantitative data økt betydelig. Institusjonene forteller derfor om et uoversiktlig landskap der de opplever et høyt rapporteringstrykk. Det er derfor et behov for en større grad av koordinering, samordning og forutsigbarhet. Her er det viktig å se evalueringene i sammenheng med løpende indikatorer og andre typer data. Da kan man identifisere merverdien av evalueringer og utvikle et evalueringssystem som både gir nødvendig informasjon og oppleves som nyttig for institusjonene selv.

Konklusjoner

Vi har intervjuet aktører med til dels svært ulike interesser og oppfatninger av evaluering av kvalitet i forskning og høyere utdanning. Likevel er det enkelte meninger og behov som ser ut til å gjelde alle eller de fleste av interessentene. For det første er tilnærmet alle vi har intervjuet enige om at evalueringer av kvalitet er både viktige og nyttige. Derfor er de også langt på vei villige til å akseptere den arbeidsbelastningen som kommer av evalueringene.

Ulike evalueringstyper kommer med både styrker og svakheter, og dekker ulike interessenters behov. Derfor peker mange på at det er en god idé å ha en viss variasjon i evalueringsmetoder, og ikke kun fokusere på én type evalueringer. Likevel trekker mange interessenter frem NFRs fagevalueringer som svært nyttige, og noe de ønsker mer av.

Det er tydelig at UH- og instituttsektoren består av svært ulike institusjoner, og det har også preget svarene vi har fått i interessentanalysen. Flere av institusjonene opplever det som noe problematisk å skulle evalueres opp mot de samme kvalitetskriteriene som institusjoner som er svært ulike dem selv. Man kan derfor diskutere om evalueringer i større grad skal ta hensyn til institusjonenes ulikheter og åpne opp for mer differensierte evalueringer.

Svært mange av de vi har intervjuet etterspør bedre koordinering og mer forutsigbarhet i evalueringssystemet. Flere aktører har etterspurt faste evalueringssykluser eller et evalueringsprogram for hele sektoren. Utviklingsprosjektets mål om å utvikle et nasjonalt rammeverk for evalueringer av forskning og høyere utdanning virker derfor både nyttig og betimelig, og noe som kan bidra til å svare på flere av aktørenes behov.

1 Innledning

Dette analyseoppdraget skal utgjøre deler av kunnskapsgrunnlaget for et utviklingsprosjekt initiert av Kunnskapsdepartementet, og som gjennomføres av en prosjektgruppe med deltakere fra Norges forskningsråd (Forskningsrådet), NOKUT, Universitets- og høyskolerådet (UHR), Forskningsinstituttene fellesarena (FFA), Olje- og energidepartementet (OED) og Norsk studentorganisasjon (NSO). Formålet 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.

Vårt oppdrag omfatter en kartlegging av de mest sentrale aktørenes behov for og bruk av evaluering av norsk forskning og høyere utdanning. Med de mest sentrale aktørene menes Kunnskapsdepartementet, Forskningsrådet og NOKUT, samt et utvalg av universiteter og høyskoler, forskningsinstitutter, departementer, direktorater og helseforetak. I tillegg skal oppdraget gi en mer generell beskrivelse av hvilken betydning denne typen evalueringer har for andre samfunnsaktører.

I denne rapporten gjengir vi resultatene fra en interessentanalyse som ble gjennomført for å svare på spørsmålet: *Hvilke behov har sentrale aktører for evalueringer, hvorfor bruker de evalueringer, og hvilken funksjon har evalueringer i de to sektorene og samfunnet for øvrig?* Denne analysen er gjennomført som en del av prosjektet *Kartlegging av evalueringer* for Kunnskapsdepartementet.

I tillegg til dette er det utarbeidet to øvrige notater som omhandler a) en typologisering av evalueringer og b) en oversikt over sentrale trekk ved evalueringssystemene i fire andre land og en oversikt over sentrale dimensjoner ved slike evalueringer. Disse notatene refereres til som *Internasjonal sammenligning* og *Typologisering*, og de er vedlagt denne rapporten som vedlegg 1 og 2.

1.1 Problemstillinger

Vi gir i det følgende en oversikt over hvilke hovedargumenter og dimensjoner som har kommet frem i interessentanalysen, og hvordan interessentene fordeler seg langs de aktuelle dimensjonene. Vi presenterer primært funn knyttet til vurdering av gjennomførte evalueringer, behov for og formål med evalueringer og evalueringenes funksjon og effekter. De konkrete problemstillingene som svares ut i denne rapporten er følgende (opprinnelige nummerering er her beholdt, og problemstillinger 1 og 5 er dekket av vedlegg 1 og 2):

2a) Hvilke styrker, svakheter og forbedringspunkter ser de ulike aktørene ved gjennomførte evalueringer av kvaliteten i forskning og/eller høyere utdanning?

2b) Hvordan vurderer aktørene forhold ved evalueringene som f.eks. frekvens, omfang, innretning, respondentbelastning og potensial for samordning?

3a) Ser aktørene udekkede behov for nasjonale evalueringer av kvaliteten i forskning og/eller høyere utdanning, og i så fall hvilke?

3b) Evalueringenes betydning for å skape en kultur for kvalitetsforbedring i norsk forskning og høyere utdanning

4a) Hvilke effekter ser aktørene av nasjonale evalueringer?

4b) Hvilken betydning har evalueringer i forhold til å styrke befolkningens tillit til forskning og høyere utdanning, og til å ivareta legitimiteten til disse samfunnsinstitusjonene i befolkningen?

1.2 Metode og utvalg

Interessentanalysen er basert på en 44 ekspertintervjuer samt en gjennomgang av relevante dokumenter (for eksempel nyhetssaker, pressemeldinger, o.l.) som kan knyttes til de forskjellige interessentgruppene.

Intervjuene ble gjennomført i perioden 20.3-13.8.2020. 53 representanter for totalt 34 organisasjoner er intervjuet, noe som innebærer at flere intervjuer er gjennomført som gruppeintervjuer. Intervjuene ble i hovedsak satt opp til én time.

Tabell 1: Oversikt over interessentene.

Type interessent		Antall org.	Antall pers.
<i>Myndigheter</i>	Departementer	7	12
	Dir	4	4
	Forskning	1	1
<i>Evaluatører</i>		2	2
<i>UH-institusjoner</i>	Universitet	3	11
	Høgskole	2	3
	Priv høgskole	1	3
<i>Forskningsinstitutter</i>		4	4
<i>Annen privat forskningsorganisasjon</i>		1	1
<i>Helseforetak</i>		1	2
<i>Interesseorganisasjoner (UH/Institutt)</i>		2	3
<i>Arbeidslivsorganisasjoner</i>	Arbeidsgiver	2	2
	Arbeidstaker/student	4	5
	Sum	34	53

Med bakgrunn i problemstillingene og dokumentgjennomgangen ble det utarbeidet intervjuguider for intervjuene. Disse har utgjort et grunnlag for intervjuene, uten at de har vært strengt førende. Intervjuene var semi-strukturerte, og vi både ønsket og tillot relativt stor variasjon i hvilke temaer som ble vektlagt i intervjuene med de forskjellige interessentene. Denne strategien ble valgt fordi interessentene er svært forskjellige, og fordi vi ønsket å vektle mulighetene for å fremskaffe ny og ukjent informasjon. Dermed har vi i stor grad benyttet oss av oppfølgingsspørsmål og oppmuntret interessentene til å forfølge tråder og temaer de oppfattet som spesielt interessante, og som de hadde spesiell kunnskap om.

Flere interessenter har hatt flere posisjoner (f.eks. deltaker i prosjektgruppen og departementsansatt, representant for interesseorganisasjon og institutt/UH-institusjon, o.l.). Vi har i Tabell 1 registrert disse ut fra årsaken til at vi rekrutterte dem til intervjuene, og i starten av intervjuene avklarte vi de forskjellige rollene. Vi åpnet så for at de snakket ut fra flere perspektiver, ved å oppfordre dem til selv å tydeliggjøre skiller mellom ulike perspektiver og ved at vi også spurte der dette fremsto som uklart og relevant. I tillegg har svært mange av interessentene en sammensatt bakgrunn, der mange tidligere har hatt stillinger, verv, osv. fra andre typer organisasjoner enn den de representerer i interessentanalysen. Mange som ikke nå arbeider i UH-sektoren har for eksempel en bakgrunn derfra. Der en interessent omtaler forhold som ikke er relevant for den type organisasjon de omtaler i intervjuer, behandler vi i hovedsak informasjonen som generell informasjon, uten at vi nødvendigvis knytter den til posisjonen vedkommende hadde ved intervjutidspunktet.

1.3 Oppbygging og struktur

Først presenterer vi funn knyttet til oppfattede *formål* med evalueringer i seksjon 2. Dette oppfattes som en del av vurderingen av eksisterende evalueringer og aktørenes behov for evalueringer. Vi bygger dermed videre på funnene i denne delen i de følgende delene.

I seksjon 3 svarer vi mer direkte på problemstillingene 2a og 2b, i det vi presenterer oppfatninger om evalueringenes frekvens, omfang, innretning og metodikk samt potensial for samordning av evaluering av forskning og utdanning. Dette knyttes i utgangspunktet til evalueringer som har blitt gjennomført, og blir dermed i stor grad bakoverskuende.

I Seksjon 4 fortsetter vi primært å se bakover, og vi presenterer her interessentenes oppfatninger om hvordan evalueringer brukes og hvilke effekter de har. Med dette svarer vi her på problemstillingene 3b og 4a og 4b.

I seksjon 5 presenterer vi mer generelle oppfatninger om hvilke *behov* for evalueringer aktørene har. Her har vi latt respondentene svare fritt, og dette er dermed en del der vi ser fremover og lar respondentene formidle sine ønsker og behov, uavhengig av eksisterende evalueringssystem. Her besvarer vi dermed problemstilling 3a.

I tillegg til problemstillingene som er beskrevet har vi også spurt interessentene om hvilke sentrale utviklingstrekk de oppfatter som viktige for utarbeidelsen av et fremtidig evalueringssystem. De sentrale funnene presenteres kort i seksjon 6.

Avslutningsvis oppsummerer vi hovedfunnene fra interessentanalysen i seksjon 7.

2 Formålet med evalueringer

Først presenterer vi funn knyttet til oppfattede *formål* med evalueringer. Dette undersøkes som en del av grunnlaget for interessentenes vurdering av eksisterende evalueringer og aktørenes behov for evalueringer. Denne delen omhandler kun unntaksvis de uttalte målsetningene med evalueringer, siden vi har spurt interessentene hva de *oppfatter* som formålet med evalueringer. Dette fremstår som relevant, uavhengig av om interessentene oppfatter de reelle formålene og intensjonene med evalueringene. Dersom det ikke er samsvar mellom oppfatninger og intensjoner, er dette viktig informasjon. Kun unntaksvis har vi fått svar som tyder på at evalueringer kun har *ett* formål, og de aller fleste interessentene har pekt på at flere av aspektene som fremheves nedenfor er sterkt gjeldende på én og samme tid.

Det er også en tydelig kobling mellom det oppfattede formålet med evalueringene og hvordan evalueringene *brukes*. De fleste aktørene gir svar som tyder på at formålet med evalueringene oppfattes å være i samsvar med deres bruk av dem. Det er imidlertid også en del aktører som ikke selv bruker evalueringer i spesielt stor grad, men som allikevel har en oppfatning om hva formålet er. For flere av disse aktørene kommer det tydeligere frem at de oppfatter formålet å være noe annet enn det de selv hadde hatt behov for. Kapittel 4 (Bruk og effekter av evalueringer) bygger delvis på denne gjennomgangen, og behov for evalueringer diskuteres i kapittel 5.

2.1 «Accountability» og kunnskapsinnhenting

En rekke aktører anser evalueringer som et verktøy Norsk forskningsråd (NFR) bruker for å holde mottakere av forskningsmidler *ansvarlig*. Det er spesielt representanter fra instituttsektoren som fremmer et slikt syn. De knytter det videre til at NFR har et behov for å vite – og vise Kunnskapsdepartementet (KD) – hva de får igjen for forskningsmidlene som bevilges. KD antas også å ha et behov for å vite om midlene som fordeles av NFR gir ønskede resultater. I tillegg har både NFR og KD et behov for kunnskap om status for norsk forskning, og dette fremstår som et helt sentralt formål. Også øvrige departementer er opptatt av evalueringer som ledd i ansvarliggjøring, og for å få innsikt i hva man får igjen for pengene som bevilges, spesielt gjennom program- og virkemiddelevalueringer.

På direkte spørsmål om evalueringer gjennomføres primært for NFR, KD eller institusjonene, svarer en instituttrepresentant at det er *mest* for NFR, og at evalueringene går for lite i dybden til at de blir nyttige for instituttet. I denne sammenheng vektlegger en annen interessent, som representerer en større høyskole, at evalueringer som oftest vil ha preg av *både* accountability- og læringsformål, og at det er viktig å være åpne og tydelige på hva som gjøres, og hvorfor. Enkelte aktører framhever behov for mer refleksjon, debatt og åpenhet rundt hva man ønsker å oppnå med evalueringer.

Flere aktører anser det som nødvendig, men også *problematiske*, at evalueringer brukes til å ansvarliggjøre de som mottar forskningsmidler. Aktører, primært fra arbeidslivsorganisasjoner, mener at dette er et element som følger av en New public management (NPM)-tilnærming til UH-sektoren, som kan være uheldig. Et fokus på accountability, hevdes det, signaliserer nemlig en viss mistillit til institusjoner og forskere.

Det er her flere som peker på at et fokus på effekter og bruk av forskningsmidler fører til at evalueringene får et preg av revisjon og accountability – de blir primært *summative*. Dette settes så i kontrast med evalueringer som har et *formativt* formål. Den grunnleggende forskjellen mellom formative og summative evalueringer diskuteres inngående i vedleggene *Typologisering* og *Internasjonal sammenligning*.

2.2 Benchmarking og sammenligning nasjonalt og internasjonalt

Tett knyttet til det foregående punktet er det et oppfattet formål å legge til rette for sammenligninger både nasjonalt og internasjonalt. Dette vektlegges av samtlige interessentgrupper. Men, som vi går nærmere inn på i kapittel 5 om behov for evalueringer, er det svært ulikt hvor viktig de synes sammenligning er, og hvem de som evalueres ønsker å sammenligne seg med. Større forskningsaktører tenderer mot å fokusere på internasjonal benchmarking, mens mindre aktører i større grad er opptatt av å sammenligne seg mot aktører på det nasjonale feltet.

Benchmarking nevnes av svært mange, og det er tydelig at dette oppfattes som viktig både for KD og øvrige aktører involvert i forskningspolitikk, forskningsprodusentene selv og andre brukere av forskning. Det er også verdt å ta med seg at flere aktører peker på muligheten til å sammenligne institusjoner *over tid*. Dette er spesielt viktig for å kunne vurdere kvalitetsarbeidet som følger evalueringer og resultatene av satsninger, annen innsats og prioriteringer som gjøres.

Fokus på *eksellens* trekkes frem av mange, og det ser ut til at mange nevner dette i sammenheng med benchmarking fordi de oppfatter det som en *relativ* karakteristikk. Man er eksellent om man skiller seg ut, er spesielt god og utmerker seg. Eksellens finnes dermed gjennom sammenligning.

Benchmarking og muligheten for sammenligning oppfattes som et *viktig* formål med evalueringer, og det er mange interessenter som er bekymret for å miste denne muligheten dersom det utvikles et nytt rammeverk for evaluering der man går bort fra de brede og nasjonale evalueringene som muliggjør benchmarking.

2.3 Kvalitetsutvikling/læring

En rekke aktører, spesielt fra UH- og instituttsektoren, ser også på kvalitetsutvikling og læring som tydelige formål med evalueringer. Som vi senere kommer tilbake til, er det forskjellige oppfatninger om hvor sterkt det formative elementet i større nasjonale evalueringer egentlig er, men et *visst* formativt element er nær sagt alle enige om at de har. Store aktører mener kvalitetsutvikling *bør* være det sentrale, og mange mener også at evalueringene har hatt klare positive konsekvenser for kvalitet i egen organisasjon. Andre igjen påpeker at det er viktig å sørge for at institusjonene har et eierskap til evalueringene. De må ha «en hånd med på rattet» om man skal klare å skape gode læringsprosesser. Dette understrekes også av arbeidslivsorganisasjoner, som vektlegger det faglige perspektivet og involvering av forskere selv. Det er viktig, sier de, at evalueringer ikke blir en ren administrativ øvelse.

På spørsmål om *hvem* som skal lære eller utvikles, er svaret at både fagmiljøene og institusjonene både *bør* og kan få utbytte av evalueringene. Her er det også flere som peker på at det er ønskelig at evalueringene ikke blir et rent øyeblikksbilde av situasjonen, men at de også følges opp og blir en del av en *kvalitetskultur*, med klare rutiner og prosedyrer for oppfølging og støtte i etterkant. Dette kommer vi tilbake til når vi omtaler effekter av evalueringer i seksjon 4.2. Det er imidlertid også klare læringseffekter av evalueringer som *ikke* følges opp slik man ideelt sett kunne ønske, og det er spesielt større universiteter og forskningsinstitusjoner som vektlegger at de kan lære, og har lært, mye av de store nasjonale forskningsevalueringene.

Det poengteres også at én måte å få til endring på er å gjennomføre *summative* evalueringer som beskriver tilstand, hvorpå institusjonene selv behandler den informasjonen de får og utvikler strategier og iverksetter endringsprosesser. Dette

innebærer at det er mulig å hevde at formålet med en evaluering er å skape endring, selv om evalueringen har et summativt preg. Dette støttes av flere, som også peker på at man *må* ha en viss konkurranse, sammenligning, og andre typisk summative elementer, for å bli bedre – *rene* formative evalueringer oppfattes ikke av alle som den beste veien mot forbedringer for sektoren. Dette er det imidlertid uenighet om, og flere mener for eksempel at når en evaluering knyttes til ekstra midler, eller at det oppfattes som viktig å få en god evaluering, faller det formative elementet i stor grad bort. Andre påpeker i denne sammenheng at summative evalueringer oppleves som noe «farligere» enn formative evalueringer, og det oppstår dermed en fare for at evaluanden er mindre ærlig om formålet oppfattes som summativt.

I denne sammenheng oppfatter flere interessenter det norske systemet som en tydelig kontrast til for eksempel det britiske. I Norge, hevdes det, er formålet å forbedre kvalitet, og ikke å måle og overvåke. I Storbritannia er det sterkere konkurranse mellom institusjoner og et til tider desperat «overgangsmarked for professorer», og gode evalueringer er viktige for å vinne fram.

2.4 Forskningspolitikk, prioritering og strategiske grep

Et annet sentralt formål som omtales av mange interessentgrupper er evalueringenes funksjon i utvikling av forskningspolitikk. Det er primært KD og NFR som nevnes som aktører med behov for dette, men også andre aktører som finansierer og arbeider med forskningspolitikk vektlegger dette. Representanter for KD bekrefter også at dette aspektet ved evalueringer er svært viktig. Evalueringer muliggjør prioriteringer og gir nødvendig kunnskap for å kunne ta strategiske grep i forskningssektoren.

Samtidig er det flere UH-institusjoner som peker på at denne type kunnskap også kan og bør utnyttes strategisk av institusjonene selv. Nasjonale evalueringer gir oversikt over forskjellige fagmiljøer, og kvaliteten på disse. Dette gjør det mulig å finne samarbeidspartnere, koordinere innsats, og også sammen finne frem til gunstige muligheter for differensiering og spesialisering. Av UH-institusjonene og øvrige forskningsprodusenter fremstår de store aktørene som mest opptatt av dette formålet med evalueringer.

2.5 Formålet med utdanningsevaluering

Utdanningsevalueringer omtales av relativt få aktører, og det er primært KD og aktører i UH-sektoren som har klare formeninger om formålet med utdanningsevaluering. De fleste forholder seg til dagens situasjon og NOKUTs virke, og kvalitetsarbeid gjennom tilsyn og utvikling av kvalitetssikringssystemer fremstår som helt sentralt.

Videre knyttes utdanningsevalueringer til et ønske om å vite om studentene får det de har krav på, eller blir lovet. Denne måten å uttrykke formålet med høyere utdanning på innebærer et *klient/bruker*-syn på studenter, noe som preges av brukere som rettighetshavere med et krav overfor institusjoner.

Det er imidlertid også mange aktører som her kommer inn på institusjonenes evne og forutsetninger for å drive egenutvikling. Dette oppfattes som et klart formål, og mange av aktørene er også tydelige på at NOKUT faktisk bidrar positivt i dette henseende. Flere sier at evalueringer skal bidra til å belyse hvilke utfordringer og hvilket utviklingsbehov institusjonene har. NOKUTs rolle oppfattes i hovedsak som klart formativt, konstruktivt, og nyttig.

Det kommer klart til uttrykk at aktørene i UH-sektoren mener at NOKUT står overfor en vanskelig balansegang mellom å være tilsynsorgan og samtidig yte hjelp og være en støttespiller. En UH-representant illustrer dette med et eksempel på hvordan forholdet til NOKUT tidvis oppfattes: «først var dere vennen vår, og nå kommer dere og skal kreve». Det er imidlertid en rekke aktører som er tett på utdanningsvirksomhet og NOKUT som forteller at NOKUT har truffet en god balanse. De forteller videre at NOKUT over tid har beveget seg mer mot å bli en støttespiller enn en ren kontrollør. Det finnes imidlertid også representanter for det motsatte synet. Det er to ting de er redd for, sier interessenten, og det er «NOKUT og riksrevisjonen». Dette knyttes videre til en oppfatning av NOKUT som et strengt og relativt rigid kontrollorgan med strenge og «harde» parametere, som ikke nødvendigvis gjenspeiler det interessenten oppfatter som sentralt for kvalitet i utdanningen. Andre forteller om et visst ubehag og en usikkerhet knyttet til det å bli «kikkert i kortene», og dette knyttes til en viss opplevd mistillit. På dette feltet forteller flere om en forskjell i oppfatningen av NOKUT mellom administrasjon og fagansatte. Administrativt ansatte, hevdes det, har større forståelse for, og ser mer direkte nytten av, det arbeidet NOKUT gjør, enn de fagansatte, som i større grad oppfatter tilsyn o.l. som utidig og tidkrevende byråkratisering av deres virke.

Oppsummering

Sentrale motsetninger og aktører med spesielle interesser

Summativ vs. formativ evaluering	KD, evaluatorene og forskningsprodusenter
Accountability vs. læring	Departementer, NFR, forskningsprodusenter
Institusjonsfokus vs. nasjonalt nivå	NFR, UH-institusjoner og KD
Konkurransen og sammenligning vs. dyp forståelse og aktiv bistand	Arbeidslivsorganisasjoner, departementer, evaluatorene og forskningsprodusenter

Hovedpoenger

- Både summativ og formative elementer er helt sentrale evalueringsformål
- Evalueringer har flere formål, og det er primært snakk om hvordan de forskjellige vektet
- Det er viktig å være tydelig på hva formålet med evalueringene er, og at dette kommuniseres
- NOKUT har en posisjon som innebærer en spenning mellom rollen som kontrollør og veileder

3 Vurdering av tidligere evalueringer

I dette kapitlet presenterer vi oppfatninger om frekvens, omfang, innretning, metodikk og potensial for samordning av evaluering av forskning og utdanning. Dette knyttes eksplisitt til evalueringer som har blitt gjennomført, og blir dermed i stor grad bakoverskuende.

Så langt har det blitt gjennomført langt flere store, nasjonale evalueringer av forskning enn av høyere utdanning. De fleste uttalelsene som blir presentert i dette kapitlet er derfor knyttet til evalueringer av forskning utført av NFR. Dersom vi omtaler utdanningsevalueringer vil vi bemerke dette, og vi vil også presentere vurderinger av noen konkrete utdanningsevalueringer gjennomført av NOKUT. Likevel dreier mange av poengene i dette kapitlet seg i stor grad om evalueringer generelt og hvordan «evalueringsregimet» som helhet oppleves av institusjonene.

Vi presenterer først synspunkter på frekvens og periodisering, omfang og belastning samt innretning og metodikk, primært rettet mot forskningsevalueringer. Deretter presenterer vi noen vurderinger av tidligere utdanningsevalueringer. Til slutt oppsummerer vi kort interessentenes syn på muligheten for å samordne forsknings- og utdanningsevalueringer.

3.1 Frekvens og periodisering

I vurderingen av eksisterende evalueringer er det svært få som opplever at det evalueres *for ofte*. Dette henger imidlertid sammen med opplevd nytte av evalueringene, som vi diskuterer i neste punkt. Bekymringer knyttet til for høy evalueringsfrekvens diskuteres i hovedsak i forbindelse med studentperspektivet, da flere aktører peker på at studenter samlet sett er utsatt for et stort undersøkelsestrykk. Dette gjelder imidlertid alle type undersøkelser, og ikke evalueringer spesielt, selv om evalueringer kan utgjøre en del av dette trykket. Samtidig er det også flere aktører som mener at vi har «ekstremt mange evalueringssystemer» og «et voldsomt måleregime» i sektoren. Det fremstår for noen som et problem at arbeid med å gjennomføre evalueringer og rapportere resultater «til departementet [KD]» kommer i veien for arbeidet med å faktisk forbedre kvaliteten i forskning og utdanning. Dette er dog et synspunkt som ikke deles av mange.

Det som derimot kommer meget tydelig frem fra en rekke aktører, spesielt fra KD og forskningsprodusenter (men også øvrige departementer og aktører), er en bekymring over at man ikke har *regelmessige* evalueringer med relativt faste og kjente intervaller. Dette diskuteres primært på forskningssiden. De som er mest opptatt av regelmessige evalueringer kobler ofte dette til de tradisjonelle *fagevalueringene* som har blitt gjennomført av NFR. Noen diskuterer intervallene mellom evalueringer, og i forbindelse med de store fagevalueringene omtales alt mellom 5- og 10 års-sykluser som en hensiktsmessig frekvens. Det er bred enighet om at oftere enn hvert femte år hadde blitt for ofte, da man hverken har behov for det, eller får muligheten til å vurdere utviklingen siden forrige evaluering. Samtidig pekes det på at det avhenger av hvor omfattende evalueringene er. Evalueringene av biologi, medisin og helseforskning, og fysikk, omtales som eksempler på store evalueringer som naturlig peker mer mot et intervall nærmere 10 enn 5 år.

Vi vil ved flere anledninger komme tilbake til at en rekke aktører peker på at rapportering på indikatorer osv. kan, og kanskje bør, håndteres separat fra de store evalueringene. Indikatorrapportering og analyse kan være et nærmest kontinuerlig arbeid eller i alle fall ha en langt høyere frekvens enn evalueringer basert på andre grunnlag. Dette omtales nærmere i seksjon 5.4.

NFRs kalender for kommende evalueringer er for øyeblikket tom, og dette fremstår som meget problematisk for en rekke interessenter. Etter flere runder med fagevalueringer har det blant flere oppstått en forventning om at dette skulle fortsette, og mange har allerede forberedt seg på, og venter på, en tredje runde med fagevalueringer. Biologi og kjemi omtales som eksempler på fagområder der man har forventet og opplever behov for neste runde med evalueringer. Uklarhet rundt hva som skjer videre skaper en viss usikkerhet og misnøye i flere miljøer. Behovet for forutsigbarhet og et ønske om regelmessige evalueringer fremkommer aller tydeligst hos flere av landets største forskningsprodusenter, spesielt de som er knyttet til naturvitenskapelige miljøer, samt hos representanter for KD. Det er også en viss misnøye med kommunikasjonen rundt denne type evalueringer. Etter to runder med evaluering opplever flere at det har blitt skapt en slags forventning om hvordan evalueringer skal gjennomføres. Når dette endres uten hverken dialog eller tydelig signaler skaper dette en viss frustrasjon. Behovet for forutsigbarhet drøftes nærmere i seksjon 5.5.

Et relatert poeng er *hvem* som skal avgjøre hvilke evalueringer som skal gjennomføres, og når. Her er det delte meninger, der en del vektlegger NFRs frie rolle og kompetanse, mens andre igjen ytrer et ønske om at evalueringssystemet som helhet tas «et nivå opp», og at KD, eller andre, i større grad sikrer forutsigbarhet og *koordinering* av store nasjonale evalueringer (se nærmere diskusjon om dette i seksjon 5.5).

3.2 Omfang og belastning

Et annet punkt som skulle belyses var oppfatninger om omfanget av evalueringene og belastningen evalueringer medfører. Dette er i all hovedsak relevant i forbindelse med UH-institusjoner, institutter og helseforetak, og det er deres oppfatninger som her fremkommer.

Store evalueringer, som for eksempel fagevalueringene, er omfattende, og innebærer mye arbeid for både evaluator og evaluand. Det fremkommer imidlertid meget tydelig at belastning alltid vurderes opp mot oppfattet *nytte* av evalueringene. Alle de vi har intervjuet erkjenner lett at evalueringene er omfangsrike, og at det medfører relativt høy arbeidsbelastning. Det er imidlertid like tydelig at nær sagt alle oppfatter omfanget og belastningen som *rimelig* og *nødvendig*. Samtidig er det her en vanskelig balanse. Når evalueringer sees sammen med øvrige rapporteringsplikter og undersøkelser kan det tenkes at man bryter balansen og ender opp med et samlet undersøkelses- og måleregime flere oppfatter som *for* omfattende. Dette må man ta hensyn til i koordineringen og ved utforming av en strategi.

Det er i det hele tatt et overordnet inntrykk at eksisterende evalueringer har ført til en god balanse mellom nytte og innsats. Ingen mener at evalueringene har vært for *lite* krevende, men det er altså også få som mener at evalueringene har vært for omfattende. Dette skyldes at institusjonene har sett en klar nytte og effekt av evalueringene.

Det er imidlertid flere grupper innad i hver institusjon, og det kan tenkes at for eksempel administrativt ansatte har et annet syn på belastningen enn de fagansatte. Vi har derfor intervjuet representanter for fagansatte i tillegg til et større antall administrativt ansatte og ledere for forskningsadministrasjonen i institusjonene. Flere representanter fra UH-sektoren uttaler at administrasjonen *tar*, og *bør ta*, en stor del av arbeidet, slik at evalueringer ikke i for stor grad går på bekostning av forskningen og undervisningen til de fagansatte. Denne bekymringen deles av studentrepresentanter. Representanter for faglige ansatte peker imidlertid på at dersom det faktisk er det *faglige* som skal evalueres, er det uunngåelig at fagansatte involveres i evalueringene.

Vi finner få indikasjoner på at det er store forskjeller i oppfatning mellom disse to gruppene, og de av interessentene som har vært involvert som fagansatte fremmer også et syn som indikerer at arbeidsbelastningen forbundet med evalueringer har vært tung, men akseptabel.

Til tross for at det kan se ut til at man har truffet godt med evalueringene som har blitt gjennomført, vil vi understreke nettopp den *balansen* vi tidligere har nevnt. Mange av respondentene oppfatter dette som en hårfin balanse, og de vektlegger at det ikke er rom for *mer* omfattende, eller flere, evalueringer. Når de uttrykker at belastningen er rimelig, sier de fleste samtidig at den er så stor at det ville blitt meget problematisk om de skulle gjøre mer, eller om for eksempel evalueringer av forskning og utdanning i sammenheng kom i *tillegg* til evalueringene de allerede kjenner.

Kostnadsspørsmålet understrekes også av flere. Det er svært kostbart å gjennomføre store evalueringer, og det vil alltid være behov for å vurdere kostnader opp mot nytte. Flere omtaler SAMEVAL som en svært kostbar evaluering, og mye arbeid gikk med til gjennomføringen av denne, både for panelene og for ansatte ved institusjonene. Kost/nytte knyttes her til et spørsmål om det er NFR eller institusjonene selv som bør være i førersetet i slike prosesser, der noen mener at institusjonene hadde fått mer igjen for arbeidet om de selv var mer involverte i prosessene.

3.3 Innretning og metodikk

Spørsmål knyttet til *innretning* knytter seg naturlig til oppfatninger om formålet med evalueringer og hvilke behov aktørene har. Evalueringenes formål ble belyst i kapittel 2, og aktørenes behov for evalueringer blir gjennomgått i kapittel 5. Vi diskuterer her innretning og metodikk på et mer overordnet nivå, og tar deretter disse diskusjonene videre i påfølgende diskusjoner om hvordan evalueringer brukes, hvilke effekter de har og hvilke behov aktørene på feltet har.

I mange intervjuer uttrykkes en bekymring for at det *summative* elementet i evalueringene vektlegges for sterkt og at det er et begrenset fokus på å komme med hjelp, forslag og anbefalinger som oppfattes som konkrete og grundige nok til at de blir nyttige for institusjonene. Hovedpoenget, som fremheves av mange interessenter, er at evalueringer som sier noe om *hvordan* ting står til, uten å bidra til å skape klarhet i *hvorfor* det er slik, er av noe begrenset nytte.

Helt generelt er det tydelig at mange oppfatter det kvalitative elementet med internasjonale paneler som svært viktig, nødvendig og hensiktsmessig. I tillegg oppfattes egevaluering som nyttig og viktig. Samtidig er det viktig at panelene består av personer med kjennskap til norske systemer. Her peker forskjellige interessenter både på særnorske forhold i primærnæringene, instituttsektoren generelt, men også den norske modellen i arbeidslivet. I tillegg til den faglige legitimiteten, som mange fra UH- og instituttsektoren vektlegger i denne forbindelse, understreker flere forvaltningsorganer at eksterne internasjonale paneler gjør at evalueringene får mer legitimitet *generelt*, ved at de gjennomføres av andre enn de som selv forsker og underviser.

I omtalen av instituttevalueringer er det flere som opplever at bibliografi har fått for mye oppmerksomhet, at man savner skikkelige dypdykk i fagområdene som dekkes og til dels at panelene ikke kjente til særnorske forhold i tilstrekkelig grad.

Flere ytrer også et ønske om *mer* bruk av egevaluering. Dette, sier de, medfører mer jobb, men det vil samtidig gjøre evalueringene enda mer nyttige for evaluanden. Her er det imidlertid andre interessenter som opplever egevalueringer som *for* omfattende

allerede, og som også opplever at mye av informasjonen man gir i egevalueringene ikke blir *anvendt* i stor nok grad. Det oppleves at det spørres om ting «for sikkerhets skyld». En av interessentene påpeker også at man må prøve å finne en god balanse mellom selvkritikk og selvskryt for at evalueringene skal bli nyttige. Som vi senere kommer tilbake til, er det imidlertid flere som er tydelige på at det oppfattes som vanskelig å bruke egevalueringer strategisk for å forskjønne situasjonen. Det kommer også frem at flere oppfatter egevalueringer som noe *ve/førende*, og at de føler at noe av årsaken til at evalueringene til dels gir lite ny informasjon er at de i stor grad gjenspeiler egevalueringene som utarbeides.

Et annet poeng i denne sammenheng er at det kan være en viss forskjell mellom NOKUT og NFR sine evalueringer, og hvordan bestillingene fra disse besvares av institusjonene. NOKUT gjennomfører også mindre evalueringer, for eksempel med relevans for arbeidslivet, og også disse er i stor grad basert på egevaluering. I denne sammenheng oppleves det som at institusjonene tar det mindre alvorlig, og i mindre grad responderer, enn hva tilfellet er på forskningssiden.

Selvrapportering i form av kvantitative indikatorer er imidlertid et aspekt mange er opptatt av, og det er en viss uenighet om disse bør være en sentral del av selve evalueringene, eller om løpende indikatorer og statusrapportering til dels kunne utgjøre et *separat* system. Dette kommer vi nærmere inn på i seksjon 5.4.

Et annet tema som diskuteres av en rekke aktører er SAMEVAL og fokuset på *forskergrupper*. Mens man i en del organisasjoner allerede arbeidet ut fra forskergrupper som passet godt med evalueringssopplegget, passet dette langt dårligere for andre organisasjoner. Resultatet var da at noen meldte inn eksisterende og aktive forskergrupper, mens andre institusjoner i større grad konstruerte grupper som passet til malene eller måtte være kreative for å få eksisterende grupper til å passe med malen. Dette oppleves som en utfordring. Samtidig oppleves det som negativt at institusjonene selv velger hvilke grupper som meldes inn. Mens noen organisasjoner forteller om et ønske om å bli evaluert fullstendig, hvorpå de meldte inn alle sine grupper, har andre i større grad meldt inn konstellasjoner av ansatte som ikke normalt arbeider sammen som forskergrupper. Man har også vært nødt til å velge om man skal melde inn miljøer som man frykter vil prestere dårlig. Dette oppleves ikke som en alvorlig mangel med evalueringer som SAMEVAL, men som en utfordring.

Det er videre mange fra Mat-Nat-miljøer ved flere organisasjoner som opplever at evalueringene de kjenner til har truffet meget godt, og at de opplever at man har vært i stand til å identifisere og beskrive sterke og svake miljøer på en god og nyttig måte.

Avslutningsvis kommer det også frem at fagforeninger etterlyser et sterkere ansattfokus i evalueringene, og generelt i arbeidet med kvalitet i forskning og utdanning. Evalueringer gjennomføres ofte på et nasjonalt nivå, og det er en oppfatning om at fagansatte i liten grad a) deltar i utforming av og gjennomføring av evalueringer, b) blir involvert i prosessene og c) omtales i oppsummeringer, konklusjoner og anbefalinger.

3.4 Hvem eller hva evalueres?

Aktørene på forskningsfeltet er til dels svært forskjellige, og *én* innretning derfor vanskelig kan treffe alle aktører like godt. De sentrale forskjellene vi har identifisert er mellom store og små UH-institusjoner, private og offentlige UH-institusjoner, forskjellige fagområder og tradisjoner og UH-institusjoner sammenlignet med institutter og helseforetak. I tillegg er det en rekke andre variabler som gjør at også instituttene skiller seg sterkt fra hverandre på flere områder. Det samme gjør for eksempel offentlige høyskoler med forskjellige strategier og ambisjoner.

Store og små UH-institusjoner er den første dimensjonen. Denne sammenfaller også til dels med forskjellen mellom høyskoler og universiteter. Spesielt de mindre institusjonene påpeker at det er noe problematisk om de skal vurderes etter samme mal som de største universitetene, og at *eksellens* skal være målet for alle aktører. Dette knyttes også forskjeller i samfunnsoppdrag og mandater. Noen institusjoner har for eksempel et tydelig oppdrag og en ambisjon om å være en pådriver for *regional* utvikling og kunnskapsdeling. Det kan være problematisk å kombinere dette med en ambisjon om å utvikle fag- og forskermiljøer i internasjonal toppklasse. Samtidig er det også tydelige forskjeller mellom universiteter som har sterkt fokus på anvendelse av forskning, innovasjon og samfunnskontakt og de som fokuser mer på grunnforskning. Avhengig av innretningen på evalueringene, og for eksempel betydningen av *impact* (akademisk eller samfunnsmessig), vil de ikke treffe alle institusjonenes mål like godt.

En annen sentral forskjell er mellom UH-institusjoner og institutter. Flere representanter for instituttsektoren opplever til dels at de evalueres på *universitetenes* premisser, og at dette i visse tilfeller slår uheldig ut for dem. Noen peker på at en årsak til dette er at mye av forskningen som produseres i instituttene er oppdragsforskning som ikke alltid er rettet mot publisering i de kanalene som teller mest i evalueringer. Andre igjen påpeker at de har svært aktivt – og eksellent – forskning, men at evalueringene allikevel ikke «treffer» helt.

Flere instituttrepresentanter er tydelige på at visse NFR-evalueringer oppleves som noe problematiske for dem, fordi de opplever at evalueringene i for stor grad gjennomføres på UH-sektorens premisser. Evalueringene treffer UH-sektoren godt, og også grunnforskningsorienterte institutter, men de er noe mindre treffsikre når det gjelder en del av instituttene som primært arbeider med oppdragsforskning og anvendt forskning. En svært viktig forutsetning for å kunne evaluere instituttene på en god måte er å sette sammen paneler som forstår, eller får tilstrekkelig innføring i, hvordan den norske instituttsektoren fungerer. Dette har til dels vært en utfordring, men flere forteller også at de mener at NFR har klart å gjøre dette på en god måte. Det er med andre ord noe delte meninger om dette temaet blant instituttrepresentantene.

Samtidig oppfattes disiplin fokuset i evalueringene som best tilpasset (deler av) UH-sektoren, noe som medfører at man i flere evalueringer har fått stor påmelding fra universiteter, mens institutter *ikke* melder seg på. Dette kan også gjelde andre veien, ved at universiteter ikke melder seg på dersom de har organisert forskningen etter disiplin, og ikke fagområde. Dette innebærer at flere miljøer *ikke* deltar i evalueringer (også fra universiteter som ikke har organisert et fagområde etter disiplin), men at det i evalueringene konkluderes *som om* hele feltet i Norge har blitt evaluert. Dette fører tidvis til kortslutninger, hevder noen, og at konklusjonene og anbefalingene ikke gjenspeiler virkeligheten.

Dette fremstår som en bekymring som gjør seg *mest* (men ikke utelukkende) gjeldende i instituttsektoren. Dette knyttes også til diskusjonen om *tverrfaglighet*, som vi diskuterer videre i seksjon 6.2.

I denne sammenheng stiller også en representant fra UH-sektoren spørsmålet om de private institusjonene og instituttene i større grad enn offentlige UH-institusjoner er avhengig av å lykkes i *markedet*, og dermed har et større insentiv til å tilnærme seg evalueringene strategisk. Dette kan medføre at man kun melder inn grupper som forventes å evalueres positivt, at man er mindre ærlig i egne evalueringer og mindre villige til å blottlegge svake sider. Dette benektes av institusjonene selv. Både representanter for institutter og UH-sektoren mener generelt at det er meget vanskelig å slippe unna med uriktige, eller forskjønnede, fremstillinger i egne evalueringer. Her fremmes de internasjonale panelene som viktige, og det eksisterer en oppfatning om at

slike paneler i all hovedsak vil gjennomskue forsøk på strategisk og manipulativ aktivitet i egnevalueringer. Samtidig er det altså flere som har reagert på at det nærmest er mulig å «gjemme» fagmiljøer man ikke ønsker evaluert, ved at institusjonene selv velger hvilke miljøer som meldes inn.

Forskjeller mellom *fagområder* og disipliner ser også ut til å henge sammen med forskjeller i oppfatninger av hvor hensiktsmessig innretningen av evalueringer er. Dette henger til dels sammen med forskjellen mellom institusjoner, der for eksempel mindre høyskoler med mange profesjonsutdanninger kan oppleve at evalueringene i større grad er tilpasset disiplinaryfag og mer forskningsbaserte breddeuniversiteter. Flere mener at de store fagevalueringene passer mindre godt til evaluering av de mindre UH-institusjonene. Dette gjelder spesielt når evalueringene er tydelig disiplinorienterte, mens en del institusjoner har et sterkt profesjonsmiljø.

Det er noe uenighet om hvor store forskjeller det er mellom forskjellige fagområder, men vi har funnet en viss tendens til å oppfatte evalueringene som har blitt gjennomført som *best* egnet til naturvitenskapelige fag, og at for eksempel sosiologi tilhører en tradisjon og en vitenskapelig tilnærming som til dels gjør denne tilnærmingen vanskelig. Dette kom tydelig frem i debatten etter SAMEVAL. Det er imidlertid andre interessenter som tar opp nettopp dette, og mener at denne innvendingen i stor grad skyldes at man ikke er *vant til* den type evalueringer, og at også disse fagmiljøene opplever det som mindre problematisk etter å ha vært gjennom det (gjørne flere ganger).

Til sist er det også viktig å peke på helseforetakene og samarbeidet disse har med universiteter, som viktige aspekter evaluator må ta hensyn til. Flere påpeker at organiseringen av samarbeidet mellom universitetssykehus og universitet gjøres på forskjellige måter. Det er helt nødvendig å forstå, og ta høyde for, disse forskjellene (og helheten) for å kunne evaluere for eksempel forskningen ved et helseforetak. Selv om St. Olav og NTNU har et samarbeid som tilsynelatende kan ligne samarbeidet mellom Oslo Universitetssykehus (OUS) og UiO, er det store forskjeller i strukturene i disse samarbeidene. Velger man å sammenligne St. Olav og OUS uten å for eksempel inkludere og ta inn over seg aktiviteten som foregår i universitetene, vil sammenligningen lett kunne bli meningsløs.

Et viktig valg blir dermed om forskjellige organisasjoner skal evalueres på *samme* premisser, eller om det er nødvendig å evaluere dem på sine *egne* premisser. Dette diskuteres også i flere sammenhenger nedenfor, og vi viser også til notatet *Internasjonal sammenligning* for en mer teoretisk gjennomgang av innretninger som sikter mot enten det *spesielle* eller det *universelle*.

3.5 Vurdering av utdanningsevalueringer

De fleste av interessentene har en langt tydeligere relasjon for forskningsevaluering enn til utdanningsevalueringer. Dette har medført at vi har fokusert klart mest på dette i mange av intervjuene som er gjennomført. Vi har imidlertid også interessenter som primært arbeider med utdanningsfeltet og utdanningskvalitet, og vi vil her ta for oss noe av det disse har fortalt. Vi understreker at det her er snakk om et begrenset antall informanter, og at det her er noe større usikkerhet forbundet med graden av bredden av synspunkter som er fanget opp.

Som tidligere nevnt, består NOKUTs virksomhet av til dels ganske ulike former for kvalitetssikring og kvalitetsvurdering. Primært arbeider de med tilsyn og akkreditering, samt analyser og data knyttet til Studiebarometeret. I tillegg har de gjennomført et begrenset antall evalueringer av utdanninger innen konkrete fagområder (som vi her kaller fagevalueringer), samt noen mindre evalueringer av konkrete temaer, som for

eksempel arbeidslivsrelevans i utdanningen. NOKUTs tilsynsarbeid er forholdsvis periodisk og *kontinuerlig* og godt innarbeidet i sektoren. Det samme gjelder Studiebarometeret, som kommer en gang i året og brukes relativt aktivt av både studenter og institusjonene. Det fremkommer få negative synspunkter på frekvensen av utdanningsevalueringer, selv om noen aktører diskuterer alternative evalueringssopplegg for utdanning som for eksempel ligner mer på NFRs fagevalueringer. Dette diskuterer vi videre i seksjon 5.2.

Selv om NOKUT selv ikke ser på tilsyn og akkreditering som en form for evaluering, mener de fleste av informantene at de faktisk innebærer en stor grad av evaluering og kvalitetsvurdering. Dette har en klar positiv effekt for institusjonene – både ved at de får en vurdering og en tilbakemelding og ved at et periodisk tilsyn tvinger dem til å ta kvalitetsarbeidet på alvor og aktivt bruke kvalitetssikringssystemene som er på plass. Som én informant formulerer det: «NOKUT-tilsynene fører helt opplagt til at institusjonene strammer opp kvalitetssystemene sine. Ingen tvil. Og de som påstår annet, de lyver». Likevel er det et par informanter som mener tilsynene er litt overfladiske – litt som å «telle hoder» – og at de er krevende administrativt. Det blir også meldt om at det kan være vanskelig å overbevise faglig ansatte om verdien av et slik kvalitetssystem med mange byråkratiske leveranser. Alt i alt virker informantene fornøyde med NOKUTs tilsyns- og akkrediteringsarbeid, særlig ettersom mange opplever at NOKUT har beveget seg noe fra «kontrollør» til «samarbeidspartner» i kvalitetsarbeidet.

NOKUT anser ikke Studiebarometeret som en evaluering. Likevel er det denne som nevnes av flest informanter når vi spør om deres kjennskap til NOKUTs arbeid innen utdanningsevaluering. De fleste informantene anser denne undersøkelsen som «en del av sektoren», og at den kan være nyttig i kvalitetsarbeidet – til en viss grad. Flere av informantene understreker at den har noen metodiske svakheter, og de fleste er enige om at undersøkelsen ikke er et godt nok mål på utdanningskvalitet alene. Det er også flere som trekker frem betydningen av tidspunkt for utsendelse av undersøkelsen, med tanke på hvor studentene er i studieløpet og når i semesteret de mottar undersøkelsen. En informant understreker at det her er mye nyttig lokal kunnskap ute i institusjonene som NOKUT ikke har – for eksempel når ulike utdanningsprogram har praksisperioder. Man har tidligere opplevd at dersom undersøkelsen sendes ut mens studenter er i praksis, får man langt lavere svarprosent. Generelt virker oppfattelsen blant informantene likevel å være at Studiebarometeret kan være en god pekepinn på hvilken vei utviklingen går og hvilke områder det er nødvendig å jobbe mer med.

Av NOKUTs evalueringer og analyser knyttet til konkrete temaer, blir kartleggingen av læringsutbyttebeskrivelser som ble gjennomført i 2015 nevnt av enkelte informanter. Det er imidlertid svært delte meninger rundt nytten av denne kartleggingen. På den ene siden mener én informant at det var mye læring i prosessen, hvor de både fikk et dypdykk i sine egne læringsutbyttebeskrivelser og ble presentert for gode eksempler på fagspesifikke læringsutbyttebeskrivelser. En annen informant er derimot kritisk til krav til utforming av læringsutbyttebeskrivelser og mener at de ofte blir litt virkelighetsfjerne og at man nærmest må «ha en doktorgrad for å tolke dem selv». Ifølge denne informanten fører dette til dobbeltarbeid for underviserne. Først må de tenke gjennom innholdet i kurset man skal gjennomføre og formulere noen mål, deretter må man «oversette» dette til de riktige ordene fra en tillatt ordliste. Når man så skal presentere det til studentene er man nødt til å oversette tilbake for at studentene skal forstå. Vedkommende mener videre at NOKUTs kartlegging av læringsutbyttebeskrivelser ikke førte til noen større forståelse for eller oppklaring av hvordan læringsutbyttebeskrivelser bør utformes.

Generelt virker det som om NOKUTs fagevalueringer har blitt tatt godt imot i sektoren, og har hatt relativt store konsekvenser. Flere informanter etterlyser flere evalueringer og

mener det er lenge siden sist, og andre uttrykker at de gleder seg til kommende evalueringer, for eksempel den pågående evalueringen av lektorutdanningene.

Flere av informantene understreker at også innen utdanningsevalueringer er det store forskjeller mellom ulike institusjoner og mellom ulike typer utdanninger. For eksempel er det stor forskjell mellom masterprogrammer på en del høyskoler der studentene tar graden på deltid mens de er i jobb og fulltids masterprogrammer på universiteter. Her er det store forskjeller i hvordan studieprogrammene er lagt opp, hvilke forutsetninger institusjonene har og studentenes behov. Dette bør dermed tas hensyn til i evalueringer.

I tillegg blir enkelte profesjonsutdanninger trukket frem som et eksempel på utdanninger som krever evalueringer med en litt annen innretning. En grunn til dette er at profesjonsutdanningene skal utdanne studenter som ofte skal ivareta sentrale funksjoner i samfunnet, hvor målet med utdanningen ikke nødvendigvis er eksellens og et høyt internasjonalt nivå. Ofte befinner profesjonene seg derimot i en særegen nasjonal kontekst. Én informant nevner for eksempel lærerutdanningene, og at disse ikke behøver å være «internasjonalt fremragende» og skape gjennombrudd i forskningen. De må heller fokusere på å skape en «god utdanning av fremtidens lærere».

En annen grunn til at særlig enkelte profesjonsutdanninger skiller seg ut er at de ofte er svært verdiladede. Dette er også knyttet til at utdanningene skal sikre nyutdannede som kan ivareta sentrale samfunnsfunksjoner. I denne sammenheng ble allmennlærerutdanningen nevnt. Den ble evaluert av NOKUT i 2006. Lærerutdanningene blir av én informant omtalt som svært «verdiladet og viktig», ettersom lærerne skal ivareta målsettingen om den norske enhetsskolen. Dette må reflekteres i evalueringen.

3.6 Samordning av evaluering av forskning og utdanning

I tillegg til å spørre interessentene om evaluering av forskning og utdanning som separate aktiviteter, spurte vi også om muligheten for å samordne disse evalueringene. Vi vil her bemerke at de fleste interessentene utenfor UH-sektoren har lite kontakt med utdanningsvirksomhet. Svarene som følger baseres dermed primært på interessenter fra UH-sektoren. Øvrige interessenter har imidlertid også klare oppfatninger om *at* det er viktig å se utdanning og forskning i sammenheng, uten at de har spesielt klare oppfatninger om *hvordan* dette skal gjennomføres.

3.6.1 Behov, fordeler og mulige utfordringer

På spørsmål om hvorvidt det ville være mulig å evaluere forskning og utdanning sammen, svarer de aller fleste aktørene at dette ville være positivt og interessant, men at det er mange utfordringer forbundet med praktisk gjennomføring. På den ene siden peker flere på at forskning og utdanning er to forskjellige verdener, men på den andre siden understreker flere at universiteter og høyskoler har både forskning og utdanning som en del av sitt samfunnsoppdrag, og at det er en del sammenhenger mellom de to virksomhetene som kunne og burde vært undersøkt grundigere.

Et eksempel som blir trukket frem er evalueringen av nordiskfaget i 2005. Her ble det først levert en evalueringsrapport, og deretter opprettet et oppfølgingsutvalg. Én informant forteller at oppfølgingsgruppen opplevde det som svært utfordrende å følge opp evalueringen ettersom den kun dreide seg om forskning. Dette ble utilstrekkelig i et fag som i stor grad retter seg mot skolen og lærerutdanningene, der studentene skal ut

å undervise i faget etterpå. Her mener informanten altså at fagets rolle og funksjon i utdanningene ikke ble tatt tilstrekkelig hensyn til.

Generelt er aktører fra UH-sektoren positive til å se forskning og utdanning mer i sammenheng. Som en representant for en privat høyskole påpeker, så er det ikke bare slik at forskningen påvirker utdanningen – utdanningen og studentenes arbeid kan også påvirke forskningen, særlig på høyere nivå. Slike koblinger er ønskelige, og bør derfor også evalueres som en helhet. En representant for et stort universitet understreker at det å koble forskning og utdanning er noe de generelt snakker mer og mer om innad på institusjonen, og at det hadde vært nyttig med undersøkelser av hvordan man kan gjøre en slik kobling. Aktøren understreker imidlertid at hvordan man ser utdanning og forskning i sammenheng kan variere mellom institusjonene, og det vil være forskjeller mellom et forskningstungt universitet og en mindre høyskole. Dermed blir det vanskelig å utvikle en evalueringsmodell som kan passe alle.

Mange informanter understreker behovet for å finne frem til gode modeller og metoder for å evaluere forskning og utdanning sammen. Det er også enkelte som kommer med eksempler på evalueringer der dette ikke har blitt gjort på en god måte. Én informant trekker frem en NOKUT-evaluering av et masterprogram, der de ønsket å undersøke i hvor stor grad undervisningen var forskningsbasert. Informanten forteller at alle underviserne på studieprogrammet fikk tilsendt en excel-fil der de ble bedt om å koble konkrete forskningsprosjekter til konkrete emner, og at dette ble oppfattet som relativt meningsløst. Vedkommende mener at man som underviser bringer med seg hele sin kompetansekapital, og at man kan trekke på all sin erfaring som aktiv forsker. Det å «vri ut relevans» fra konkrete forskningsprosjekter opplevde personen som en «kunstig prosess». I tillegg kan det være mange eminente forskningsmiljøer, enten på universiteter eller i forskningsinstitutter, som ikke er koblet til utdanning. Man må også vurdere hvordan disse kan passe inn i en felles evalueringsmodell. Representanter fra instituttsektoren har også uttrykt bekymring rundt deres plass i et evalueringssystem som i større grad ser forskning og utdanning i sammenheng. Det er en fare, mener flere, for at et nasjonalt rammeverk som i større grad knytter forskning til utdanning fører til at evalueringene blir gjennomført på *UH-sektorens* premisser.

Flere aktører nevner KOMBEVAL fra 2018. Det ser ut til å være en bred enighet om at det var et interessant forsøk, men at man ikke fikk så mye ut av det. En representant fra KD mener det kan skyldes at det var et pilotprosjekt og at NFR og NOKUT ikke har mye erfaring med å samarbeide. Også representanter fra UH-sektoren uttaler at de opplever NFR og NOKUT som litt ulikt «skrudd sammen» og at det kan være nødvendig å jobbe videre med å utvikle en god metode for slike evalueringer. De fleste er likevel positive til å fortsette med denne utviklingen.

Som tidligere nevnt poengterer mange interessenter i denne sammenheng at slike kombinerte evalueringer ikke må komme i *tilllegg* til eksisterende evalueringer. De må enten erstatte, eller integreres i, øvrige evalueringer, om ikke belastningen for evaluandene skal oppleves som for stor. Samtidig peker andre på potensielle *stordriftsfordeler* ved å kombinere forsknings- og utdanningsevaluering. Dette kan gjelde både for NOKUT og NFR, men også for sektoren mer generelt, dersom kombinerte evalueringer gjennomføres på en helhetlig og koordinert måte som reduserer behovet for å rapportere samme eller lignende data flere ganger, til flere instanser.

3.6.2 Mulige måter å gjøre det på/løsninger

Et grunnleggende problem med kombinerte evalueringer er hva som skal være *evalueringsobjektet*. Forskning og utdanning organiseres på forskjellige måter, og man kan for eksempel ikke basere seg på *forskergrupper* dersom man ønsker å uttale seg

om utdanningskvalitet på programnivå. Forskere bidrar i forskjellig grad, og på forskjellige måter, inn i undervisningen på programmer, og de aller fleste interessentene sliter med å finne gode løsninger på denne nivåproblematikken.

En mulig løsning på utfordringen med å se forskning og utdanning i sammenheng er å evaluere fagmiljøene. Som en representant for KD uttalte «det er to verdener, men med de samme ansatte». Dette gjelder ikke i alle tilfeller, men ofte vil det være de samme ansatte som står for forskningsvirksomheten og som har ansvaret for utdanning innen et fagmiljø. Dermed kan man undersøke flere aspekter ved fagmiljøet, som for eksempel hvor godt samarbeidet innad er, hvor god ledelsen er, hvor godt miljøet jobber med å styrke seg selv og hvordan det samarbeider med andre fagmiljøer på tvers av institusjonelle og faglige skillelinjer. Det er imidlertid uklart i hvilken grad en evaluering basert på *fagmiljø* vil gi resultater som er nyttige på programnivå.

Blant UH-institusjonene mener mange at en mulig løsning er å inkludere utdanningsvirksomheten som en del av forskningsevalueringene, ettersom de to går hånd i hånd og ikke kan sees som to helt separate aktiviteter. Her kan man evaluere fagmiljøer, og inkludere hvordan miljøet forholder seg til utdanning, men ikke nødvendigvis evaluere utdanningsvirksomheten ned på enkeltemnenivå. Da kan det bli vanskeligere å finne frem til en modell og metodikk som kan favne både forskningens og utdanningens karakteristikk. Med en slik løsning er det imidlertid også viktig å beholde, eller utvikle et rammeverk som gjør det mulig å vurdere forskergrupper- og miljøer *uten* særskilte koblinger til undervisningsaktivitet.

Videre er det flere som etterspør mer koordinering på feltet, og at slike evalueringer bør koordineres av KD, og ikke av NFR eller NOKUT, slik at man kan inkludere et utdanningsperspektiv, noe som ikke inngår i NFRs mandat. Her påpekes det at KD bør ha dialog med aktørene, særlig institusjonene, på forhånd, slik at det er en felles enighet rundt premisset for evaluering. Slik kan man fange helheten i stedet for å dele ansvaret evalueringene mellom flere direktorater med ulike mandater. Dette kan sees i sammenheng med det generelle behovet for koordinering på feltet som ble diskutert i avsnitt 3.1, og som diskuteres videre i seksjon 5.5.

Oppsummering

Sentrale motsetninger og aktører med spesielle interesser

Små vs. store institusjoner	UH-institusjoner
UH-sektor vs. andre	UH-sektoren, forskningsinstitutter, helseforetak
Grunnforskning vs. anvendt forskning	Forskningsinstitutter, departementer, arbeidslivsorganisasjoner
Disiplinbasert forskning vs. annen organisering	Universiteter og forskningsinstitutter

Hovedpoenger

- Frekvens og omfang vurdere opp mot nytte, og nytten oppleves i hovedsak som god
- De fleste interessentene vurdere eksisterende evalueringssystem som nyttig og hensiktsmessig
- Det kan være behov for bedre koordinering av forskjellige typer evalueringer og annen rapportering
- Det er ønskelig, men vanskelig, å kombinere evaluering av forskning og utdanning

4 Bruk og effekter av evalueringer

I dette kapittelet presenterer vi interessentenes oppfatninger om og erfaringer med hvordan evalueringer *brukes* og hvilke *effekter* de har. Mange av aktørene bruker evalueringer aktivt, og svarer for eget bruk og erfarte effekter. Andre igjen omtaler hvordan *andre* bruker evalueringer. Først fokuserer vi på bruken av evalueringer og deretter går vi over til effektene av evalueringene, inkludert i hvilken grad evalueringer bidrar til å styrke tilliten og legitimiteten til forskning og høyere utdanning.

4.1 Bruk av evalueringer

De fleste aktørene sier at de bruker evalueringer som et kunnskapsgrunnlag, men det varierer relativt mye hva de bruker dette kunnskapsgrunnlaget til. Dette kommer delvis an på hva slags aktør det er snakk om. De som jobber med forskningspolitikk bruker evalueringer i dette arbeidet, mens de som jobber på UH-institusjoner i større grad bruker evalueringene internt på institusjonene. Andre igjen bruker evalueringer mer strategisk, for eksempel i konkurranse om begrensede midler, og i forsøk på å påvirke forsknings-, innovasjons- og utdanningspolitikk. I tillegg trekker mange frem evalueringer som en nyttig form for benchmarking og sammenligning mellom institusjoner.

Forskningspolitikk

På det forskningspolitiske nivået blir evalueringer brukt som et kunnskapsgrunnlag for beslutninger og strategiske grep. For KD gjelder dette både i styringsdialogen med NFR og institusjonene og i mer langsiktig planlegging av politikk for forskning og høyere utdanning. De store, nasjonale evalueringene og undersøkelsene kan ofte peke på utfordringer som gjelder sektoren generelt, og som dermed kan gi utslag i politiske endringer og strategiske grep. En representant fra KD trekker frem NFRs evalueringer av bio-, helse- og medisin og av teknologifag som to evalueringer som har fått konsekvenser for den langsiktige forskningspolitikken. I tillegg brukes evalueringer til å legitimere, begrunne og påvirke reformer. Her trekkes både kvalitetsreformen og strukturreformen frem som eksempler. En del av bakgrunnen for strukturreformen var at mange av NFRs evalueringer viste at Norge var preget av små og fragmenterte fagmiljøer. Man ser også at NFRs fagevalueringer blir sitert i bl.a. stortingsmeldinger, og at de dermed blir brukt politisk. Andre trekker også frem evalueringen av medisin for ca. 10 år siden som en evaluering med klare og tydelige konsekvenser for feltet som helhet, og for de evaluerte institusjonene. Evalueringen førte også til at svakheter på feltet ble skikkelig belyst, og dette har ifølge respondentene ført til en massiv opptrapping, med økt ressurstilgang, noe som har ført til en betydelig kvalitetsforbedring på feltet.

Flere departementer peker også på hvordan evalueringer med klare anbefalinger, fra internasjonale paneler med stor tyngde, har fått tydelige konsekvenser for praksis både i sektoren generelt og i finansiering av forskning og prioritering av satsningsområder. Selv om de har hørt det samme direkte fra institusjonene og forskningsmiljøene, veier det tungt å få dette bekreftet med eksterne evalueringer – det gir legitimitet. Samtidig påpeker flere interessenter som ikke selv er forskningsprodusenter at det er viktig med legitimitet, siden de opplever en del oppdragsforskning som nærmest rene bestillingsverk, og at det dermed er viktig å få belyst hvem som er gode og pålitelige.

Én aktør trekker også frem at nasjonale kandidatundersøkelser viser en synkende interesse for å jobbe i akademia og gå inn i en forskerkarriere. Dette har nå ført til at KD lager en strategi for rekruttering, slik at sektoren i større grad skal kunne beholde gode kandidater.

En representant for en fagforening uttaler at det har vært en rask og sterk utvikling mot at det snakkes over hodet på de fagansatte i sektoren, og at evalueringer inngår i denne utviklingen. Det blir pekt på at disse undersøkelsene ofte blir gjort «ovenfra og ned», og at de deretter brukes til å gjøre politiske, strategiske og ressursmessige beslutninger som ikke involverer de fagansatte. Ifølge aktøren viser dette en mangel på forståelse for verdien av forskerne og underviserne i sektoren.

Kvalitetsutvikling/endring (formativ bruk)

Ifølge representanter fra KD er de formative delene av evalueringene mest rettet mot institusjonene og fagmiljøene selv. KD er ikke involvert i oppfølgingen av disse delene av evalueringene ute i institusjonene utover at evalueringer til dels kan tas opp i styringsdialogen med institusjonene.

Innad på UH-institusjonene og instituttene brukes imidlertid evalueringene i stor grad som kunnskapsgrunnlag i kvalitetsutvikling og som en katalysator for endring. Svært mange av aktørene beskriver hvordan evalueringer peker på forbedringspunkter eller negative forhold man allerede var klar over, men at det blir lettere å ta tak i problemene når man får en slik ekstern bekreftelse. Det kan enten være fordi det gjør behovet for endring enda tydeligere, eller fordi man kan «slå i bordet» med evalueringen internt og få satt i gang prosesser man tidligere ikke har fått satt i gang. En representant for et institutt sier at det blir lettere for ledelsen å gjennomføre tiltak hvis man kan vise til noe mer objektivt, slik at det ikke tolkes som ledelsens subjektive mening eller oppfattelse. Evalueringer kan på den måten være med på å minske motstanden mot endring i fagmiljøer som omtales som noe konservative.

En undersøkelse som svært mange av aktørene i UH-sektoren trekker frem i denne sammenheng er Studiebarometeret. Denne undersøkelsen er noe de aller fleste intervjuobjektene fra UH-sektoren har et forhold til, og de fleste anser det som en viktig tilbakemelding som blir aktivt brukt, men at det ikke kan brukes alene som et mål på utdanningskvalitet. Det varierer også mellom institusjonene hvordan det brukes. En representant for et universitet understreker at Studiebarometeret ikke er en premissgiver, men at man ofte følger med på barometeret etter at man har gjort endringer i et studieprogram for å se om det gir utslag i studenttilfredsheten. En representant for en høyskole fortalte imidlertid at de har foretatt undersøkelser dersom noen utdanningsprogrammer har fått svært negative resultater i Studiebarometeret. Da har man for eksempel funnet at det har vært konflikter i fagmiljøet eller lignende som påvirker studentenes tilfredshet. En representant fra et annet universitet forteller at Studiebarometeret brukes til å utfordre selvforståelsen av ulike studieprogrammer. Dersom et program får dårlige resultater blir ikke dette tolket som en «fasit», men som en indikator på at man bør ta en nærmere kikk på programmet og se om resultatene gjenspeiler virkeligheten.

En ting som fremstår som viktig for at evalueringer skal kunne brukes til kvalitetsutvikling er at de inneholder noe mer enn rene talldata eller en rangering. Dette understrekes særlig av mindre høyskoler der man ofte både har mindre ressurser og erfaring med kvalitetsarbeid. En representant for en høyskole nevner en evaluering der høyskolen kom dårlig ut, men understreker at det allikevel ble oppfattet som positivt fordi evalueringen ga så mye «kjøtt på beina» at det var lettere for aktuelle avdelinger å gjøre endringer i etterkant.

Til sist er det mange av aktørene som beskriver hvordan institusjonene bruker evalueringer og tilsyn som anledninger til å sette ekstra søkelys på kvalitetsarbeidet – uavhengig av resultatene av evalueringen. Ofte vil prosessen med å forberede seg til en evaluering, samt innhenting av dokumentasjon, være nyttig i seg selv og gi

institusjonen bedre selvinnsikt i hvordan de arbeider, hva som fungerer og hva som eventuelt kan forbedres. På denne måten kan evalueringssprosessen – ikke kun resultatene - bidra til læring og utvikling. Her understrekes også betydningen av egnevalueringer som et element i evalueringene. Evalueringer oppleves som lærerike og bevisstgjørende uansett, men dette gjelder spesielt når institusjonene selv er nødt til å vurdere sentrale elementer ved egen virksomhet.

Dette kan også være en delforklaring på hvorfor så mange av interessentene forteller at evalueringer er nyttige, men at rapportene sjelden gir *ny* kunnskap for institusjonene. Ved å evaluere seg selv, og mer generelt være en del av et evalueringssystem, virker det som om mange av institusjonene har fått relativt god oversikt over egen virksomhet – spesielt når det gjelder lett målbare aspekter som ytelsen og robustheten til forskergrupper, publiseringer, o.l.

Det er derfor viktig å påpeke at mange av interessentene sier at evalueringer er nyttige, og brukes, men at det også er visse begrensninger knyttet til bruksmuligheter. Om vi fokuserer på fagevalueringer skyldes dette primært at fokuset på *et bredt felt* nødvendigvis fører til at evalueringene i begrenset grad kan gå svært detaljert inn i hver institusjon. På overordnet nivå, hevder en annen interessent, vil det være vanskelig for toppledelsen ved institusjoner å ta for eksempel strategiske grep basert på kunnskap om hvilke sterke og svake forskningsmiljøer man har. De fleste fremmer imidlertid ikke dette som en kritikk, eller som et ønske om en annen innretning. Det fremstilles heller som *kostnaden* ved å oppnå de fordelene man får ved fagevalueringer, og en erkjennelse av at valg av innretning medfører at avveininger må gjøres og valg må tas. Disse avveiningene har imidlertid fått fornyet fokus når NFR nå vurderer mer institusjonsrettede evalueringer, noe vi kommer tilbake til i kapittel 5.

Benchmarking og måling av kvalitet (summativ bruk)

Selv om mer nyanserte analyser fremstilles som nyttig for kvalitetsutvikling, er det også mange aktører som trekker frem målinger, sammenligning og «benchmarking» som en viktig funksjon ved evalueringene. En representant for KD uttaler at evaluering av forskning er viktig nettopp for å sjekke kvaliteten etter internasjonale standarder og å få en benchmarking nasjonalt slik at de som ikke når helt opp kan strekke seg. For KD er de summative delene av evalueringene mer sentrale enn de formative, siden sistnevnte retter seg mer mot institusjonene. Et viktig aspekt for KD er å vite hvordan det står til med norsk forskning, hvor mange fremragende miljøer som finnes innenfor ulike fag, og hvordan ressursene som investeres i utdanning og forskning blir brukt. Det siste poenget blir også trukket frem av andre som finansierer forskning, og de fleste opplever at dagens evalueringer treffer godt med tanke på denne typen ansvarliggjøring.

Benchmarking blir trukket frem av mange av UH-institusjonene og øvrige forskningsprodusenter. Det varierer imidlertid hvem de ser behov for å sammenligne seg med. Blant de større universitetene er internasjonal benchmarking viktig, mens den nasjonale ikke vektlegges like sterkt. Imidlertid trekker en representant for et universitet frem at nasjonal benchmarking kan være særlig relevant innen disiplin-fag som finnes i relativ lik form på flere institusjoner i Norge. Flere mindre høyskoler legger stor vekt på både internasjonal og nasjonal benchmarking, og dette kobles ofte til ambisjoner om å vokse eller å søke om universitetsstatus. For mindre høyskoler kan nasjonal benchmarking være nyttig for å vise hvilke områder de er gode på. En representant for en privat høyskole legger spesielt stor vekt på benchmarking, ettersom dette blir brukt aktivt i markedsføringen av høyskolen. Ettersom inntektsgrunnlaget hovedsakelig kommer fra studieinntekter er benchmarking og sammenligning på utdannelsesområdet og studenttilfredshet svært viktig.

En representant fra KD setter spørsmålstegn ved evalueringers rolle i benchmarking ettersom det allerede samles inn så mye data og statistikk at benchmarking og sammenligning på tvers av institusjoner uansett finner sted. Man må dermed spørre seg om hva evalueringer kan bidra med, ut over ren formidling av talldata og statistikk.

Dette fremstår som et sentralt tema, og et av hovedfunnene i interessentanalysen: mange aktører har en klar oppfatning om at det finnes et potensiale for å utnytte løpende data og indikatorer bedre, og heller fokusere *evalueringene* noe mer på forhold som *ikke* belyses av dette datamaterialet. Dette temaet behandles derfor i mer detalj i seksjon 5.4. Mange, og tunge, aktører som representerer både UH-institusjoner, institutter og helseforetak deler en interesse for denne tematikken.

Strategisk bruk (internt og i ekstern kommunikasjon)

Flere av aktørene i UH-sektoren forteller at evalueringene blir brukt strategisk internt til å ta beslutninger, gjøre prioriteringer og til å profilere seg på ulike fagområder. Et stort universitet forteller om hvordan fagevalueringene har vært viktige innspill i strategiutvikling, spissing av innsats, samt at de har ført til at visse områder har blitt nedprioritert, satt på vent, osv.

En aktør fra instituttsektoren viser til et eksempel der en evaluering pekte på at de burde satse mer på å være koordinator for forskningsprosjekter, i stedet for å bare inngå som partner. Dette var noe som ble fulgt opp i etterkant og som fikk positive konsekvenser for instituttet. Representanter fra to departementer har også nevnt at de har brukt evalueringer som et kunnskapsgrunnlag i «strategisk instituttsatsing», der de har valgt å samarbeide med fagmiljøer som får gode resultater i evalueringene.

I tillegg kan resultatene av ulike undersøkelser og evalueringer brukes strategisk utad for å enten skape blest om temaer eller som en del av en pågående debatt. En representant for en studentorganisasjon forteller blant annet at de bruker Studiebarometeret aktivt for å rette fokus på studentenes trivsel og legge press på institusjonene dersom det kommer frem negative resultater. En representant for en privat høyskole understreker at all omtale av dem, inkludert evalueringer, er viktig for deres omdømme og dermed også inntektsgrunnlag.

Evalueringer brukes også aktivt av departementer, direktorater og arbeidslivsaktører. Mange av disse aktørene (KD ikke inkludert) har forholdsvis lite kjennskap til evalueringene som NFR har gjennomført *generelt*, men det er tydelig at evalueringene som mest tydelig er relatert til egen sektor leses med stor interesse. Disse får i visse tilfeller konsekvenser for aktørenes virksomhet. Det fremstår imidlertid som om evalueringene blir brukt mest strategisk av disse aktørene, ved at de henviser til evalueringene der de gir støtte for argumenter i budsjettforhandlinger, debatter, o.l.

Utfordringer med å bruke evalueringer

Selv om mange av aktørene forteller at de bruker evalueringer aktivt, er det også mange som mener evalueringer ikke alltid er så relevante eller enkle å bruke. For eksempel sier en representant for et universitet at de som en stor, desentralisert organisasjon ikke får så nyttige resultater på institusjonsnivå. Dermed er de avhengige av at resultatene følges opp lokalt av de relevante fakultetene og instituttene.

Det er generelt også flere som påpeker at det oppleves som lite motiverende, og vanskelig, å skulle få en generell karakter på det arbeidet en gjør, uten spesielt gode begrunnelser, og noen ganger på det som oppleves som svært sviktende grunnlag. Dette gjelder spesielt de ansatte, og til dels også enkeltmiljøer som blir evaluert. På et mer overordnet nivå, påpeker interessenten, fungerer fagevalueringer relativt godt. Andre

forteller også at de i stor grad ikke forstod hvordan panelene hadde kommet frem til de konklusjonene de la frem, og dette oppleves opplagt som en utfordring.

En sentral utfordring som pekes på av en representant for instituttsektoren er at evalueringer *brukes feil*. Evalueringene kan være meget gode, om de brukes korrekt og forstås. Men, sier interessenten, mange trekker funnene for langt, glemmer premissene for evalueringene, og bruker dermed evalueringen på måter de ikke har grunnlag for. Dette fører også til urimelig kritikk av evalueringene. Et eksempel som trekkes frem er en teknologievaluering som konkluderte med at man i Norge manglet «Blue sky»-forskning. I realiteten, hevder en interessent, skyldes dette at evalueringens premisser var lagt opp slik at mange miljøer (inkludert UiO-miljøet som har mest forskning på dette feltet) ikke ønsket å delta i evalueringen.

Et relatert poeng knyttet til premisser er at evalueringer basert på disipliner blir både konservative og forutsigbare. Om et panel med disiplinekspertter skal evaluere, hevder én, vil de komme frem at det finnes gode fagmiljøer (og at de bør få mer midler til fri forskning innenfor feltet). Disse evalueringene oppfattes som tilpasset UH-sektoren og spesielt de som er organisert etter disiplin. Resultatet er at instituttene passer dårligere inn, og kanskje enda viktigere: svært mye forskning og utvikling i næringslivet oversees fullstendig. En interesseorganisasjon peker på at dette a) fører til at viktig forskning overses og b) at mange forskere har insentiver til å forske «trygt», fremfor å søke nye og radikale løsninger og innovasjoner som kan løse større samfunnsproblemer. Disiplinorientering, og insentivsystemer knyttet til publikasjoner, o.l. kan med andre ord legge en demper på innovasjon og nyteknik. Dette kan skyldes at de mest attraktive tidsskriftene er a) disiplinspesifikke og b) kan være noe konservative.

En representant for en privat høyskole mener at institusjonen selv er ekstremt opptatt av kontinuerlig kvalitetsforbedring, og at enkeltevalueringer kommer i tillegg til, og litt på siden av, det. Denne høyskolen har et omfattende egeninitiert evalueringsregime som består av rene interne prosesser, i tillegg til at de deltar i internasjonale nettverk og organisasjoner som innebærer at de også vurderes/evalueres av internasjonale paneler. Denne type evalueringer, som er skreddersydd for organisasjonen, og *type* organisasjon, fremstår som et mer effektivt læringsverktøy enn de mer overordnede evalueringene som gjennomføres av NFR. Flere forteller at de nasjonale evalueringene leses når de kommer, og at man deretter har et oppfølgingsmøte der man diskuterer resultatet. Etter dette brukes de i mindre grad systematisk av institusjonen. Én forteller at de stort sett får vite ting de vet fra før. Deretter deles anbefalingene internt – kanskje resultatene deles på Twitter, osv – men det ender her, så lenge evalueringene ikke er knyttet til *konsekvenser*, og så lenge de ikke oppleves som tilpasset nok til at de tilfører ny og ukjent kunnskap. De «legges i skuffen», forteller flere, men dette gjelder dette på langt nær alle (jf. avsnitt 4.2).

En representant for en statlig høyskole mener også at de ulike institusjonene har svært forskjellige forutsetninger for å bruke resultatene. De store universitetene har ofte både høyere kompetanse på og mer ressurser til den typen oppfølgings- og kvalitetsarbeid som kreves. Representanten understreker at det kreves kompetanse både for å gjennomføre en evaluering og for å analysere resultatene. Særlig hvis resultatene kommer som statistikk og tallmateriale kreves det metodisk kompetanse og ressurser for å få gode svar ut av råmaterialet. Også representanter for store universiteter etterspør imidlertid mer interne ressurser knyttet til analyse og innsikt. Dette er altså ikke begrenset til mindre eller nyere institusjoner, og det fremstår som et viktig tema for mange UH-institusjoner. Samtidig er det tydelig at noen av institusjonene har satset mer på å bygge opp interne analysemiljøer og bruk av for eksempel *learning analytics* (som vi også diskuterer videre i kapittel 6). Det er imidlertid en tendens hos andre til å oppfatte de tyngste UH-institusjonene som noe mer selvstendige, og noe mindre overbevist over

verdien av NOKUT- og NFR-evalueringer. Dette inntrykket kan til dels avkrefte av vår analyse, til tross for at vi finner et visst belegg for at disse institusjonene har et større og mer solid internt miljø for kvalitetssikring og analyse, både på forsknings- og utdanningsiden.

I tillegg kan de mindre institusjonene ha utfordringer med å få nok respondenter til at undersøkelsene gir gyldige resultater. Her trekker aktøren spesielt frem NIFUs kandidatundersøkelse som noe som i teorien er flott, men som i praksis er lite nyttig for institusjonen. En siste ting som oppleves som problematisk er at institusjonene ofte blir bedt om å redegjøre for hvordan de har brukt evalueringene og resultatene. Dermed blir man målt på hvordan man bruker resultater som oppleves som lite nyttige. Dette oppleves som lite hensiktsmessig.

Til sist fremheves det også som problematisk når indikatorer og parametere som er utviklet for institusjonsnivået blir brukt til å evaluere forskere individuelt. Dette gjelder for eksempel publiserings- og siteringsindikatorer. Ifølge KD selv skal dette brukes til å fordele penger mellom institusjoner. Imidlertid blir de noen ganger brukt til å fordele penger mellom enkeltforskere. Dette understrekes også av en fagforening som mener det er viktig å skille mellom samfunnsoppdraget på institusjonsnivå og hva som forventes av den enkelte forsker. En UH-institusjon skal som helhet levere på både forskning, undervisning og innovasjon, men dette kan ikke forventes av hver enkelt forsker.

Et annet viktig poeng er at en del organisasjoner over tid opplever store endringer internt i organisasjonen. Dette gjelder blant annet et forvaltningsorgan som har vært gjenstand for omorganiseringer der forskjellige funksjoner flyttes både inn og ut av virksomheten. Denne type endringer gjør det nærmest umulig å bruke evalueringer til å si noe om endring over tid, og det gjør det også vanskelig å bruke forslag til endringer som fremmes i evalueringene, når disse over tid ikke lenger er gyldige. Dette er imidlertid ikke en utfordring som knyttes direkte til evalueringene, men heller til et generelt behov for stabilitet dersom man ønsker å isolere effekter av tiltak over tid. Et viktig poeng her kan være at en evalueringsmodell som baserer seg mer på institusjoners særegenheter, og interne kvalitetsarbeid, kan gjøre institusjonene bedre i stand til å håndtere denne type endringer. Prisen man betaler, ved at man i mindre grad får sammenlignet seg med andre eller seg selv i fortiden, kan til dels rettferdiggjøres ved at evalueringene bidrar til å styrke institusjonene og deres evne til å håndtere og styre endringsprosesser.

Andre aktører, som departementer (ikke KD), direktorater og øvrige organisasjoner som ikke har egen forsknings- eller undervisningsaktivitet opplever i hovedsak at nasjonale evalueringer, som for eksempel fagevalueringer, har *begrenset* nytteverdi for dem. Aktørene er som nevnt ovenfor i prinsippet interesserte i kunnskap om forskning og utdanning, men evalueringene oppleves i hovedsak som mindre relevante for denne målgruppen. Mange av disse aktørene vektlegger imidlertid elementer som i økende grad får oppmerksomhet i evalueringer i form av *impact*. Departementer og interesseorganisasjoner, for eksempel, deler et sterkt ønske om at *relevans, anvendelse* og *samfunnsnytte* skal vektlegges og tydeliggjøres.

4.2 Effekter av evalueringer

I vurderingen av *effekter* av evalueringer peker mange av interessentene på historiske evalueringer og observerte effekter av disse. I diskusjonen som følger vil vi dermed til dels knytte effekter til evalueringene som nevnes. Dette er imidlertid kun et tilfeldig utvalg, og vi tar ikke her stilling til om de evalueringene som nevnes har hatt større, eller andre, effekter enn øvrige evalueringer. Det er allikevel tydelig at flere av evalueringene blir nevnt nettopp fordi interessentene opplever at de har hatt klare effekter. Samtidig er

det nærliggende å anta at evalueringer som SAMEVAL og HUMEVAL nevnes oftere enn eldre evalueringer, siden mange av våre informanter forteller at de er relativt nye i sine stillinger, og at det dermed primært er de nyere evalueringene de kjenner til. Det er også naturlig å anta at nyere evalueringer er lenger fremme i interessentenes bevissthet enn eldre evalueringer.

Organisatorisk

Et viktig aspekt ved evalueringer er at de aldri vil være *nøytrale*, hevder flere interessenter. De er et resultat av et spesielt perspektiv på kvalitet og hva som er hensiktsmessig organisering av arbeid, og dette vil uunngåelig påvirke institusjoner og individer som evalueres. Dette gjelder både forsknings- og utdanningsevalueringer. En interessent sier at en verden i rask endring, med stadig nye kompetansebehov, egentlig innebærer at god utdanningskvalitet nødvendiggjør store og radikale endringer. Men det er en risiko for at institusjonene ikke tør å endre seg, fordi dette kan innebære at de går på tvers av kravene for NOKUT-godkjenning. På samme måte vil for eksempel vurdering av publiseringskanaler og vektlegging av publiseringspoeng påvirke enkeltforskernes prioriteringer sterkt. Dette er overordnede og relativt abstrakte effekter, og de fleste interessentene diskuterer ikke dette temaet i særlig grad.

Innretningen av evalueringer har effekter på organiseringen av forskning. Dette er en påstand som i forskjellige former fremmes av en rekke aktører i UH- og instituttsektoren. Dette knyttes spesielt til SAMEVAL, og fokuset på større forskergrupper. Ifølge representanter for store universiteter, er det store forskjeller mellom fagmiljøer, og spesielt representanter fra sosiologi-miljøet ble sentrale i debatten om hvilke konsekvenser innretningen av evalueringer har.

Mange peker på at den mest opplagte effekten av fokus på større forskergrupper var at institusjonene i større grad begynte å organisere og oppfordre til opprettelsen av og deltakelsen i slike forskergrupper. Dette førte som nevnt til store debatter i flere miljøer, der mange peker på at en slik organisering av forskning er naturlig i «Mat-Nat» (matematisk-naturvitenskapelige fag), men *ikke* i, for eksempel, flere av samfunnsfagene.

I løpet av interessentanalysen har vi imidlertid gått lenger inn i denne debatten, og SAMEVAL og innretningen av denne evalueringen fremstår nå som noe mindre kontroversiell, også for de miljøene som var aktive i debatten. Flere peker på at noen av utfordringene var knyttet til at evalueringsformen var *ny* for disse fagmiljøene. Dette innebar mange nye arbeidsoppgaver, som man ikke var vant til, noe som skapte en del usikkerhet. Poenget her er at det å bli evaluert til dels er en *treningssak* – noe man både blir vant til, og blir bedre til. Dette er også noe av grunnen til at mange ønsker forutsigbarhet og faste evalueringer som gjentas regelmessig. Første gang man evalueres, sier én, er det litt som i en «skyttergrav», man er usikker, forsvaret seg, og evalueringsformen må «sette seg litt» for at den skal bli nyttig. Interessenten forteller at evalueringene oppleves som meget nyttige når de *har* satt seg. Faren er imidlertid om det setter seg *for godt*, og rett og slett blir rutine. Interessenten antar at det kanskje er dette som er årsaken til at NFR nå ønsker å se på mulighetene for å endre innretning.

Samtidig peker flere på at det er et reelt poeng at organiseringen av forskning blir påvirket. Men, mange av aktørene mener at også samfunnsfag som sosiologi uten problemer kan organiseres i forskergrupper, og at det er en klar utvikling fra enkeltstående forskere til mer aktive forskernetverk innenfor alle fagområder.

En annen posisjon knyttet til det foregående er at SAMEVAL var et resultat av, og til dels satt fokus på, betydningen av robuste fagmiljøer. Spesielt de store aktørene mener at det finnes mange små forskermiljøer av variabel og tidvis dårlig kvalitet. For å oppnå

forskning av god kvalitet – oppnå *eksellens* – er det nødvendig med *konsentrasjon*. Flere hevder at NFRs evalueringer i for *liten* grad fokuserer på dette spørsmålet. Oppfatningene om dette spørsmålet er tydelig delte.

I tillegg er det viktig å påpeke at mange av de tyngste forskningsaktørene sier at de har vært meget fornøyde med fagevalueringene som har vært gjennomført, og at disse har gitt grunnlag for å gjennomføre organisatoriske endringer og å ta strategiske grep. Dette gjelder spesielt Mat-Nat-miljøer, men også helseforetak. Evalueringen av biologi, medisin og helseforskning (2011) og fysikk (2010) trekkes frem som evalueringer som hadde klare effekter – både for intern organisering i institusjoner og for forskningspolitikk på feltene.

Det er i denne sammenheng relevant at evalueringene kun unntaksvis omtales som *overraskende* eller at de i stor grad ga *ny kunnskap*. En av hovedgevinstene som omtales av interessentene som har blitt evaluert, og ledere i disse institusjonene, er at institusjonene og fagmiljøene har fått en ekstern bekreftelse av et situasjonsbilde de i stor grad kjenner. Dette brukes aktivt for å organisere arbeidet internt i organisasjonen, ved at visse miljøer blir satset mer aktivt på, mens andre miljøer kanskje legges ned eller omorganiseres. Evalueringene har også blitt brukt aktivt for å endre forskningspolitikk, fordeling av forskningsmidler, og prioritering av satsningsområder.

En representant for UH-sektoren er også tydelig på at evalueringer *bør* ha klare effekter, og kanskje også mer direkte *konsekvenser*, ved at det for eksempel knyttes til finansiering og ressurser. Men, forteller interessenten, det er to mulige måter å tilnærme seg dette på: a) premiere gode prestasjoner, eller b) identifisere behov for hjelp. Begge tilnærminger kan være legitime, og dette ble primært tatt opp som et poeng interessenten mente man var nødt til å diskutere i større grad. Flere representanter vektlegger dette, og peker på at mulighetene til å søke *oppfølgingsmidler* i etterkant av evalueringene er viktig for at institusjonene og miljøene skal kunne følge anbefalinger og klare å gjennomføre forbedringer. Én representant tar det lenger, og sier at dersom formålet utelukkende er å si hvem som er flinke og ikke, har evalueringene liten interesseverdi – de må medføre *konsekvenser*.

Det er annet poeng, relatert til dette, som dukket opp ved flere anledninger. Er det slik, sier flere, at de som klager på innretningen og metodene som brukes gjør det fordi de har kommet dårlig ut i evalueringen? Og motsatt, er det slik at de som gjør det bra liker evalueringen? Vi har ikke undersøkt denne sammenhengen på noen måte som lar oss konkludere på dette punktet, men det fremstår i det minste som et relativt tydelig mønster at store og tunge Mat-Nat-miljøer, og de som utfører forskning på en måte som *passer* evalueringene, er svært fornøyde med evalueringene. Øvrige interessenter har flere reservasjoner, men vi har i det hele tatt et overordnet inntrykk av at alle interessentene har vært i stand til å identifisere både styrker og svakheter ved den type fagevalueringer som har fått mest fokus i diskusjonen av effekter, styrker og svakheter. Vi har også eksempler på interessenter som har fått gode evalueringer, og som har brukt dette som en mulighet til å kritisere den underliggende logikken og de implisitte antakelsene om hva god forskning er og hvordan den kan dyrkes frem.

Forskningspolitikk, dimensjonering og spesialisering

Som en fortsettelse av det forrige punktet vil vi understreke evalueringers effekter på *dimensjonering* og *spesialisering* av forskning. Dette fremheves særlig av de største forskningsaktørene og av KD.

Nasjonale fagevalueringer har gitt en tilstandsrapport på forskjellige felt. Dette har i visse tilfeller utløst satsninger og spesiell innsats rettet mot områder der man ser spesielle

behov. Det er eksempler på satsinger både for å rette opp en skute med sterk slagside og for å gå fra å være god til å bli bedre.

I tillegg påpeker en del av de store institusjonene at evalueringene kan gi grunnlag for dialog mellom institusjonene. Spesielt de aktørene som er opptatt av *konsentrasjon* og at fagmiljøer må ha en viss størrelse og tyngde for å bli gode, peker på dette. Evalueringer, sier de, kan fortelle oss at en institusjon er god på noe, mens en annen har et annet fortrinn. Evalueringer kan, og *bør*, hevder flere, brukes til å identifisere og utnytte slike komparative fortrinn. Institusjonene kan selv gjøre dette, men de peker også på at KD, og for eksempel NOKUT, har en rolle i å bruke evalueringer til å sørge for en hensiktsmessig utvikling på feltene. Her nevnes NOKUT, spesielt fordi de oppfattes som en aktør som kan bidra til å bremse oppblomstringen av små, «populistiske» og lite robuste studier og fagmiljøer. Det er spesielt de større interessentene som vektlegger dette, og de fremmer disse synspunktene fordi de oppfatter det som meget uheldig at studentene får lite relevante, og lite gode, studietilbud flere steder.

Tillit og legitimitet

Evalueringer kan påvirke tillit til forskning og høyere utdanning. Mange av interessentene mener dette er et meget viktig poeng, og at det også kommer til å bli stadig viktigere. Dette knyttes i stor grad til forskningens og vitenskapens verdi i en verden som preges av diskusjoner om «fake news», «flere sannheter» og en tendens mot økende usikkerhet om hva som er sant og sikkert. Forskning og utdanning er sentrale motvekt mot denne utviklingen, hevder mange, men samtlige som omtaler temaet oppfatter dette som et særdeles intrikat og vanskelig problem å løse.

Når det gjelder NOKUT, NFR, og nasjonale evalueringer, er de aller fleste interessentene enige om at evalueringene i *meget* begrenset grad når ut til folk flest. De som evaluerer, blir evaluert, og arbeider med forsknings- og utdanningspolitikk er opptatt av evalueringer. Vår undersøkelse viser at vi ikke skal bevege oss langt bort fra forskningsproduksjon og høyere utdanning før evalueringer er noe man *kjenner til*, men ikke *kjenner godt*.

Svært få av våre respondenter kjenner *flere* evalueringer, og meget få utenfor UH-sektoren kjenner til evalueringer av både forskning og undervisning. Det er en klar tendens til at de som har blitt evaluert omtaler evalueringen(e) de selv har deltatt i, og at kjennskapet til øvrige evalueringer er relativt begrenset. Det er videre mange som har deltatt i evalueringer for få år siden, og som nå – relativt kort tid etterpå – ikke har et spesielt nært forhold til evalueringene.

Vi har sett at en del tyngre forskningsaktører har sett positive effekter av evalueringer, og disse fremstår også som meget godt kjent både med evalueringer og NFRs virke generelt. Dette er imidlertid en relativt liten gruppe. For mange i UH- og instituttsektoren er det oppmerksomhet om evalueringer over en kort periode i forkant og i etterkant av evalueringene. Dette gjelder altså interessenter som selv har blitt evaluert, og som arbeider med forskning og utdanning.

Utenfor UH- og instituttsektoren og helseforetak er det mye mindre kjennskap til og kunnskap om evalueringer. Dette gjelder direktorater, departementer, interesseorganisasjoner og arbeidslivsorganisasjoner. Det finnes en del unntak, men dette skyldes i stor grad at mange av interessentene i disse gruppene selv tidligere har vært i UH- eller instituttsektoren. Generelt fremkommer det at disse gruppene i liten grad følger med på evalueringer *generelt* og at kjennskap til evalueringer er relativt svak og knyttet til evalueringer som har vist seg å være spesielt relevant for visse sektorer, for eksempel ved å vise svakheter i forskningen på et felt, anvendelse av forskningen, e.l.

Vi kommer senere tilbake til dette poenget, når vi omtaler hvilke behov disse gruppene har for evalueringer. Poenget her er derimot å illustrere at selv aktører som antas å være nær forsknings- og utdanningssystemet, og regnes som interessenter, i liten grad leser og får med seg evalueringene som gjennomføres. Dette er en klar indikasjon på at evalueringene ikke når spesielt langt ut, og ikke har en spesielt stor rolle i å påvirke tillitt og legitimitet til forskning og høyere utdanning i befolkningen.

Dette inntrykket deles også av nær sagt alle interessentene: Evalueringene når i svært liten grad ut til folk flest. Flere av informantene har tenkt på muligheten for å gjøre noe med denne situasjonen. Én mulighet er at NFR eller andre kunne jobbet mer aktivt med *formidling* av den kunnskapen som genereres ved evalueringer. Dersom dette var en del av NFRs mandat kunne de jobbet mer aktivt med informasjons- og formidlingsarbeid. Det pekes i tillegg på detaljer som at NFRs nettsider oppfattes som lite brukervennlige, og vanskelige å finne frem på. I denne sammenheng omtales NOKUT som mer vellykkede, i den forstand at for eksempel Studiebarometeret *blir* brukt aktivt av befolkningen, og dermed i større grad bidrar til å informere om tilstanden i høyere utdanningen. Det er imidlertid, som nevnt i notatet *Typologisering*, usikkert om dette bør anses som en evaluering i ordets rette forstand.

Utdanningsevaluering

På spørsmål om effekter av evaluering av kvalitet i *utdanningen*, dukker Studiebarometeret svært ofte opp. Til tross for at dette ikke nødvendigvis bør omtales som en evaluering, er det tydelig at Studiebarometeret brukes svært aktivt, og har reelle og tydelige effekter i institusjonene.

Studentrepresentanter omtaler Studiebarometeret som viktig, og fremhever at undersøkelsen er stor og bred, samt at det er positivt at man kan se på utvikling over tid. Dette er spesielt viktig dersom man gjør endringer på et program, og da kan følge med på utviklingen av tilfredshet i etterkant. Det er imidlertid stor enighet blant samtlige interessentgrupper om at subjektive oppfatninger om tilfredshet *ikke* er en evaluering, eller tilstrekkelig for å uttale seg om kvaliteten i utdanningen.

Vi ser imidlertid at også ledere på forskjellige nivåer ved UH-institusjonene anser studiebarometer som *nyttig*. Som med andre type evalueringer gir de imidlertid ikke alltid spesielt overraskende resultater, men det har altså en verdi å få stadfestet ting man allerede vet. Det er også tydelig at Studiebarometeret fanger *ett* aspekt ved kvalitet i utdanningen. Det er dermed et viktig, men ikke et tilstrekkelig, verktøy. Det gir en god indikasjon på hvordan studentene opplever studiekvaliteten, men det må tas med «en klype salt». Man må også være oppmerksom på at tallene er noe konjunkturutsatt.

Når det gjelder *bruk* av Studiebarometeret forteller interessentene om store variasjoner. Resultatene brukes i liten grad på institusjonsnivå, og det er i hovedsak opp til instituttledere, studieledere, osv., å følge opp resultatene og iverksette tiltak. Dette innebærer at effektene av Studiebarometeret varierer, men det er tydelig at man ved de aller fleste institusjoner i det minste benytter Studiebarometeret som et innspill i løpende kvalitets- og utviklingsarbeid. Noen peker her på muligheter for å utvikle felles retningslinjer for hvordan resultatene kan og bør anvendes og følges opp.

Som tidligere omtalt finnes det få rene evalueringer av utdanning, men NOKUT har tidligere gjennomført noen slike. Ett eksempel er evaluering av allmennlærerutdanninger i 2006 og av ingeniørutdanninger i 2008 (etter oppdrag fra KD). I tillegg nevner flere respondenter en form for evaluering/tilsyn av sykepleierutdanningen ved institusjonen de tilhører, og denne sammenligner de med og omtaler som en evaluering.

Denne type evalueringer trekkes frem av interessentene som eksempler på at NOKUT *har* gjennomført evalueringer som til dels ligner de evalueringene NFR gjennomfører på forskningssiden. Disse evalueringene hadde klare effekter på utdanningene som ble evaluert. En av respondentene forteller at både institusjonen og fagmiljøene var meget responsive i etterkant av NOKUTs gjennomgang, bl.a. basert på konklusjoner om mangler og svakheter knyttet til akademisk miljø, forskningskompetanse, o.l. Dette førte til en rask og sterk satsning på akademisering av et fag som tidligere var meget profesjonsrettet og tradisjonelt. I løpet av relativt kort tid hadde utdanningen endret seg dramatisk. Dette brukes som et eksempel på hvordan evalueringer kan være særdeles effektive *styringsverktøy*, samtidig som interessenten også påpeker at man i etterkant opplevde at pendelen muligens svingte litt for mye i løpet av kort tid, slik at en del i fagmiljøet opplevde svakere kontakt med profesjonen.

Som nevnt i seksjon 3.5, fikk også NOKUTs evaluering av allmennlærerutdanningene i 2006 også store konsekvenser. Sammen med flere andre undersøkelser, pekte denne evalueringen på en del mangler i utdanningene, noe som satte i gang omfattende reformer av lærerutdanningene og resulterte i omleggingen av allmennlærerutdanningen til grunnskolelærerutdanningen i 2010. En av våre informanter understreker at utdanningsevalueringer på denne måten kan bli politisert og brukt til å understøtte visse endringer som bestemmes politisk. Dette er ikke negativt i seg selv, men det er ifølge informanten viktig å balansere politisk styring og faglig autonomi, samt å sørge for at evalueringer av utdanningskvalitet ikke blir brukt politisk på uhensiktsmessige måter.

Evalueringers rolle i kultur for kvalitetsforbedring

Neste tema er rollen evalueringer spiller i å skape en kultur for kvalitetsforbedring i norsk forskning og høyere utdanning. Tilnærmet alle aktørene i UH- og instituttsektoren trekker frem evalueringer som sentrale for kvalitetsarbeidet og som en viktig del av en såkalt kultur for kvalitetsforbedring. Dette gjelder aktører på alle nivåer. KD bruker evalueringer i sin styringsdialog med NFR, og i noe mindre grad overfor institusjonene. Ifølge KD selv ble fagevalueringer brukt mer systematisk overfor institusjonene tidligere, mens det nå brukes på et mer overordnet nivå.

Flere av UH-institusjonene bruker evalueringer aktivt i sine egne kvalitetssystemer. Generelt trekkes det frem at evalueringer har en skjerpene effekt – det bevisstgjør institusjonene og fagmiljøene og bidrar til at man får gjort opp status med jevne mellomrom. Én høyskole beskriver hvordan evalueringer er bygget inn i institusjonens kvalitetssystem, der nye evalueringer som kommer blir trukket inn i kvalitetsarbeidet, knyttet til programmer og til dekanens rapporter.

Enkelte av institusjonene trekker også frem *tilsyn* som en viktig del av kvalitetsarbeidet. Både i forkant av og i tilsynsprosessen får institusjonene anledning til å gjennomgå sine egne kvalitetssystemer og ha interne diskusjoner rundt kvalitetsarbeidet. En aktør nevner at når man først har laget og innført et kvalitetssystem, blir det kanskje tatt litt for gitt og ikke fulgt opp. Tilsyn skaper en sterkere bevissthet rundt hvordan kvalitetssystemene fungerer i praksis.

NOKUT fokuserer i stor grad på utvikling av kvalitetssikringssystemer. Dette arbeidet oppfattes av mange som både nyttig og viktig, men det er gradforskjeller i synet. Vi ser en tendens til at mindre og nyere institusjoner uttaler at de har større behov for og nytte av NOKUTs tilsynsarbeid enn større institusjoner. Dette er en oppfatning som deles av både store og små institusjoner. Dette knyttes også til et utbredt ønske om mer hjelp, bistand og støtte fra NOKUT, som vi kommer tilbake til i kapittel 5.

Både evalueringer og tilsyn oppleves altså som viktige for kvalitetsforbedring. Som det ble beskrevet av en aktør, er det «et nødvendig onde». Selv om det ofte oppfattes som

tids- og ressurskrevende mener de fleste institusjonene at det får positive konsekvenser for dem som institusjon. De opplever ikke bare at de skal levere dokumentasjon til en evaluator – de lærer også noe av prosessen. En organisasjon understreker at verdien av evalueringer avhenger av hvor godt prosessene er forankret i sektoren. Dersom man er bevisste på hva man ønsker å oppnå med evalueringene og planlegger godt, er det lettere å høste av *prosessen*, ikke kun resultatet.

For UH-institusjonene er imidlertid evalueringer kun én av faktorene som påvirker kvalitetsarbeidet. De er også underlagt andre føringer gjennom lovverk og ulike forskrifter. I tillegg gjøres det mye kvalitetsarbeid og utvikling av kvalitetssystemer internt på institusjonene, særlig på de større universitetene og høyskolene. En representant for et stort universitet sier at den viktigste driveren for utvikling av kvalitet er det de klarer å skape internt på institusjonen og at det er sterkt forankret i ledelsen. Her varierer det mellom de store og de mindre institusjonene hvor mye ressurser og kompetanse som finnes internt. En av de mindre høyskolene trekker for eksempel frem at de gjerne vil ha mer læring og erfaringsutveksling på tvers av institusjoner gjennom for eksempel læringskonferanser i etterkant av evalueringene.

For noen aktører, for eksempel en privat høyskole, fremkommer det tydelig at de oppfatter *interne* og *internasjonale* kvalitetssikrings- og akkrediteringssystemer som viktigere for kvalitet i egen institusjon enn evalueringene som gjennomføres av NFR og NOKUT. De internasjonale systemene de deltar i er sammensatt av lignende institusjoner, og innebærer at andre deltakere evaluerer institusjonen, mens de selv også besøker og deltar i evalueringer av andre institusjoner. Dette oppfattes som svært nyttig, og dette er også tett knyttet til egne interne kvalitetssikringssystemer. Det kan se ut til å være en viss oppfatning blant de private interessentene vi har intervjuet at disse er fullstendig avhengig av å prestere godt i et konkurranseutsatt marked, og at dette medfører et noe annet fokus på kvalitetsutvikling og -arbeid enn man kan finne i andre institusjoner som er mindre konkurranseutsatt.

Oppsummering

Sentrale motsetninger og aktører med spesielle interesser

Formativ vs. summativ evaluering	Forskningsprodusenter
Institusjonsutvikling vs. forskningspolitikk	KD, NFR
Institusjonsutvikling vs. anvendelse av forskning	Forskningsprodusenter, departementer, arbeidslivsorganisasjoner

Hovedpoenger

- Evalueringer brukes aktivt og verdsettes høyt av mange tunge forskningsprodusenter
- Evalueringer brukes aktivt i utvikling av forsknings- og utdanningspolitikk
- Mindre tydelig nytte for en del mindre institusjoner
- Marginal bruk og små effekter utenfor kjernemålgruppen for evalueringer

5 Behov for evalueringer

I det følgende presenterer vi hvilke behov aktørene har gitt uttrykk for, både når det kommer til konkrete evalueringer og evalueringssystemet mer generelt. Her har vi latt respondentene svare fritt, og dette er dermed en del der vi ser fremover og lar respondentene formidle sine ønsker og behov, uavhengig av eksisterende evalueringer og evalueringssystem. Likevel bygger de fleste av respondentene sine ønsker og behov på tidligere erfaringer med evalueringer og hvilke evalueringer de har oppfattet som interessante og nyttige. Generelt er aktørene vi har intervjuet positive til evalueringer, og mener at evalueringer er nyttige. Det er likevel delte meninger om hva slags type evalueringer det er behov for, og visse skillelinjer mellom de ulike interessentgruppene fremtrer her.

Vi vil her først presentere de ulike aktørenes behov for evalueringer av forskning, og hva slags typer evalueringer de etterspør, før vi går videre til behov for evalueringer av utdanning. Deretter presenterer vi aktørenes syn på hvordan evalueringer kan og/eller bør ta hensyn til ulike institusjoners formål og forutsetninger, og hvordan dette kan balanseres mot et behov for sammenligning og benchmarking. Til sist diskuterer vi aktørenes behov for mer koordinering og samordning på evalueringfeltet. Mange aktører etterspør en mer *helhetlig* tilnærming til hva evalueringer er og hva de skal brukes til samt mer koordinering mellom de ulike aktørene på feltet.

5.1 Behov for evalueringer av forskning

Fagevalueringer

Som nevnt i de foregående kapitlene, har en stor andel av informantene trukket frem NFRs fagevalueringer som svært nyttige, og mange har følgelig understreket at det er slike evalueringer de ønsker og har behov for. Dette gjelder informanter på tvers av alle interessentgruppene – både universiteter, høyskoler, institutter, helseforetak og til dels departementer og direktorater. Vi har allerede diskutert verdien av denne type evalueringer og gjentar her noen av hovedpunktene.

Flere av informantene nevner fagevalueringer som et viktig supplement til store, kvantitative undersøkelser og bibliometri. Sammenligning av institusjoner (generelt, og ikke fag mot fag) blir av mange oppfattet som for overordnet. Det kan bli som å sammenligne «epler og pærer». Det blir oppfattet som mer nyttig og interessant for de som underviser og forsker selv med mer dyptgående sammenligning innenfor et fagområde. En informant mener at fagevalueringene også kunne vært brukt i enda større grad i sektoren, med oppfølgingsseminarer i etterkant og erfaringsutveksling rundt hvordan man arbeider med kvalitetsforbedring etter en evaluering.

Fra et myndighetsperspektiv understreker mange av informantene vi har intervjuet fra departementer og direktorater at fagevalueringer kan gi en god oversikt over forskningen på deres fagfelt. Representanter for ett av departementene mener det er stor enighet om nytten av fagevalueringer og at slike evalueringer kan bidra til en oversikt over helheten i forskningen på et felt i stedet for mer oppstykket rapportering fra mange ulike prosjekter og programmer finansiert av ulike aktører.

Det er delte meninger blant informantene om hvor brede fagevalueringene skal være. Enkelte informanter ser det som positivt at man har gått fra å evaluere enkeltdisipliner til å dekke bredere fagområder, som for eksempel i SAMEVAL. En representant for en privat høyskole sier imidlertid at de syntes SAMEVAL ble for bred, og at de heller kunne ønske seg nasjonale evalueringer og sammenligninger på mindre eller mer konkrete

fagområder. Informanten uttrykker likevel forståelse for at NFR ikke vil ha kapasitet til dette.

Den vanskelige balansegangen mellom bredde og frekvens tas opp av flere aktører. Mange synes at det kunne være nyttig med evalueringer av mer konkrete temaer eller mindre fagområder. Dette gjelder særlig departementer og direktorater som etterspør evalueringer av norsk forskning innen konkrete temaer – ofte temaer knyttet til politiske satsningsområder. I intervjuene blir temaer som vold i hjemmet og dyrehelse nevnt. Interessentene ønsker her å vite mer om forskningen på feltet samt hvilke kunnskapshull som finnes. Slike temaer vil imidlertid ofte være tverrinstitusjonelle og tverrfaglige, og det kan være utfordrende å avgrense hvem som skal dekkes av en slik evaluering. Å gå inn på smale temaområder vil imidlertid også føre til flere og mer detaljerte evalueringer, noe som igjen kan føre til at syklusene blir lengre og at det tar lengre tid mellom hver gang et fagområde blir evaluert. De fleste mener at det ikke er nødvendig med evalueringer hvert 5. år, men at hvert 10. år likevel blir litt sjeldent.

En annen utfordring med fagevalueringene, som nevnes av flere, er problemene rundt det å inkludere tverrfaglighet på en god måte. Dette har vi diskutert i seksjon 3.4 og vi diskuterer det også i neste kapittel om utviklingstrekk. På bakgrunn av disse utfordringene ønsker flere aktører – særlig institutter og mer tverrfaglig orienterte institusjoner – at evaluator må finne en bedre måte å håndtere tverrfaglighet på dersom man skal få en reell vurdering av *all* forskningen som foregår i Norge.

Til sist etterspør én aktør bedre nasjonale synteser på bakgrunn av fagevalueringene. Vedkommende mener at man utvikler generelle beskrivelser og synteser på et litt for tynt grunnlag, som deretter får stor betydning for utvikling av forskningspolitikk og styring av sektoren. Det bør legges mer vekt på å utvikle gode nasjonale synteser som bygger på et bredt spekter av ulike evalueringer.

Institusjonsevalueringer

Etter flere store runder med fagevalueringer, har NFR nå ytret ønske om å foreta en dreining mot institusjonsevalueringer. De ønsker at institusjonene tar mer eierskap til evalueringene, og at evalueringene skal være nyttige for institusjonene selv. Dette kan innebære å sette en relativt fast ramme, men også ha noe fleksibilitet der institusjonene selv kan sette indikatorer for sammenligning ut ifra hvor de forventer å være og hvem man ønsker å sammenligne seg med. I tillegg kan man i institusjonsevalueringer i større grad evaluere institusjonenes ulike virksomheter sammen, og dermed vurdere utdanning og forskning i sammenheng samt evaluere tverrfaglighet på en bedre måte enn i de tradisjonelle fagevalueringene.

Ifølge KD vil institusjonsevalueringer bli en annen type evaluering der man i større grad vurderer de organisatoriske betingelsene for å drive fremragende forskning, med fokus på systemer og prosesser. En informant fra KD sier at dette kan gi annen informasjon enn fagevalueringene og være til nytte for institusjonene.

Flere av informantene vi har snakket med stiller seg også positive til institusjonsevalueringer. En leder ved et stort universitet understreker at en slik evaluering vil være mer nyttig og lærerik for ledelsen enn en tradisjonell fageevaluering. Som leder for et fakultet er det mer nyttig å gjennomgå en evaluering der både forskning og utdanning samt ledelse, organisering og strategi, blir sett i sammenheng. Andre trekker frem at slike evalueringer kan bli enda mer formative og dermed i større grad bidra til institusjonenes selvutvikling. De kan i større grad enn de nåværende fagevalueringene ta hensyn til institusjonenes ulike mål og rammebetingelser, noe som av mange fremheves som positivt (se nærmere om dette i seksjon 5.3).

En informant fra et forskningsinstitutt sier at det ville være svært nyttig for instituttet å få en detaljert vurdering av hvordan instituttet fungerer, for eksempel hvor i organisasjonen det er dynamikk og fornyelse, om man bruker potensialet i kontakter og nettverk, osv. Likevel mener samme informant at dette er urealistisk med tanke på tilgjengelige ressurser og at et slikt detaljnivå – og hva det krever metodisk – vil koste for mye.

Det er som vi tidligere har sett også en del av informantene som har uttrykt bekymring eller skepsis når de blir spurt om en slik eventuell overgang til institusjonsevalueringer. Det er særlig det at man kan miste sammenlignbarheten som vekker bekymring. En representant for en mindre høyskole fremhever at sammenligning nasjonalt er svært viktig for dem for å vite hvordan de hvordan de når opp på ulike fagområder og hvordan de kan utvikle seg videre. Fra departementenes (utenom KD) og direktoratenes perspektiv er institusjonsevalueringer mindre relevante, ettersom institusjonenes kvalitet er KDs ansvar. Disse er mer opptatt av hvordan det står til med forskningen i deres fagområder, og fagevalueringer vil gi et bedre helhetsbilde på dette.

Videre er flere av informantene vi har snakket med usikre på hvordan slike institusjonsevalueringer vil skille seg fra andre typer evalueringer og kvalitetsarbeid. Mange trekker paralleller til NOKUTs tilsynsvirksomhet der man ser på institusjonenes egne kvalitetssikringsrutiner. Enkelte mener at det kan være nyttig å se på hva institusjonene gjør for å legge til rette for at de ansatte lykkes, mens andre uttrykker bekymring og mener at slike evalueringer i hvert fall bør sees i sammenheng med det NOKUT gjør for å unngå overlapp. Flere representanter for store universiteter stiller seg kritiske til forslaget og mener at man ikke bør evaluere forskning på samme måte som NOKUT evaluerer utdanning. De mener derimot det er nødvendig å evaluere resultater, ikke bare prosesser, for å kunne vurdere forskningskvalitet. Det blir også nevnt at institusjonsevalueringer blir som en slags «gratis konsulentarbeid» der man får noen utenfra til å komme og vurdere organisasjonens styrker og forbedringsområder. Dette kan være svært positivt, men man må avgrense NFRs arbeid mot ren konsulentvirksomhet. Til sist peker noen informanter på internasjonale initiativer i EU og OECD som også evaluerer institusjoner og hvor man må unngå overlapp og dobbeltarbeid.

Det er også forskjeller mellom informantene når det kommer til hva de gjør av evalueringer internt på institusjonen og hvordan dette vil passe sammen med eventuelle institusjonsevalueringer fra NFR. Informanter fra et stort universitet forteller for eksempel at det har vært diskutert innad i institusjonen om de skulle foreta mer helhetlige institusjonsevalueringer etter modell fra Sverige, men at dette ble møtt med stor motstand og ikke ble gjennomført. En informant fra en annen institusjon forteller derimot at de allerede har innført jevnlig evalueringer internt, og at institusjonsevalueringer gjort av NFR ikke ville gi noen merverdi utover dette. Ifølge denne informanten kan man godt vedta at alle institusjoner må gjennom en institusjonsevaluering for eksempel hvert 5. år, men at institusjonene i stor grad kan ta initiativ til dette selv uten at NFR nødvendigvis trenger å spille en stor rolle i dette. Et slikt evalueringssystem kan sammenlignes med systemet i Nederland, der institusjonene selv initierer evalueringer basert på en detaljert protokoll utarbeidet av det nederlandske forskningsrådet. Innføring av et slikt system i Norge vil kreve endringer både i institusjonene og i NFR. Dersom institusjonene selv skal ha ansvar for evalueringer kreves det at de har kapasitet til dette, og at dette arbeidet prioriteres. Det er her viktig å påpeke at interessentene ser ut til å omtale egeninitierte evalueringer med eksterne eksperter, og ikke interne evalueringer. NFR vil i så tilfelle kunne få en rolle som tilrettelegger og kvalitetssikrer av evalueringene. Institusjonene selv kan enten tilegne seg tilstrekkelig kompetanse og kapasitet til at de selv kan koordinere evalueringen, eller de kan tenkes å sette ut dette til andre aktører. Én

interessent omtalte for eksempel NIFU som en aktuell samarbeidspartner, og NIFU har også tidligere vært tungt inne i administrasjonen og koordineringen av evalueringer.

Generelt viser interessentanalysen at informantene forbinder positive aspekter og utfordringer med både fagevalueringer og institusjonsevalueringer. Det kan virke som at flere er positive til fagevalueringer enn institusjonsevalueringer, men dette kan være fordi de allerede er kjent med fagevalueringene og har sett nytten av dem. Mye av skepsisen rundt en eventuell overgang til institusjonsevalueringer skyldes usikkerhet om hvordan disse evalueringene vil fungere og hvordan de vil passe inn i evalueringssystemet generelt. Flere understreker imidlertid også at disse to evalueringstypene kan gi ulike resultater og ha ulik nytteverdi og at man ikke er nødt til å begrense seg til én type. Det å variere mellom dem kan være nyttig i seg selv.

Internasjonale eller nasjonale evalueringer

Som nevnt under avsnittet om benchmarking i seksjon 4.1, er det forskjeller mellom UH-institusjonene når det gjelder hvem de ønsker å sammenligne seg med og om de primært har et nasjonalt eller internasjonalt perspektiv. Et mønster som tydelig kommer frem er at de store breddeuniversitetene samt institusjoner som anses som ledende innen sitt område legger stor vekt på internasjonale evalueringer og sammenligninger og at det er dette som oppfattes som viktigst. Dette begrunnes med at disse institusjonene har et «høyt ambisjonsnivå». En privat høyskole sier for eksempel at de selv har igangsatt jevnlig forskningsevalueringer med internasjonale komitéer. Denne type løsning kan også sees i sammenheng med forslaget om et evalueringsregime som i større grad legger opp til at alle forskningsprodusentene initierer og har ansvar for å gjennomføre evalueringer. En informant fra et stort universitet sier at internasjonal benchmarking oppfattes som det viktigste, men at de ikke er så glade i de store internasjonale rangeringene. I stedet ønsker de seg benchmarking mot noen utvalgte universiteter i andre land.

Alle institusjonene fremhever likevel at nasjonale evalueringer også er relevante. Flere representanter for mindre og/eller regionale høyskoler har sagt at de har behov for nasjonale sammenligninger fordi det er her de oppnår høyest grad av læring. Likevel kan også disse institusjonene ha behov for internasjonal referansetesting for å vise at de er på et visst nivå internasjonalt, for eksempel i forbindelse med universitetssøknader. Dette har vært et tema for en rekke mindre UH-institusjoner i forbindelse med strukturreformen.

Spørsmålet om nasjonalt versus internasjonalt perspektiv handler ikke bare om hvem som evalueres, men også om hvem som *evaluerer*. Her fremhever de fleste at det er positivt å bruke internasjonale paneler i nasjonale evalueringer. På denne måten kan man sikre seg at norsk forskning ikke foregår i en «boble», men at man er påkoblet det som skjer i andre land. Internasjonale paneler har en positiv effekt på evalueringen i seg selv ved at det kan bidra til læring og erfaringsutveksling innad i panelene rundt hva som evalueres og hvordan. Likevel påpeker enkelte institusjoner at det også er nødvendig å sikre at internasjonale paneler har tilstrekkelig kunnskap om og forståelse for den norske konteksten.

Noen av respondentene understreker at et norsk rammeverk for evalueringer også må ta høyde for hva som gjennomføres av evalueringer internasjonalt, særlig i regi av EU og OECD. For eksempel trekkes det frem som et alternativ å lage et nasjonalt rammeverk som kan kobles på det som allerede gjøres av internasjonale sammenligninger. På den måten kan man få til både nasjonal og internasjonal benchmarking uten dobbeltarbeid. Andre aktører trekker også frem at det finnes andre former for internasjonal vurdering

enn evalueringer – for eksempel gjennom EU-søknader og andre søknader om internasjonal finansiering og internasjonalt samarbeid.

I tillegg til at enkeltinstitusjoner blir evaluert i en internasjonal sammenheng, kan det også være interessant å se hvordan ulike fagområder i Norge står seg sammenlignet med andre land. Her vises det blant annet til sammenligninger utført av OECD, som for eksempel har vist at norsk samfunnsvitenskap er i verdensklasse. Det komparative perspektivet kan være nyttig – særlig i forbindelse med omdømmebygging internasjonalt som i sin tur kan ha positive konsekvenser for internasjonale prosjektsøknader og samarbeid.

Evaluering av samfunnsrelevans

Et aspekt som særlig kommer til uttrykk blant myndighetsaktører og arbeidslivsorganisasjoner er *relevans* og behovet for å vurdere dette i evalueringer. Dette er også tett knyttet til hvordan de ulike aktørene definerer hva som er kvalitet i forskning. Blant myndighetsaktører er det tydelig at relevans er et viktig kvalitetstegn, noe som blant annet ble uttrykt slik av en representant for et av departementene: «kvalitet er noe vi kan bruke til noe». Kvalitet blir dermed ikke sett som et mål i seg selv, men heller en viktig forutsetning for å kunne bruke forskningen.

Der representanter for UH-institusjoner i større grad legger vekt på grunnforskning og at forskning er et mål i seg selv, uttrykker mange myndighetsaktører at forskningen er til for å svare på samfunnets behov. Én informant sier at det naturligvis er viktig at forskning er grundig fagfellevurdert, men at den også bør være anvendelig for aktører som skal løse problemer i samfunnet. Her bør NFR fokusere mer på implementering og bruk av forskningsresultater og hva som kommer ut av forskningen utover publiseringer. Dette handler ikke bare om at forskning må være tematisk relevant, men også om at det må være en viss dynamikk og responsivitet i forskningsmiljøer slik at de kan klare å svare på samfunnets behov når nye utfordringer dukker opp. Her trekkes COVID-19 frem som et eksempel på en situasjon der mange forskningsmiljøer har vært raske til å omstille seg og produsere nyttig kunnskap for beslutningstakere.

5.2 Behov for evalueringer av utdanning

Evaluering av utdanningskvalitet er et relativt nytt fenomen sammenlignet med evalueringer av forskning. En representant for KD uttaler at utdanning og undervisning tradisjonelt har vært opp til den enkelte lærer, men at det har skjedd mye de siste årene for å få det litt mer ut i åpent landskap slik at man kan evaluere og lære av hverandre. Vedkommende mener at utdanningsevalueringer er like viktige som forskningsevalueringer, men at de i større grad er et verktøy for NOKUT selv enn for KD. Dette kan skyldes at NOKUT er et faglig uavhengig forvaltningsorgan som tar selvstendige beslutninger om eventuelle konsekvenser av utdanningsevalueringene, for eksempel tilsyn. Representanten mener videre at nytten av utdanningsevalueringer forutsetter at man utvikler en god metodikk.

Landskapet innen evalueringer av utdanning virker å være mer fragmentert enn på forskningssiden. Det finnes som nevnt en del større nasjonale undersøkelser som Studiebarometeret og NIFUs kandidatundersøkelser. I tillegg finnes det andre studentundersøkelser som ikke går direkte på kvalitet, men som likevel er en del av helheten, slik som levekårsundersøkelser og studentenes helse- og trivselsundersøkelse (SHOT-undersøkelsen). Institusjonenes egne kvalitetssystemer skal sikre kvalitet i utdannelsen. NOKUT utfører jevnlig tilsyn. I tillegg driver NOKUT akkreditering av institusjoner og utdanningsprogrammer, som også innebærer en viss

form for evaluering. NOKUT har også gjort enkelte overordnede evalueringer av fagområder, som i større grad har fulgt samme modell som NFR bruker på forskningssiden.¹ I tillegg, som en representant fra KD nevner, kan man også se på eksamen og studentenes resultater som en evaluering av studentenes læringsutbytte.

Det at evalueringer av utdanningskvalitet er relativt nytt og at landskapet er mer uoversiktlig, gjør at det er vanskelig å ha en klar definisjon eller forståelse av hva som er evalueringer og ikke på utdanningssiden. NOKUT selv anser ikke Studiebarometeret eller tilsyn og akkreditering som evalueringer, men i våre intervjuer har mange nevnt dette når de har fått spørsmål om evalueringer av utdanningskvalitet.

I hvor stor grad trenger man evalueringer av utdanningskvalitet?

Det virker som om mange av de store, kvantitative undersøkelsene som Studiebarometeret, NIFUs kandidatundersøkelser, SHOT-undersøkelsen og andre er godt innarbeidet i sektoren, selv om utformingen av dem er gjenstand for diskusjon. På samme måte er NOKUTs akkrediterings- og tilsynsarbeid en form for vurdering og evaluering som institusjonene legger stor vekt på.

Det er delte meninger blant våre informanter om man behøver evalueringer av utdanningskvalitet utover dette og i så fall hva slags evalueringer det er behov for. Enkelte aktører mener at man ikke trenger flere evalueringer av utdanningskvalitet – med mindre man får noen indikasjoner på at kvaliteten ikke er tilstrekkelig. Disse aktørene mener at utdanningskvalitet er et institusjonelt ansvar og etterspør større tillit til institusjonene og deres kvalitetssikringsarbeid.

Andre aktører peker på at mange av institusjonene etter hvert har blitt i stand til å evaluere og analysere en del ting selv – gjennom bruk av allerede foreliggende data og bruk av analyseprogrammer, som for eksempel Tableau. En representant for et stort universitet understreker at de ved bruk av Tableau har fått langt bedre muligheter til å gjøre analyser selv og at de ikke ønsker noen flere store, nasjonale undersøkelser. NOKUT deler også delvis dette synet og legger vekt på at det er viktig å tenke kost-nytte i evalueringsarbeidet. Det er viktig å ikke ilegge institusjonene en større arbeidsmengde. Ifølge NOKUT har mange institusjoner kommet så langt på dette området at de ikke trenger et stort regelverk, styring og kontroll, men heller tillit og muligheten til å ivareta sin egen art.

Andre aktører uttrykker tvil om utdanningsevaluering og -analyser bør overlates helt til institusjonene selv. Én mener for eksempel at det også er nødvendig med et nasjonalt evalueringsrammeverk for å ivareta nødvendig sammenligning og benchmarking. En annen mener at selv om institusjonene helt klart evner å drive kvalitetssikring og utviklingsarbeid, så kan de også være konservative og la være å gjennomføre endringer før de må. Her kan eksterne evalueringer utfordre institusjonenes eget arbeid og være viktige katalysatorer for endring.

Flere aktører uttaler også at store, kvantitative undersøkelser ikke er tilstrekkelig for å måle kvalitet i utdanningen og at man også må gå litt dypere og mer kvalitativt til verks. Samtidig er det svært forskjellige syn på hvordan slike evalueringer bør gjennomføres og hva som skal måles.

¹ Så langt har NOKUT gjennomført fire slike fagevalueringer: evalueringen av allmennlærerutdanningene i 2006, ingeniørutdanningene i 2008, førskolelærerutdanningene i 2010 og KOMBEVAL i 2018. Våren 2020 startet arbeidet med en evaluering av lektorutdanningen for trinn 8-13.

Fagevalueringer

På spørsmål om hva forskjellige interessenter har behov for av evalueringer av utdanning, havner diskusjonene raskt over på mulighetene for evalueringer som ligner NFRs fagevalueringer, som NOKUT allerede har gjennomført noen av. Dette ville innebære at NOKUT brukte ekspertpaneler for å evaluere alle læresteder som tilbyr utdanning innen et gitt fagområde. Denne type periodiske fagevalueringer for utdanning virker å være noe mange ønsker seg, og flere uttrykker at det kunne være svært interessant. Én aktør mener at det kan gi en god oversikt over alle utdanningsprogram innen et fagområde, slik at man kan få informasjon om forskjellene mellom dem, både i tematikk, undervisningsmetoder og kvalitet. Dette kan gjøre det lettere både for studenter som skal velge studieprogram og for arbeidsgivere som skal rekruttere ansatte. Interessenten peker på at man her ofte baserer seg på en overbevisning om at de gamle, store UH-institusjonene er bedre enn for eksempel små, profesjonsrettede høyskoler, men at dette ikke nødvendigvis stemmer.

En annen aktør, en privat høyskole, sier det er svært viktig for dem å delta i nasjonale fagevalueringer fordi slike evalueringer kan gi et inntrykk av hva «bransjestandarden» er. Dersom de skal ta betalt for tjenester som studentene kan få gratis et annet sted, må de kunne vise til at de også kan tilby høyere kvalitet.

Til slutt forteller en representant for en arbeidslivsorganisasjon at flere av de ulike profesjonsutdanningene ønsker flere evalueringer for å synliggjøre forbedringspunkter og et behov for mer ressurser. Spesielt trekkes sykepleier- og lærerutdanningene frem som eksempler på utdanninger som ofte scorer dårlig på Studiebarometeret og hvor ansatte opplever utfordrende arbeidsforhold. Evalueringer kan bidra til å synliggjøre mangler ved ulike fagområder.

Når vi da går over på hva de ønsker at slike fagevalueringer skal legge vekt på handler svært mye om *relevans* i utdanningene. Spesielt departementer og arbeidslivsorganisasjonene er åpne for at fagområder i større grad evalueres på nasjonalt nivå for å kartlegge kvalitet i utdanning, med spesielt fokus på relevans og kobling til fremtidige samfunns- og arbeidslivsbehov. Departementer og direktorater er særlig opptatt av høy kvalitet i utdanningene innen deres sektorer for å sikre at fremtidige ansatte – både i egen organisasjon og i sektoren – har den nødvendige kompetansen. For eksempel er både KD og Udir opptatt av lærerutdanningene og hvordan disse sikrer høy kvalitet og relevant kompetanse for praksisfeltet. Også forskjellige forskningsinstitusjoner som ikke selv har undervisning er positive til en slik idé. Institutter er avhengige av at UH-sektoren utdanner relevante forskere, og spesielt innenfor helsesektoren er det mange aktører som har behov for at de mange helsefagene er gode og relevante.

Selv om mange interessenter altså mener at fagevalueringer også ville være nyttige på utdanningssiden, er det få som anser det som realistisk at NOKUT eller andre skal kunne gjennomføre slike evalueringer på alle utdanningsområder uten en radikal endring av evalueringssystemet og evalueringsorganene.

Kvantitativ eller kvalitativ innretning

En stor utfordring i denne sammenheng er *hvordan* man skal evaluere kvalitet i utdanning. I dag er dette i stor grad basert på «papirøvelser» der man for eksempel vurderer undervisernes pedagogiske kompetanse og forskningskompetanse, pensumliste, læringsmål og studieplaner. I tillegg anvendes data og indikatorer fra ulike spørreundersøkelser, eksamensresultater og lignende.

Noen sentrale aktører omtaler også muligheten for å utvikle et utvidet kvalitetsbarometer for utdanning, à la Studiebarometeret, men mer omfattende. Her kunne for eksempel arbeidsmarkedsundersøkelser, eksamensresultater, og muligens også andre type indikatorer og evalueringer, vært grunnlag for mer omfattende kvalitetsindikatorer for forskjellige utdanningssteder og utdanninger.

I tillegg til kvantitative data og indikatorer, er det imidlertid også mange som etterspør mer «kvalitativ innsikt» i hva som gir kvalitet, og hvor det er god/dårlig kvalitet. En representant for en studentorganisasjon mener for eksempel at man burde gå i dybden på hva som fører til god kvalitet i en undervisningssituasjon. Hva fungerer for eksempel i en stor forelesning med 300 studenter versus et masterseminar med 10 studenter? Her etterspør de eksperter med pedagogisk kompetanse som kan observere undervisningen, men uttrykker også at de innser at dette er praktisk vanskelig å gjennomføre i stor skala.

Studentrepresentanter trekker også frem frafall som et viktig tema. Det finnes statistikk for hvor mange som faller fra utdanningen, men svært lite kunnskap om hvorfor. Det er behov for en dypere analyse av årsaker for å kunne gjøre noe med problemet. Slike mer kvalitative dypdykk i ulike temaer kan være et viktig supplement til de mange kvantitative studentundersøkelsene.

Flere aktører peker også på at utdanningsevalueringer mangler kvalitative vurderinger av fagmiljøene som tilbyr utdanningen. Flere respondenter peker på dette som et «kjempeshull» i NOKUTs virke, ved at man har mange prosedyrer, omfattende forskrifter, studiebarometer, osv., men mangler kvalitative vurderinger som gir inngående kunnskap om faktisk kvalitet i utdanningen.

Nasjonalt/internasjonalt

Aktørene har forskjellige meninger om hva som er viktigst av nasjonalt eller internasjonalt fokus i evaluering og benchmarking av utdanning. Representanter for et stort universitet sier de har behov for internasjonal sammenligning også på utdanningsområdet, men er usikre på om det da bør være i NOKUTs regi eller om det bør legges under EU. Ettersom det er mye samarbeid knyttet til utdanning i EU, og mange EU-land har tilsvarende kvalitetssystemer, kan det være en god idé at slik internasjonal sammenligning gjøres i en EU-kontekst.

En representant for KD mener at evaluering av utdanning primært bør være et verktøy for institusjonene, med noe benchmarking nasjonalt. En representant fra en statlig høyskole sier også at de har nytte av nasjonale evalueringer og at disse skaper mye læring hos institusjonen. Her understrekes det at NOKUT har skapt gode modeller for hvordan dette kan gjøres og at dersom NOKUT går opp nye spor med tanke på metodikk og modeller for utdanningsevaluering, er det lettere for institusjoner å følge etter i eget kvalitetsarbeid.

Studentenes rolle i evalueringer

I diskusjoner rundt utdanningskvalitet er mottakerne – altså studentene – sentrale. Likevel er det forskjellige oppfatninger om hvordan disse skal inkluderes i utdannings-evalueringer og i hvilken grad studentene i det hele tatt kan uttale seg om kvalitet i utdanningen.

I den ene enden av skalaen finner vi en representant for en studentorganisasjon, som mener at det er studentene som bør ha førsterett til å evaluere utdanningen ettersom det er de som er mottakerne og det er deres læring som er målet. Likevel er representanten noe kritisk til om studentene faktisk vet hva de har krav på eller hva utdanningskvalitet faktisk innebærer. Som eksempel nevnes målet om at utdanning skal være

forskningsbasert. Vedkommende mener at studentene ikke i stor nok grad vet hva dette innebærer. Én informant fra et stort universitet understreker behovet for å inkludere studentene i utdanningsevalueringer, spesielt for å ivareta et godt samhandlingsklima mellom studenter og akademia. Andre informanter er kritiske til at studenter skal være så sentrale i evalueringer av utdanningskvalitet, men at de kan uttale seg om *opplevd kvalitet* og at dette er en dimensjon som bør inkluderes. Likevel mener mange at studentenes vurdering av kvalitet ofte kan være påvirket av hva som oppleves som «lettest» og at mer krevende – men beviselig mer effektive – pedagogiske metoder dermed vil score dårligere på Studiebarometeret enn tradisjonelle forelesninger.

Både studentorganisasjonen og andre informanter er enige om at de store kvantitative undersøkelsene om studenttilfredshet ikke gir tilstrekkelige mål på kvalitet og at de må suppleres med mer dyptgående evalueringer og analyser. For eksempel kan man se at det har skjedd endringer i studenttilfredsheten for et program eller studium i Studiebarometeret og dermed iverksette mer omfattende evaluering av hva som har skjedd.

Arbeidslivsrelevans og dimensjonering av studietilbudet

På samme måte som med forskning, er aktører som departementer, fagforeninger og til dels institutter mest opptatt av *relevans*. De har ikke så sterke meninger om utdanningskvalitet, utover at det er viktig for dem som arbeidsgivere at de får gode kandidater med oppdatert og relevant kunnskap om feltet de skal arbeide med.

Arbeidslivsorganisasjoner er opptatt av at utdanningssystemet er lydørt overfor arbeids- og næringsliv og at det er stor grad av samarbeid mellom UH-institusjonene og næringslivet, for eksempel i form av praksis for studenter.

En representant for et departement uttaler at samfunnsutfordringer er «kompasset» for hvordan institusjonene skal jobbe. Interessenten spør seg om man for eksempel kan inkludere overordnede temaer og samfunnsproblemer i læreplaner og/eller læringsmål i høyere utdanning, slik som det nå gjøres i grunnskole og videregående skole. Dette kan for eksempel være temaer som bærekraft og digitalisering. Informanten mener at man bør evaluere i hvor stor grad utdanningsinstitusjoner inkluderer slike store samfunnsutfordringer i utdanningen.

Enkelte informanter etterspør også evalueringer av dimensjoneringen av utdanningstilbud og om det finnes områder der man mangler kompetanse eller studieprogrammer. For eksempel uttaler en informant fra et stort universitet at det kunne være nyttig å få en evaluering av hvorvidt institusjonens studieprogramportefølje svarer til deres samfunnsoppdrag på aggregert nivå. Ett institutt nevner videre at det er enkelte fagområder de dekker som ikke er en del av utdanningstilbudet i Norge og at de derfor sliter med å rekruttere norske forskere. Derfor kunne det vært nyttig med et helhetlig blikk på utdanningstilbud og fagområder for å undersøke mulige hull i tilbudet.

En representant for en arbeidslivsorganisasjon sier at utdanningstilbudet kanskje er litt i overkant søkerstyrt. Man kan derfor vurdere en evaluering der man ser på om utdanningstilbudet i Norge dekker både arbeidslivets og samfunnets behov og om det er «dimensjonert for fremtiden».

5.3 Forskjellige formål og forutsetninger

Svært mange av informantene vi har snakket med trekker frem utfordringene ved å evaluere til dels svært forskjellige institusjoner i samme evalueringssystem og etter samme kriterier. Vi har beskrevet forskjellene mellom de ulike typene institusjoner i seksjon 3.4. På bakgrunn av dette har flere av de vi har snakket med ytret et behov for

at institusjonenes ulike formål, roller, forutsetninger og rammebetingelser i større grad tas hensyn til i evalueringer.

Et eksempel som trekkes frem av mange er for eksempel instituttens rolle i de store fagevalueringene til NFR. I SAMEVAL var det obligatorisk for UH-institusjonene å delta, mens det var frivillig for instituttene. Her forteller et institutt at det er vanskelig velge om de skulle bli med i evalueringen. På den ene siden anså de det som positivt å bli evaluert, både for å få en sammenligning med andre samfunnsvitenskapelige miljøer i Norge og for å kunne få tips til forbedringspunkter for instituttet. På den annen side vil et institutt i en slik evaluering kun bli evaluert på deler av sin virksomhet, ettersom oppdragsforskning ikke blir inkludert. Her er det også en aktør som trekker frem at panelene ofte er «smale» og svært akademisk orientert, slik at samfunnsmessig «impact» og anvendt forskning ikke tillegges tilstrekkelig vekt. Som institutt vil man dermed bli målt etter kriterier som ikke er tilpasset deres organisasjon og virksomhet. For eksempel er det vanskelig å måle et institutt som er avhengig av å drive mye med oppdragsforskning opp mot et universitet som i større grad kan fokusere på grunnforskning.

På grunn av slike utfordringer har det derfor vært uttrykt tvil blant våre informanter om man bør evaluere UH-institusjoner og institutter sammen. Dersom man fortsetter med et lignende evalueringssystem som i dag, mener flere at institutter bør holdes utenfor. Instituttene på sin side mener det er nyttig for dem å være med i evalueringer, men de vektlegger at de ulike forutsetningene og rammebetingelsene da må tas hensyn til. Én representant for et institutt understreker også at den delen av et institutts virksomhet som foregår i markedet, som oppdragsforskning, også blir implisitt evaluert gjennom konkurransen med andre aktører. Man overlever ikke i markedet dersom man ikke leverer god forskning som forstås og kan brukes. Vedkommende mente derfor at rangering og evaluering er viktigere for offentlige UH-institusjoner som lever av offentlige midler.

På samme måte ser man forskjeller mellom store, internasjonalt orienterte breddeuniversiteter og mindre, regionale profesjonsrettede høyskoler. En representant for et av de største universitetene i Norge understreker at det er viktig å reflektere over hvem man skal sammenligne seg med, og at breddeuniversiteter bør sammenlignes med hverandre, ikke med mindre høyskoler. En representant for en mindre høyskole sier også at det ville være positivt om evalueringene tok hensyn til at ulike institusjoner både har ulike omgivelser, forutsetninger og ambisjoner.

Til sist er det også store forskjeller mellom private og offentlige UH-institusjoner. Private høyskoler har, på samme måte som institutter, ofte en helt annen økonomisk situasjon enn offentlige UH-institusjoner. Som én representant for en privat høyskole sier, så tilbyr de tjenester som studenter kan få gratis andre steder. De er derfor helt avhengige av å kunne tilby høy kvalitet, og de blir dermed også svært avhengige av å ha en sterk kvalitetskultur og mange evalueringer av både forskning- og utdanningskvalitet for å vinne frem i markedet. Dette behovet har ført til at de har initiert mange evalueringer selv, internt på institusjonen, fordi evalueringene som gjennomføres av NFR og NOKUT ikke er detaljerte eller treffende nok.

Som nevnt i seksjon 5.1, har en overgang fra fagevalueringer til institusjonsevalueringer blitt fremmet som en løsning på denne utfordringen. I en institusjonsevaluering kan man i større grad ta utgangspunkt i institusjonens egne ambisjoner og formål. Det at institusjonene selv kan velge hvilke andre institusjoner og mål de ønsker å evalueres eller benchmarkes mot kan også være en løsning. Et annet alternativ, som flere informanter trekker frem, er at man kan utvikle et rammeverk for evalueringer som fungerer som en slags verktøykasse for institusjonene. Her kan man velge noen måter

å evaluere på som passer egne forutsetninger, rammebetingelser og ambisjoner – og så kan man sammenlignes med andre institusjoner som evalueres på tilsvarende måte. Slik kan man på den ene siden sørge for at alle jobber med kvalitetsutvikling og oppfyller noen grunnkrav, og på den annen side at det er muligheter for fleksibilitet ut ifra institusjonenes egne behov og formål. Slik kan man ivareta en grunnleggende kvalitetsstandard, men samtidig støtte opp under institusjonenes egenart. Konsekvensene av en slik modell ble kort diskutert i seksjon 5.1, under «Institusjonsevalueringer».

Et aspekt, som særlig fagforeningene trekker frem, er at man ikke bare må ta hensyn til de ulike institusjonenes forutsetninger i en evaluering, men også *evaluere* de ulike forutsetningene og rammebetingelsene. Her peker de særlig på de ansattes arbeidsbetingelser og kompetanse og hvilke muligheter de har for kompetanseutvikling og faglig påfyll. De ansattes arbeidshverdag og deres mulighet til å fokusere på kvalitet i eget arbeid er sentralt for kvaliteten i forskning og undervisning generelt.

5.4 Løpende indikatorer og foreliggende data

Vi har flere ganger omtalt et ønske om mer effektiv bruk av foreliggende data, og dette oppfattes som et behov av både KD og alle typer forskningsprodusenter. Det finnes stadig mer data, og forskningsprodusentene rapporterer allerede til en rekke ulike instanser. Mange ser gode muligheter for å sammenstille og utnytte disse dataene bedre. Dette kan forstås som et klart og tydelig behov fordi a) det ville ført til mer effektive evalueringer, og b) løpende indikatorer og evalueringer basert på indikatorer ville gi nyttig informasjon kontinuerlig og raskt gi informasjon om utvikling og trender.

Én hevder at det er behov for å se på evalueringers rolle i sammenheng med data, statistikk og annen informasjon som for eksempel vurderinger knyttet til forskningsprosjektsøknader. Det finnes en lang rekke evalueringspunkter, og dersom man hadde vært i stand til å samle og utlede informasjon fra denne helheten kunne man potensielt oppnådd store effektivitets- og innsiktsgevinster.

NFRs årsrapport for hvert instituttområde brukes som eksempel på en kilde til svært mye god informasjon. Flere aktører mener materialet disse er bygget på i stor grad er nok til å dekke deres behov for benchmarking, osv. Dersom man hadde klart å utnytte denne kilden bedre, hadde det åpnet for at de tradisjonelle evalueringene i større grad kunne vris mot et institusjonsfokus uten at sentrale evalueringsspørsmål forble ubesvart. Benchmarking kan bygge på løpende indikatorer, og da kan kanskje evalueringer være noe annet, spør én interessent.

Gevinstene forbundet med innsikt er altså ett poeng, og en annen grunn til at mange fokuserer på dette er at rapportering av nøkkeltall oppfattes som ressurskrevende, og det er ønskelig å få maksimalt ut av dette.

Dersom man skal utnytte dette potensialet, kreves det en koordinert satsning på å vurdere hvilke tall som finnes, hva man fremdeles ikke har tall på samt hvordan dette skal analyseres, sammenstilles og rapporteres. Mange av de større institusjonene har analysekapasitet nok til å få mye nyttig informasjon ut av rådataene i slike undersøkelser. En representant for en høyskole understreker imidlertid at de har et stort behov for å få slik informasjon ferdig analysert, eller for å øke sin egen kompetanse og kapasitet på feltet for at de skal kunne nyttiggjøre seg dataene slik de eksisterer i dag. Informanten etterspør for eksempel et årlig seminar der NOKUT gjennomgår resultatene fra årets Studiebarometer og hvordan disse dataene kan analyseres og brukes. Dette mener vedkommende kunne bidratt til kompetanseheving i sektoren. I tillegg påpeker flere at det er svært viktig med bedre kvalitetssikring av tallene dersom de skal brukes mer aktivt.

Den mest fremoverlente interessenten, fra UH-sektoren, foreslår halvveis på spøk å avskaffe de tradisjonelle evalueringene, samle inn forskningsdata, siteringsdata, o.l. og årlig komme med en oversikt: «This is the state of Norwegian research this year!» Dette forslaget kommer i en diskusjon om hva man ville tapt med en slik løsning. Interessenten ser at man opplagt hadde tapt noe, men ikke nødvendigvis svært mye.

5.5 Behov for koordinering og forutsigbarhet

Et av de behovene det har vært størst enighet om blant våre informanter er behovet for mer koordinering, samordning og forutsigbarhet på evalueringsfeltet. Dette omhandler både arbeidsdelingen mellom ulike aktører, hva som evalueres når og hvilken rolle evalueringer skal ha i samspill med andre former for vurderinger, målinger og løpende indikatorer.

Behov for koordinering og arbeidsdeling

Flere av aktørene vi har snakket med mener at det gjennomføres for mange ulike evalueringer, som også til dels er overlappende og fører til dobbeltarbeid i sektoren. Som en representant fra KD uttrykte det: «NFR har sine evalueringer, NOKUT har sine og institusjonene utvikler sine egne. Det er nødvendig å se det i en større sammenheng slik at man får en god og omforent arbeidsdeling». Dette signalet fra KD har også evaluatorene selv merket, og NOKUT forteller at de har oppfattet at KD ønsker at NOKUT og NFR skal jobbe mer sammen. NOKUT er også enig i at dette er nødvendig ettersom de også må se på fagmiljøene når de skal evaluere forskningsbasert utdanning. Dette er relatert til forskningen, og her er det nødvendig å samarbeide med NFR for å få til gode evalueringer.

Det generelle inntrykket vi har fått fra informantene er at hver enkelt evaluering og rapportering forståvidt er fornuftig, men at totaliteten kan oppleves som overveldende og uoversiktlig. De ulike evalueringene kan ha litt forskjellig fokus og dermed konkurrere litt mot hverandre. Dette oppstår særlig fordi det finnes mange organer med ulike styringsparametre. I tillegg til NFR, NOKUT og institusjonenes egne evalueringer nevner også flere aktører at de rapporterer mye til NIFU, og at de nye direktoratene DIKU, og i mindre grad UNIT, også krever noe rapportering. Dette omtales av flere aktører som en «direktorativering» av feltet.

Det virker særlig å være en del uklarhet rundt arbeidsfordelingen mellom NOKUT og DIKU, som begge har utdanningskvalitet som en del av sitt ansvarsområde. De færreste av de vi har intervjuet har en klar oppfatning av hvordan ansvars- og rollefordelingen er mellom disse to aktørene.

Mange mener at det er behov for koordinering på feltet, og det foreslås blant annet at ansvaret for evalueringer av forskning og utdanning bør flyttes opp til KD. Flere mener at KD burde ha et koordinatoransvar og deretter delegere evalueringsoppdrag og -mandater til de ulike direktoratene og organene som jobber med evaluering. Dette kan skape en bedre helhet i evalueringene, i stedet for at de ulike organene evaluerer med bakgrunn i hvert sitt mandat. Dette kan gi mulighet for å ha færre og bredere evalueringer der man koordinerer de ulike parameterne.

Andre mener derimot at NFR og NOKUT gjør en god jobb, og at KD ikke bør få muligheten til å detaljstyre evalueringsvirksomheten. Flere er også bekymret for at et nytt nasjonalt rammeverk ikke nødvendigvis vil føre til en forenkling og mer koordinering, men heller til et evalueringsbyråkrati. En representant for en fagforening mener at apparatet rundt evalueringer bør holdes minst mulig, slik at ressursene kan gå til institusjonene og de som faktisk gjør grovarbeidet i slike evalueringer. En informant fra

en høyskole understreker at man, hvis man skal lage et koordinert rammeverk, bør ta utgangspunkt i hvilke systemer som allerede er på plass og som fungerer godt. Det er viktig at et nytt rammeverk *forenkler* og skaper forutsigbarhet. Som mange har uttrykt det, bør et slikt rammeverk komme *i stedet for*, ikke i tillegg til det som allerede gjøres.

Behov for forutsigbarhet

Behovet for bedrekoordinering og klarere arbeidsdeling henger tett sammen med behovet for forutsigbarhet. Flere aktører etterspør et evalueringsprogram der man får informasjon om hvilke evalueringer som kommer og når. Et slikt evalueringsprogram kan gjøre det enklere for institusjonene å planlegge egne evalueringer og undersøkelser slik at de unngår dobbeltarbeid. Et slikt program bør også ta hensyn til eventuelle internasjonale undersøkelser eller evalueringer, for å kunne hindre kollisjoner og dra nytte av hva som kommer frem i slike prosesser. En representant fra et universitet etterspør også medvirkning fra institusjonene i designet av et slikt evalueringsprogram. Her bør sektoren involveres aktivt slik at de får et større eierskap over prosessene.

Flere aktører etterspør faste evalueringssykluser, slik at institusjonene vet både når og hvor ofte de vil bli evaluert. En representant fra en privat høyskole forteller at de innførte egne periodiske evalueringer nettopp for å få en slik forutsigbarhet. Deres erfaring med de store nasjonale evalueringene var at de ofte dukket opp én gang, og at de fikk en interessant rapport, men at den kun ble lest en gang og deretter havnet i en skuff. De opplevde en mangel på både oppfølging og videre evalueringer i etterkant.

Én aktør foreslår å gi institusjonene mer eierskap til prosessen ved å innføre noen grunnkrav (for eksempel at alle institusjoner eller fagområder skal evalueres hvert 10. år), og så la institusjonene selv planlegge og ta initiativ ut ifra hva som passer best med deres tidsplan og strategi. Dette kan gi institusjonene både forutsigbarhet og en viss fleksibilitet. Dette er en modell som blant annet brukes i Nederland (se *Internasjonal sammenligning* for mer informasjon).

Et annet aspekt som trekkes frem er forutsigbarhet rundt hva som måles og hva man, som både enkeltforsker og institusjon, får uttelling for. En representant for en fagforening understreker at hvis man endrer kriteriene og indikatorene man blir målt på, kan dette gi negative utslag for forskere og fagmiljøer som har investert i noe som ikke lenger gir uttelling. Dette er spesielt relevant i sammenheng med åpen forskning og hvordan dette kan påvirke for eksempel nivåinndelingen og hvordan man teller publikasjonspoeng. Her er det også viktig å tenke på hvordan man kan evaluere på en måte som gir forskerne forutsigbarhet i karrieren.

Evalueringers rolle i helheten

Til sist er det flere av informantene som mener det er viktig å tenke gjennom hva evalueringer er og skal være sammenlignet med annen rapportering, statistikk og vurdering av forskning- og utdanningskvalitet. Her bør man også inkludere NOKUTs tilsyns- og akkrediteringsvirksomhet, den kvalitetsvurderingen som finnes i søknadsskriving og vurdering av søknader om for eksempel forskningsfinansiering, samt den vurderingen som ligger i det resultatbaserte finansieringssystemet. Det å se alt dette i sammenheng er nødvendig både for å redusere det totale evaluerings- og rapporteringstrykket i sektoren, og for å få frem den informasjonen man har behov for ved bruk av best mulige virkemidler. Som nevnt over etterspør flere bedre bruk av løpende indikatorer. Dersom man klarer dette, er det viktig å tenke igjennom hva evalueringer skal være – hva kan man få ut av evalueringer som man *ikke* får ut av statistikk og løpende indikatorer? En representant for et universitet forteller at han synes evalueringene treffer, men at det er mulig de kunne kommet frem til samme resultater

på en mye enklere måte, for eksempel ved bruk av eksisterende statistikk og datamateriale. Han understreker imidlertid at man da ville mistet mye av eierskapet institusjonene føler for evalueringene. I tillegg ville det begrenset læringen som følger av å gå gjennom en evalueringsprosess der man må skrive egevalueringer, impacthistorier og beskrivelser av egne resultater. Det å beskrive hvilke publikasjoner en selv anser som sentrale og viktige, og *hvorfor*, bidrar til læring og eierskap, sammenlignet med bare å telle opp publikasjoner og siteringer.

Flere aktører understreker betydningen av å ha en klar intensjon med utformingen av et nytt nasjonalt rammeverk for evalueringer – og deretter sørge for at resultatet følger denne intensjonen. Tilnærmet alle er enige om at to viktige aspekter ved et slikt rammeverk må være å forenkle evalueringssystemet og sørge for at evalueringer gir nyttig og relevant informasjon til mottakerne. Derfor er det viktig å først tenke på hvordan man ønsker at helheten skal se ut, og deretter se på hvilke evalueringer det er behov for. Som en av aktørene uttaler: «evalueringene må tilpasses systemet, systemet kan ikke tilpasses evalueringene». Dette gjelder også evalueringer av kvalitet. Her er det viktig å først avklare hvordan man forstår kvalitet, og deretter sette opp indikatorer og evalueringer for å måle dette – ikke omvendt.

Det er viktig å vurdere hva som er merverdien av evalueringer, særlig i en tid der rapporteringstrykket vokser og aktørene får bedre tilgang på både datamateriale og analyseverktøy. Dette har vi diskutert i seksjon 5.4, og vi vil også komme litt mer inn på det i neste kapittel om utviklingstrekk.

Oppsummering

Sentrale motsetninger og aktører med spesielle interesser

Forskjellige forutsetninger og bruk av evalueringer	Små og store UH-institusjoner
Nasjonale vs. internasjonale evalueringer	Små og store forskningsprodusenter
Formative vs. summative evalueringer	Forskningsprodusenter
Institusjonsfokus vs. nasjonalt fokus/fagevaluering	KD, NFR, forskningsprodusenter

Hovedpoenger

- Fagevalueringer anses som nyttige og etterspørres av alle interessentgrupper
- De fleste interessentgruppene mener at sammenligning og benchmarking er viktig
- Forskjellige oppfatninger om det bør sammenlignes nasjonalt eller internasjonalt

6 Utviklingstrekk

Avslutningsvis vil vi i dette kapitlet presentere en rekke temaer vi omtaler som *sentrale utviklingstrekk*. Dette er temaer som har kommet opp i intervjuene, både i samtale om evalueringer generelt, men også på direkte spørsmål om hva interessentene oppfatter som viktige utviklingstrekk som kan påvirke evalueringer av kvalitet i forskning og utdanning. Dette omfatter utviklingstrekk som kan ha konsekvenser for hvordan evalueringer bør gjennomføres, og omhandler ikke hvordan evalueringer i seg selv har utviklet seg over tid.

6.1 Åpen forskning

Det første temaet vi ønsket å undersøke er utviklingen mot mer åpen forskning. Dette innebærer både økt tilgjengelighet av forskningsartikler/-produkter, men også mer åpenhet rundt forskningsprosesser og datamateriale. Vi fokuserer i denne sammenhengen på generelle temaer knyttet til dette, og strukturerer ikke diskusjonen rundt konkrete planer og initiativer som DORA, Plan S, osv. Dette som en følge av at interessentene i stor grad omtaler dette som generelle fenomener.

Grunnen til at dette temaet diskuteres er at krav til åpen forskning kan medføre endringer i hvordan forskningskvalitet vurderes, samt hvordan enkeltforskere vurderes (og for eksempel ansettes). Mange av interessentene mener det er et *viktig* utviklingstrekk, men de aller fleste ser allikevel ut til å være av den oppfatning at denne utviklingen ikke nødvendigvis fører til store endringer i hvordan evalueringer kan eller bør gjennomføres.

Et vanlig standpunkt i denne sammenheng er at den norske tilnærmingen med avtaler om åpen publisering med de store forlagene gjør at situasjonen som helhet ikke endres nevneverdig. Norske forskere, hevdes det, vil fremdeles i hovedsak kunne velge å publisere i journaler som tradisjonelt sett oppfattes som prestisjetunge, og eksisterende nivåsystem og vurderingskriterier kan dermed i stor grad opprettholdes.

6.2 Tverrfaglighet

Et annet punkt vi velger å omtale som et utviklingstrekk er økende behov for og betydning av tverrfaglighet. En hovedinnvending mot eksisterende fagevalueringer har vært at de har vært *disiplinbaserte* og at de i for liten grad har premiert god *tverrvitenskapelig* forskning. Grunnen til at vi nevner dette her er at flere av interessentene mener at tverrvitenskapelig forskning vil bli stadig viktigere, og at dette må føre til endringer i hvordan forskning evalueres. Samtidig knyttes det til en oppfattet svakhet ved tidligere fagevalueringer: Innretningen har vært mot disiplin, panelene består av representanter for disipliner, og anbefalingene er da ofte at miljøene bør rette seg mer mot disiplinene. Dette er spesielt problematisk for alle som oppfatter seg som tverrfaglige, og noen opplever at de blir anbefalt å opprette «disiplinære grupper for å løse disiplinære problemer, i stedet for å løse samfunnets problemer og sette sammen forskergrupper».

Argumentene som brukes er at man i større grad bør se på anvendelse av forskning, og samfunnsnytte, og at den type forskning som i størst grad bidrar (direkte) til dette ofte er sterkt avhengig av tverrfaglighet. Økt fokus på bærekraftsmålene fremheves som en årsak til økt behov for tverrfaglighet. Samtidig hevder flere at et mer komplisert, digitalisert og teknologibasert samfunn generelt medfører større behov for tverrvitenskapelig forskning.

Dette synet har imidlertid også blitt tydelig kritisert av flere interessenter. For det første pekes det på behovet for grunnforskning *og* anvendt forskning, av disiplinbasert

forskning og tverrfaglig forskning. Begge deler er nødvendig, og flere oppfordrer til forsiktighet dersom evalueringssystemer skal dreies mot et større fokus på impact og tverrfaglighet. For det andre utfordres selve premisset om at det er mer tverrfaglighet nå enn tidligere. Anvendt forskning har *alltid*, hevder flere, vært tverrfaglig. Dette er med andre ord kanskje ikke et helt nytt fenomen, og det er viktig å vurdere nøye om økt tverrfaglighet er et faktisk eller tilsynelatende utviklingstrekk.

6.3 Impact og samfunnsnytte

Et utviklingstrekk som omtales av mange interessenter er økt vektlegging av *relevans* og *samfunnsnytte* innen både forskning og utdanning. Dette omtales til dels i forbindelse med evalueringer, men det omtales også av mange interessenter som *ikke* kjenner til «impact»-begrepet og økt vektlegging av dette i evalueringer.

Impact har blitt en stadig større del av evalueringer (SAMEVAL nevnes her av flere som en foregangsevaluering), og denne utviklingen støttes av nær sagt alle interessentgruppene. Impact, bredt forstått som samfunnsnytte og effekter av forskning, er et tema alle er opptatt av, men få er enige om *hvordan* man skal kunne måle dette. Hvordan rapporterer man på impact, hvordan defineres det, utover at det skrives såkalte «impact cases»?

Mange etterlyser her tydelighet, og en større debatt om hva begrepet er, hvordan det bør forstås og hvordan man kan tilnærme seg impact-målinger metodisk. En av interessentene bruker SAMEVAL som eksempel, og undrer seg over en overraskende god vurdering av egen institusjon på dette området. Dette skyldes, mener interessenten, at akademisk og samfunnsmessig impact til dels ble sammenblandet i den aktuelle evalueringen. Det etterspørres derfor en tydeligere tilnærming til dette i kommende JUREVAL.

Det er tydelig at mange fagmiljøer opplever en viss usikkerhet rundt vektleggingen av impact, men flere forteller om en innledende motstand, som gradvis gikk over til en erkjennelse av at det er svært nyttig å forholde seg aktivt til hvordan egen forskning anvendes, både på kort og lang sikt. Én aktør beskriver arbeidet med impact som en «ekstremt viktig øvelse», som førte til at fagmiljøene innhentet informasjon om og forstod konsekvensene av egen forskning på en helt ny måte.

Departementene og arbeidslivsorganisasjoner skiller seg ut som spesielt opptatt av økt oppmerksomhet på impact i evalueringer. For at dette skal bli nyttig for flere bør det imidlertid tydeliggjøres hva impact er, og tydelig formidles hva man kommer frem til i evalueringene – også for et publikum utover kjerneaktørene i forskningsverdenen.

6.4 «Fake news», utdanning og tillit til forskning

Vi omtalte tidligere hvilke effekter evalueringer har på befolkningens tillit til forskning (se 4.2). Flere interessenter har på eget initiativ understreket hvor viktig forskning er i dagens samfunn, og at forskningens rolle blir *enda* viktigere i en verden med økt usikkerhet rundt hva som er *sant* og hva som ikke er forskningsbasert kunnskap, herunder «fake news».² Utviklingstrekket som her diskuteres knyttes i stor grad til større fenomener som økt

² Kalsnes, B. (2020). Falske nyheter: Løgn, desinformasjon og propaganda i den digitale offentligheten. Oslo. Cappelen Damm akademisk.

polarisering og effektene av økt bruk av algoritmer og kunstig intelligens i sosiale nettverk.³

Utdanning og forskning fremstår begge som sentrale bolverk mot en uønsket utvikling i retning av «mine fakta og dine fakta», og dette er grunnen til at flere interessenter tar opp temaet. Evalueringer, sier de, kan ideelt sett bidra til å styrke tilliten til forskning, og forklare hva som er *god* forskning, og helst også hva som kjennetegner god og legitim forskning. Et slikt fokus kan imidlertid innebære en utvidelse av NFRs mandat, og interessentene er ikke låst til at det er NFR som skal ha ansvar for denne type folkeopplysning. *Noen* bør imidlertid gjøre det, og evalueringer fremstår for flere som et naturlig grunnlag for å vurdere og informere om hvilke miljøer, og hva slags forskning, som fortjener tillitt og oppmerksomhet. En interessent svarer på spørsmål om hva de trenger av evalueringer: «At vi vet om det – at det blir tilgjengeliggjort». Dette gjenspeiler et ønske om noe mer aktiv formidling av viktige funn, resultater og evalueringer fra NFRs side.

Samtidig er tillitt til utdanning viktig, til dels for å skape større legitimitet til den forskningen som er et resultat av høyere utdanning og forskerutdanning. *Kunnskap* og forståelse og respekt for hva dette er og hvordan det fremskaffes, bringer utdanning og forskning sammen. Samtidig påpekes det at det å *ta* høyere utdanning også bør innebære at man blir rustet til å vurdere hvilken informasjon man kan stole på og ikke. I denne forbindelse kan det også være naturlig å trekke inn tematikken rundt *kombinasjonen* av evaluering av forskning og utdanning.

6.5 «Learning analytics» og økende tilgang til data

Tett knyttet til ønsket om bedre og mer effektiv bruk av løpende indikatorer og eksisterende data, er utviklingen rundt *stordata* mer generelt. I alle deler av samfunnet registreres, lagres og analyseres mer data enn noen gang tidligere, og UH-sektoren og forskningsverdenen generelt er intet unntak. Institusjonenes og organisasjonenes behov for og ønske om mer effektiv bruk av data er beskrevet i seksjon 5.4.

Et sentralt *utviklingstrekk* er at man innenfor UH-sektoren i større grad samler inn det som kalles *læringsdata* (learning analytics). Ved bruk av læringsmiljøsystemer (LMSer) går svært mye kontakt mellom studenter og institusjoner gjennom digitale systemer, og også eksamener og forelesninger foregår i stadig større grad på digitale plattformer. Dette åpner en ny verden av muligheter for datainnsamling og analyse, og institusjonene har i varierende grad begynt å benytte seg av mulighetene som ligger i denne type systemer. Det er imidlertid også flere aspekter knyttet til både personvern og *åpenhet* når det gjelder denne type løsninger, siden mange institusjoner bruker proprietære løsninger som innebærer at læringsdata eies og håndteres av private internasjonale selskaper.

I denne forbindelse kan man med evalueringer både se på potensialet for å benytte seg av mulighetene som ligger i dette datamaterialet, men man kan også se for seg at institusjonenes håndtering av personvern og etiske problemstillinger evalueres og blir en del av tilsynsvirksomheten. Vi har funnet en klar tendens til at store institusjoner i større grad har avdelinger og ansatte med nødvendig analysekompetanse, og mindre institusjoner uttrykker et ønske om hjelp og støtte til å enten analysere datamaterialet selv, eller få bistand til analyse og eksterne gjennomganger av datamaterialet.

³ Sætra, H. S. (2019). The tyranny of perceived opinion: Freedom and information in the era of big data. *Technology in Society*, 59, 101155.

6.6 Øvrige utviklingstrekk

Avslutningsvis vil vi kort peke på et utvalg øvrige utviklingstrekk som har blitt fremhevet i løpet av interessentanalysen.

Noen trekker frem at økt fokus på *livslang læring* medfører behov for endringer i sektoren og evaluering og kvalitetsarbeid rettet mot UH-sektoren. Livslang læring kan for eksempel innebære at flere ønsker å ta mindre kurs, og tilleggsutdanninger, og at utviklingen mot sterkere fokus på fullstendige *grader* o.l. kan komme i veien for dette. Samtidig medfører slike løsere og mer fleksible utdanningsmodeller behov for å innrette evalueringer på en ny måte. Hvordan skal man evaluere og sikre kvalitet her, dersom «follow the student» blir prinsippet, fremfor at studentene skal gå inn i planlagte og ferdige løp?

Andre igjen trekker frem *internasjonalisering*, og sikter da i hovedsak til et økt fokus på internasjonale samarbeid, for eksempel universitetskonstellasjoner. Samtidig er flere opptatt av at man i evaluering av utdanning i større grad ser til internasjonale initiativer og kobler seg på eller initierer samarbeid med for eksempel EU. Også OECD nevnes i denne sammenheng, og det fremstår generelt som ønskelig med mer internasjonalt samarbeid på evalueringsfeltet, spesielt siden en del institusjoner ønsker å utvide og styrke sine internasjonale forbindelser. Dette vil være lettere dersom de forskjellige institusjonene ikke evalueres på svært forskjellige måter.

Vi har også nevnt fokuset på *arbeidslivsrelevans* i høyere utdanning, og dette temaet tas også opp av flere interessenter. Spesielt aktører *utenfor* UH-sektoren vektlegger dette, mens flere andre aktører frykter at man går for langt i å la markedet og tanker om relevans styre UH-sektoren.

7 Oppsummering

I denne rapporten har vi presentert funnene fra intervjuer med 53 interessenter med tilknytning til forskning og høyere utdanning. Vi har sett at forskjellige interessentgrupper har forskjellige oppfatninger av evalueringer som tidligere har blitt gjennomført, og at de har forskjellige behov for evalueringer.

Kunnskapsdepartementet (KD) fremstår som en aktør med spesielt sterke interesser på feltet, og med spesielle behov for informasjon og kunnskap om status på feltet som helhet. Dette er naturlig og følger av departementets overordnede ansvar for politikktutvikling på og kunnskap om feltet. Nasjonale evalueringer, som NFRs fagevalueringer, fremstår som en meget viktig kunnskapskilde for KD, og dersom NFR går bort fra denne type evalueringer ser det ut til å være behov for en alternativ løsning som dekker behovet for oversikt over fagfelt på et nasjonalt nivå.

Øvrige departementer har i større grad behov for kunnskap om forskningen i egen sektor, og fokuserer i hovedsak på programmene de selv er med på å finansiere. Evalueringen av programmer og virkemidler er her sentrale. Departementene har også en relativt klar arbeidsdeling, der det er KD som har ansvar for høyere utdanning og forskning *generelt*, mens øvrige departementer i langt større grad fokuserer på anvendelse av forskning, innovasjon og samfunnsnytte. Direktoratene har også en viss forskningsaktivitet, men direktoratene og øvrige departementer er kun i begrenset grad opptatt av de store nasjonale evalueringene av kvalitet i forskning og utdanning i sitt daglige virke.

UH-sektoren, institutter, helseforetak og øvrige forskningsprodusenter er de som i hovedsak *blir* evaluert. Vi får inntrykk av at mange opplever de nasjonale evalueringene som disiplinorienterte, noe som medfører at de passer best til evalueringer av UH-institusjoner, og at noen institutter derfor opplever evalueringene som mindre relevante. Til dels medfører dette at de ikke deltar i alle evalueringer, og til dels deltar de og opplever at de evalueres på noe urimelige premisser, ved at for eksempel anvendt forskning og sterkt tverrfaglig forskning lider av et for sterkt disiplinfokus. Det er imidlertid også store forskjeller mellom UH-institusjoner, og også mellom institutter, så bildet fremstår som meget sammensatt. Det er derfor nødvendig å gå inn i disse nyansene for å få et fullstendig bilde av hvordan evalueringene treffer og oppfattes av disse aktørene. Her er institusjonsstørrelse, privat/offentlig, type organisasjon, fagområde, mm. sentrale dimensjoner som henger sammen med til dels meget forskjellige oppfatninger.

Interesseorganisasjoner og arbeidslivsorganisasjoner har varierende grad av interesse for store nasjonale evalueringer. Studentorganisasjoner og fagforeninger har en sterk interesse, og begge disse interessentgruppene påpeker at det er viktig at fagansattperspektivet blir en større del av evalueringer av kvalitet i forskning og utdanning. Øvrige arbeidslivsorganisasjoner som representerer arbeidsgiversiden ligner til dels på øvrige departementer (unntatt KD), ved at de har et sterkt fokus på samfunnsnytte, relevans, innovasjon og nyskaping.

Et sentralt tema i interessentanalysen er fremtidig innretning av store nasjonale forskningsevalueringer. Hvilken innretning som er mest hensiktsmessig: en modell basert på eksisterende fagevalueringer eller mer institusjonsorienterte og formative evalueringer? Vi har vist at det er gode argumenter som føres for begge alternativer, og interessentene har delte meninger om hvilken tilnærming som vil være best. Et helt sentralt poeng for mange er at regelmessige evalueringer av fagområder på nasjonalt nivå har store og positive effekter. De oppfattes som helt nødvendige for å utforme god forskningspolitikk, samtidig som de gir muligheter for benchmarking, noe som oppleves som viktig for mange. Andre igjen er åpne for at en institusjonsrettet evaluering vil gi mer nyttig informasjon til evaluanden, og at dette kan ha positive konsekvenser.

Generelt ser vi også at mange aktører mener at det finnes et potensial for å utnytte eksisterende data på en bedre måte, og utforme et system for *løpende evaluering* basert på kvantitative indikatorer, hvorpå de større evalueringene i større grad kan være kvalitative og mer målrettede. Det er generelt et sterkt ønske om mer koordinering av evalueringsvirksomheten knyttet til forskning og høyere utdanning. Samspeillet mellom indikatorer og evalueringer er viktig i denne sammenheng. Andre igjen peker på et behov for koordinering på et høyere nivå enn NFR og NOKUT, slik at evalueringer i større grad kan spille sammen, og ta hensyn til hverandre.

Utdanningsevalueringer er et tema som relativt få av interessentene har et sterkt forhold til. UH-institusjonene er imidlertid meget opptatt av NOKUTs virke. De forteller om få rene evalueringer, men et omfattende kvalitetssikringsarbeid og tett samarbeid med NOKUT. Det er også mange som omtaler spenningen mellom NOKUT som støttespiller og NOKUT som kontrollør, men de fleste mener at NOKUT har funnet en god balanse. Det er også en tydelig oppfatning at NOKUT over tid har utviklet seg i retning av å bli mer støttespiller og mindre kontrollør, noe som oppfattes som positivt. I tillegg opplever flere interessenter at NOKUTs fagevalueringer har vært viktige og hatt sterke virkninger. Det har imidlertid blitt gjennomført relativt få slike, og en rekke interessenter er positive til at det initieres flere slike evalueringer også på utdanningsiden.

Hovedtrekk ved interessentgruppene

Kunnskapsdepartementet	forskningspolitikk, utdannings- og forskningssektoren mer generelt
Evaluatorene	oppleves som genuint opptatt av å bidra til å fremskaffe kunnskap og bistå institusjoner (vanskelig å dekke begge deler like godt?)
UH-institusjoner	Preget av store interne forskjeller
Store universiteter	Mye intern kompetanse, komplekse og svært sammensatte organisasjoner som ikke har ett behov
Store høyskoler	Ambisjoner og benchmarking mot universiteter
Mindre høyskoler	Fokus på samfunnsoppdrag og nisje, behov for å evalueres på dette premisset
Private institusjoner	Konkurranserettete, sterkt kvalitetsfokus, mer egeninitiert evaluering (oppleves som mer direkte relevant og nyttig)
Institutter	Stor variasjon, generelt mer fokus på anvendt forskning og relevans - behov for å evalueres på dette premisset
Øvrige dep. og arbeidsgiverorg	Fokus på relevans, innovasjon, anvendelse
Direktorater	Fokus på egne felt, ønske om mer spissede evalueringer
Studenter og forskere	Meget opptatt av kvalitet, og opptatt av at deres perspektiv i større grad ivaretas

Vedlegg 1: Internasjonal sammenligning

Notat: Prosjekt *Kartlegging av evalueringer*

Innhold:

1	INNLEDNING	1
2	INNRETNING AV EVALUERING AV FORSKNING	2
2.1	Storbritannia.....	2
2.2	Sverige.....	4
2.3	Finland	6
2.4	Nederland	7
2.5	Sammenligning og sentrale dimensjoner i vurderingen av forskningsevalueringer	8
3	KVALITET I UTDANNING	14
3.1	Storbritannia.....	14
3.2	Sverige.....	15
3.3	Finland	16
3.4	Nederland	18
3.5	Sammenligning og sentrale dimensjoner i vurderingen av utdanningsevalueringer.....	19
4	REFERANSER	27

1 Innledning

Evaluering av forskning og utdanning gjennomføres på en rekke forskjellige måter, og i dette notatet viser vi hvordan kvalitet i forskning og utdanning evalueres i *Nederland, Sverige, Finland* og *Storbritannia*.

I tillegg til å presentere og redegjøre for den grunnleggende innretningen av evaluering i de fire landene avslutter vi hver hoveddel med en beskrivelse av det som fremstår som de sentrale dimensjonene for å vurdere og sammenligne forskjellige evalueringssystemer. Dette er til dels basert på vår egen gjennomgang og til dels på eksisterende forskning på forsknings- og utdanningsevaluering.

Vi svarer her på følgende problemstilling fra konkurransegrunnlaget:

Hvordan er nasjonale evalueringer av kvalitet i forskning og/eller høyere utdanning organisert og innrettet i utvalgte andre land, f.eks. Nederland, Storbritannia, Sverige og ev. ett til to land til som oppdragstaker vurderer som særlig relevante? I denne sammenhengen er vi primært interessert i evalueringene som kvalitetsvirkemiddel, og ikke så mye i ev. Kobling til finansieringssystemer.

2 Innretning av evaluering av forskning

I dette kapittelet gir vi en kort oversikt over innretningen av evalueringer av forskning i Storbritannia, Sverige, Finland og Nederland. Alle disse landene har systemer for evaluering av kvalitet i forskning, men det er store variasjoner mellom landene når det gjelder organisering av evalueringene, hvilke enheter og nivåer som blir evaluert, hvem som utfører evalueringene og i hvilken grad evalueringene kobles til finansieringssystemet. I dette kapittelet beskriver vi hvordan kvalitet i forskningen evalueres i de ulike landene. Vi vil konsentrere oss om evalueringer knyttet til UH-sektoren og inkluderer ikke instituttsektoren ettersom denne i mindre grad er sammenlignbar på tvers av landene.

2.1 Storbritannia

I Storbritannia er forskningsrådene fordelt etter fagområder, men ligger alle under *United Kingdom Research and Innovation* (UKRI). I tillegg til fagspesifikke forskningsråd, er det også etablert egne nasjonale finansieringsorganer med ansvar for henholdsvis England, Skottland, Wales og Nord-Irland.¹

Hva finnes av evalueringer?

De nasjonale finansieringsorganene har sammen etablert Research Excellence Framework (REF), som er Storbritannias system for å evaluere forskningskvalitet blant høyere utdanningsinstitusjoner. Ordningen blir administrert av Research England, en organisasjon under UKRI som jobber med den delen av UKRIs arbeid som er rettet mot England.

REF ble første gang gjennomført i 2014, men ettersom den overtok etter den tidligere ordningen Research Assessment Exercise (RAE), er det den syvende evalueringen av sitt slag i Storbritannia. REF 2014 omfattet forskningsproduksjonen ved 154 universiteter i perioden 2008-2013. Den neste evalueringen er planlagt gjennomført innen 2021, men våren 2020 ble det besluttet at prosessen skulle utsettes til det ble klart hvor store konsekvenser COVID-19 hadde fått for de britiske universitetene. Ettersom REF 2021 ikke er gjennomført ved dette notatets ferdigstilling, beskriver vi her innretning og gjennomføringen av REF 2014.²

Formålet med evalueringen

Storbritannia var ett av de første landene som introduserte indikatorer for produksjon blant høyere utdanningsinstitusjoner, og hadde allerede på tidlig 90-tall innført resultatbasert finansiering av forskning.

Formålet med forskningsevalueringer i Storbritannia er å gjennomføre en revisjonslignende gjennomgang av hva Storbritannia får ut av finansieringen av forskning (rundt NOK 20 milliarder i 2014). Forskningsmidlene tilsvarer 30 prosent av all finansiering i UH-sektoren. REF skal dermed gi et informasjonsgrunnlag for allokering av offentlig forskningsfinansiering, og kan potensielt få store økonomiske konsekvenser for institusjonene.

¹ Geuna & Martin (2003) gir innføring i historien til det britiske evalueringssystemet, og beskriver utviklingen frem til 2003.

² Beskrivelsen av REF 2014 er basert på informasjon fra evalueringens hjemmeside, <https://www.ref.ac.uk/2014/>. For en kortfattet oversikt over evalueringen, se brosjyren «REF2014 Key Facts», <https://www.ref.ac.uk/2014/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf>.

Innretning og gjennomføring

I REF blir høyere utdanningsinstitusjoner evaluert på 36 fagområder (kalt «units of assessment») (for eksempel juss, økonomi og økonometri, klinisk medisin).³ Innen hvert av disse fagområdene er det etablert ekspertpaneler som består av både nasjonale og internasjonale akademikere, samt personer som bruker forskning i sitt eget arbeid. Hvert universitet som ønsker å delta i evalueringen bestemmer selv hvilke enheter de ønsker å sende inn til evaluering. Enhetene ble i 2014 definert innenfor ett av de 36 fagområdene.

Antall fagområder er redusert over tid. I 2001 var det hele 68 units of assesment, og dette var ca. likt som i 1992 og 1996 (Geuna & Martin, 2003).⁴ I REF 2021 er det planlagt å bruke 34 fagområder.⁵

Fagområdene som blir sendt inn til evaluering blir vurdert etter resultater («output») (65 prosent), impact (20 prosent) og forskningsmiljø (15 prosent). Resultater inkluderer for det meste ulike typer vitenskapelige publikasjoner, men kan også være andre former for resultater, som design, utstillinger, produkter, forestillinger og lignende. Resultater blir evaluert ved at institusjonene sender inn de fire beste publikasjonene/resultatene for hver ansatt de har valgt å inkludere i evalueringen. Resultatene skal være fra evalueringsperioden (som i REF 2014 var 2008-13).

I 2014 introduserte man for første gang effekt («impact») som evalueringskriterium. Impact defineres i REF 2014 som «any effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia». For å bevise impact av egen forskning kan institusjonene sende inn enten impact case studies eller impact templates. En impact case study er et dokument på fire sider som beskriver hvordan institusjonens forskning har hatt en impact i evalueringsperioden. En impact template er et dokument som beskriver hvordan det aktuelle fagområdet har lagt til rette for impact og dets fremtidige impact-strategi.

I REF 2014 ble forskningsmiljø definert som strategien, ressursene og infrastrukturen som støttet opp om forskningen i de ulike fagområdene. Her sendte hvert fagområde inn en environment template, som er en beskrivelse av fagområdets forskningsstrategi, inntekter, infrastruktur og fasiliteter, støttefunksjoner for fagansatte og studenter, samarbeid med andre samt fagområdets bidrag til disiplinen som helhet. I tillegg ble det levert inn statistikk over finansiering fra ulike kilder og antall doktorgrader som ble godkjent i evalueringsperioden.

Disse tre faktorene ble deretter vurdert av det aktuelle ekspertpanelet på en skala fra verdensledende (4), internasjonalt utmerket («excellent») (3), internasjonalt anerkjent (2) og nasjonalt anerkjent (1). Deretter ble de tre områdene vektet og slått sammen til en overordnet karakter for hele fagområdet som helhet. Hver institusjon fikk dermed en karakter fra 1-4 på de ulike fagområdene der de hadde valgt å sende inn materiale til vurdering. I overgangen fra RAE ble det vurdert å gå fra indikatorer og fagfellevurdering til ren bruk av indikatorer, men dette ble stoppet etter massive protester (Simon & Knie, 2013).

Fremtidige evalueringer

REF 2014 ble kritisert av flere grunner, særlig på grunn av de høye kostnadene. Totalt 1100 forskere i Storbritannia ble betalt for å vurdere totalt 191 232 forskningsartikler som ble levert inn for vurdering. Tidskostnaden, og alternativkostnaden vurderes som høy ettersom forskere får liten tid til egen forskning. Det anslås at den totale kostnaden for

³ <https://www.ref.ac.uk/2014/panels/unitsofassessment/>

⁴ Dette gjaldt da tidligere rammeverk RAE.

⁵ <https://www.ref.ac.uk/panels/units-of-assessment/>

REF 2014 var på rundt £250 millioner eller om lag 1,2 prosent av årlige forskningsbevilgninger.⁶

Likevel er det bestemt at den neste runden med evaluering, REF 2021, i stor grad skal følge REF 2014. Arbeidsmetodene for vurdering av forskning er lagt opp til å være relativt like i REF 2021 som i REF 2014, og beskrives som følgende:

*The assessment process is based on expert review. [...] The process is objective, and evidence-based but is not and cannot be purely algorithmic.*⁷

En endring fra REF 2014 er at institusjonene skal ha mindre rom til å selv velge ut hvilke ansatte de vil inkludere i evalueringen. I REF 2021 er det dermed bestemt at alle aktive forskere med et betydelig bidrag til forskningen skal inkluderes, ikke bare et utvalg (Sivertsen 2017).

2.2 Sverige

I Sverige blir mesteparten av forskning finansiert av det svenske forskningsrådet, Vetenskapsrådet, som også opptretr som rådgivingsinstans i forskningspolitiske spørsmål. Vetenskapsrådet har som en del av sitt mandat å evaluere forskning, inkludert kvalitet i forskningen og forskningens bidrag til samfunnet. I mandatet står det spesifisert at det skal vurderes hvorvidt myndighetenes investeringer resulterer i høyere kvalitet i forskningen.

Formålet med evalueringer

Gjennom flere typer analyser, undersøkelser og evalueringer ser Vetenskapsrådet på utviklingen i svensk forskning og hvordan det svenske forskningssystemet fungerer på et overordnet nivå. De analyserer for eksempel forskerkarrieren, kjønnsbalanse, hvordan forskningsmidler blir fordelt og hvordan initiativer fra myndighetene påvirker forskningen og forskningssystemet. I tillegg evaluerer de kvalitet i forskningen og forskningens betydning for samfunnsutviklingen. Alle disse undersøkelsene og analysene danner grunnlaget for anbefalingene Vetenskapsrådet kommer med i forskningspolitiske spørsmål.⁸

Hva finnes av evalueringer?

I Vetenskapsrådet finnes det tre emneråd, ett råd og fem komiteer som dekker ulike fagområder, for eksempel emnerådet for humaniora og samfunnsvitenskap og komiteen for kliniske studier. Disse jobber med fordeling av forskningsmidler og oppfølging av forskningen innenfor sine respektive områder. Oppfølgingen innebærer blant annet at kvaliteten på forskningen vurderes. Både emneråd og komiteer består av svenske forskere som sitter i rådet i en periode på 2 år, og som møtes mellom 4 og 6 ganger i året.⁹ Emnerådene sitter med mandatperiode på tre år, og skal ved slutten av mandatperioden publisere en rapport som blant annet beskriver forskning på det bestemte emnet de siste årene, men også følge opp tidligere publiserte rapporter.

⁶<https://blogs.lse.ac.uk/impactofsocialsciences/2015/08/03/why-did-the-2014-ref-cost-three-times-as-much-as-the-2008-rae-hint-its-not-just-because-of-impact/>

⁷ https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf

⁸ Se oversikt på Vetenskapsrådets hjemmesider, <https://www.vr.se/english/analysis/we-analyse-and-evaluate.html>.

⁹ Informasjon om emneråd, råd og komiteer i Vetenskapsrådet, se <https://www.vr.se/om-vetenskapsradet/organisation/amnesrad-rad-och-komiteer.html>

Ett eksempel på en slik rapport er Research overview 2019: Humanities and social sciences,¹⁰ som følger opp forrige rapport fra 2015. Her blir finansieringssøknader analysert, i tillegg til strukturelle analyser av de relevante institusjonene innen humaniora og samfunnsvitenskap i Sverige. Ett av funnene i rapporten er at forskningen i liten grad tilfredsstillter forventningen om samfunnsmessig påvirkning og relevans, og at det er liten grad av samspill mellom forskning og utdanning innenfor samme fagfelt. Det vil si at utdanningen ikke kobles til forskningsfronten.

For det andre blir forskningen i Sverige oppsummert i et forskningsbarometer annethvert år. Barometeret er basert på 30 forskjellige indikatorer. Dette skal gi en oversikt over forskningen i Sverige, i tillegg til hvordan den står seg i sammenligning med andre land.¹¹ Rapporten gir i stor grad et overordnet bilde av situasjonen i Sverige eller på enkelte universiteter, og gir i liten grad et innblikk i de forskjellige fagområdene. I den siste rapporten fra 2019 ble kun noen få sider viet til forskjellige fagfelt, og kun på et overordnet nivå. For eksempel ble siteringsdata brukt til å måle svensk forsknings betydning/gjennomslag («citation impact») innen de ulike fagområdene sammenlignet med gjennomsnittet i Europa og verden.

For det tredje sammenfatter Vetenskapsrådet data om forskning i Sverige og gir anbefalinger til myndighetene. Datamaterialet består av

- Undersøkelse om forskning, som understreker styrker og svakheter i fagfelt, forskningens påvirkning, utfordringer og muligheter for utvikling i fremtiden
- Forskningsinfrastruktur, bestående av informasjon om forskeres behov for laboratorier, databaser, programmer o.l.
- Analyse av sammensetningen av ansatte på svenske universiteter, blant annet kjønnsbalanse, karrieremuligheter for stipendiater, mobilitet o.l.
- Statistikk over hvordan tildelinger fra Vetenskapsrådet har blitt allokert mellom kvinner og menn, og hvorvidt tildelinger har vært konsentrert på noen fagfelt ved noen universitet, eller om det har vært en større spredning.

Fremtidige evalueringer

I 2020 gjennomfører Vetenskapsrådet et pilotprosjekt der de evaluerer statsvitenskapelig forskning i Sverige. Evalueringsmodellen fokuserer på forskningsresultater og består i at et internasjonalt ekspertpanel vurderer universitetenes forskning (i form av publikasjoner og bibliometri) og impact (i form av impact cases, som er beskrivelser av effektene av forskning) innen statsvitenskap. Denne modellen skal sørge for minimal innsats fra institusjonene. Dersom pilotprosjektet blir vellykket, forventer Vetenskapsrådet å gjennomføre en-to slike evalueringer årlig. Disse resultat-fokuserte evalueringene skal komplementere evalueringene som foretas av Universitetskanslersämbetet (UKÄ), som evaluerer institusjonenes egne kvalitetssikringssystemer og fokuserer mer på *prosess* (se seksjon 3.2 om evaluering av utdanningskvalitet i Sverige).

¹⁰https://www.vr.se/download/18.6969eb1a16a5bec8b5989f/1557492643057/Research-overview-humanities-and-social-sciences_VR_2010.pdf

¹¹<https://www.vr.se/english/analysis/reports/our-reports/2019-09-16-the-swedish-research-barometer-2019.html>

2.3 Finland

I Finland er det universitetene som selv er ansvarlig for å evaluere forskningsaktiviteten, og i henhold til lovgivningen må de jevnlig delta i eksterne evalueringer og kvalitets-sikringssystemer.¹² I tillegg er det lovfestet at universitetene må publisere resultatene av slike evalueringer. Det legges ikke noen klare føringer på hvordan den eksterne evalueringen skal gjennomføres, det vil si hverken innretningen eller omfanget av evalueringene. Dette begrunnes med at universitetene skal ha en høy grad av autonomi.

Hva finnes av evalueringer?

I utdanningslovgivningen står det spesifisert at finske universiteter skal innrettes på en måte som sikrer høy internasjonal standard i forskning, utdanning og forskningsintegritet. På grunn av mangelen på sentral styring av evalueringene og variasjonene mellom de ulike institusjonene, vil vi nedenfor beskrive hvordan Universitetet i Helsinki har gjennomført evaluering av kvalitet i forskning.¹³

Universitetet i Helsinki gjennomførte en omfattende evaluering av forskning i 2018-2019. Det var første helhetlige evaluering siden perioden 2010-2012. Universitetet i Helsinki skriver at de baserte sin evaluering på innretningen i Nederland (beskrevet i kapittel 2.4), i tillegg til noen aspekter hentet fra REF i Storbritannia (beskrevet i kapittel 2.1). Evalueringen ble gjennomført som en kombinasjon av bibliometrisk analyse, egenevaluering og panelintervju. Målet var å gi en sammenfatning av forskernes resultater som ga en oversikt over hva som hadde blitt produsert i en gitt tidsperiode, men først og fremst hva som var forbedringspotensialet for den kommende perioden.

Universitetet i Helsinki fremhever at en slik evaluering innebærer mye administrativt arbeid, og at tidskostnaden er høy, men at det likevel gir en helhetlig oversikt over nivået på forskningen og verdifull informasjon om universitetets forbedringspotensiale. Evalueringen skulle være tett knyttet mot fakultetenes og instituttene egne strategimål, og følgelig ble både sammensetningen av forskere og enhetenes levedyktighet vesentlige momenter i evalueringen. De enhetene som fikk den beste evalueringen fikk i tillegg ekstra finansiering, mens det ikke var lagt opp til noen umiddelbare konsekvenser for de enhetene der hele eller deler av evalueringen viste seg å ikke gi tilfredsstillende resultater.

Tilsvarende evalueringer har blitt gjennomført av alle universitetene i Finland, med litt forskjellig innretning. University of Eastern Finland har for eksempel valgt en modell der de evaluerer forskningsmiljøer, ikke organisatoriske enheter ved universitetet. Her identifiseres forskningsmiljøer som tematiske, multidisiplinære forskningsgrupper med internasjonalt nettverk og et klart identifiserbart forskningsfokus.¹⁴

I tillegg til universitetenes egeninitierte evalueringer oppsummeres forskningsaktiviteten av det finske forskningsrådet (Academy of Finland) i form av bibliometriske analyser og vurderinger av forskningskvalitet. Dette er en øvelse som de siste 10 årene har blitt gjennomført annethvert år, mens det ble gjennomført hvert tredje år fra 1997 til og med 2012.¹⁵ I denne oversikten presenteres i tillegg anbefalinger basert på funnene, som også inneholder oversikt over finansiering av forskning, sammensetningen av ansatte (blant annet midlertidige og faste ansatte) og tilslag på EU-prosjekter.

¹² Se seksjon 87 punkt 1: <https://www.finlex.fi/fi/laki/kaannokset/2009/en20090558.pdf> (Side 33).

¹³ https://www.helsinki.fi/sites/default/files/atoms/files/rauh_final_report_060919_full_0.pdf

¹⁴ <https://www.uef.fi/en/research-development-and-evaluation>.

¹⁵ <https://www.aka.fi/en/research-and-science-policy/state-of-scientific-research-in-finland/aiemmat-arvioinnit/>

2.4 Nederland

I Nederland skal alle forskningsinstitusjoner evalueres hvert 6. år. Hvordan dette gjennomføres beskrives i Standard Evaluation Protocol (SEP), som utvikles av det nederlandske forskningsrådet (*Nederlandse Organisatie voor Wetenschappelijk Onderzoek*, NWO) sammen med Association of Universities in the Netherlands (VSNU) og Royal Netherlands Academy of Arts and Sciences (KNAW). SEP har vært i utvikling i over 20 år, og beskriver hvilke metoder, kriterier og dokumentasjon som skal brukes i evalueringene. Nåværende protokoll gjelder for perioden 2015-2021, og det er besluttet at neste protokoll (2021-2027) i stor grad skal følge denne.

Formålet med evalueringer

Ifølge den nåværende protokollen, er formålet med SEP å avdekke og bekrefte forskningens kvalitet og relevans for samfunnet, samt å forbedre dette ved behov. Vurderingen omfatter både vitenskapelige, samfunnsmessige og strategiske aspekter ved forskningen. Evalueringene retter seg mot flere målgrupper, både forskerne selv og ledelsen ved de ulike forskningsinstitusjonene, men også myndigheter, privat sektor og samfunnet generelt.

Innretning og gjennomføring

I Nederland må universiteter og institutter selv sørge for at de får alle sine forskningsenheter evaluert hvert 6. år, og at denne evalueringen gjennomføres i tråd med SEP.¹⁶ Evalueringene gjøres ikke på en gang, men følger et rullerende skjema med 6-årssykluser. Styret på hver institusjon bestemmer når en evaluering skal finne sted og hvilke enheter som skal evalueres. Dette kan variere fra gang til gang, og kan være alt fra forskningsgrupper til hele fakultet.¹⁷ Det kan i noen tilfeller være hensiktsmessig å evaluere institutter samlet, mens det i andre tilfeller kan være hensiktsmessig å evaluere forskningsgrupper eller tverrfaglig samarbeid. Styret setter deretter opp en tidsplan og informerer de aktuelle enhetene i god tid. De skal også, i samråd med hver enhet, sette passende mål («benchmarks») og innretning («Terms og Reference») for evalueringen. Terms of Reference kan gi informasjon om forskningsenheten som evalueringskomitéen må ta hensyn til i sin vurdering. Dette kan være informasjon rundt for eksempel strategiske spørsmål eller spesielle oppgaver enheten har.

Styret er også ansvarlige for alt det praktiske rundt gjennomføringen av en evaluering, og for å sette sammen en evalueringskomité. Disse komitéene er ofte internasjonale og skal ha kunnskap både om fagområdet og om det nederlandske forsknings- og evalueringssystemet.

Evalueringen foregår ved at forskningsenheten først gjennomfører en selvevaluering og samler inn dokumentasjon. Når komitéen har mottatt egenevalueringen og annen dokumentasjon, besøker de institusjonen («site visit»). I protokollen er det beskrevet hva et slikt besøk innebærer. Det legges opp intervjuer med et gitt antall personer som skal representere bredden på enheten, og det er forventet at evalueringskomitéen presenterer foreløpige funn avslutningsvis. På bakgrunn av dette foretar komitéen en evaluering av

¹⁶ Informasjon om innretningen av evalueringer i Nederland er godt beskrevet i SEP 2015-2021: <https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf>. Beskrivelsen under er basert på den inneværende protokollen.

¹⁷ Se krav til hva som kan telle som en «research unit» på s. 10 i den inneværende protokollen, <https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/SEP2015-2021.pdf>.

tre overordnede vurderingskriterier: forskningskvalitet, forskningens relevans for samfunnet og levedyktighet (i hvilken grad enheten er forberedt på fremtiden).

Vurderingen av kvalitet i forskningen baserer seg på enhetens bidra til fagfeltet, samt omfanget av forskningsresultater (publikasjoner, instrumenter, infrastruktur og annet som er utviklet av enheten). Samfunnsrelevans baserer seg på kvaliteten, omfanget og relevansen av enhetens bidrag til samfunnet. Dette kan være bidrag til politikktutforming, bidrag til den offentlige debatten eller lignende. Komitéen skal vurdere bidrag på områder som enheten selv har definert som relevante og der de har hatt et mål om å bidra, og skal dermed evaluere om slike mål er oppnådd. Til slutt vurderes enhetenes levedyktighet, noe som innebærer at evalueringene også ser fremover i tid. Her evalueres enhetens strategier, og hvorvidt de er i stand til å oppnå målene sine. Evaluering av strategien kan bygge bl.a. på vurderinger av utsiktene til å skaffe finansiering, av rekruttering og av forskningsledelse.

Endelig rapport skal foreligge senest et halvt år etter komitéens besøk og skal inneholde både en kvalitativ og en kvantitativ tilbakemelding på hvert av de tre evalueringskriteriene. Den kvantitative tilbakemeldingen er basert på fire kategorier, «excellent», «very good», «good» og «unsatisfactory». I tillegg skal komitéen komme med anbefalinger for fremtiden og en overordnet kvalitativ vurdering av enheten. I evalueringen skal komitéen ta hensyn til målene forskningsenhetene har satt for seg selv. Deretter sendes rapporten til styret på institusjonen, som setter seg inn i rapporten og den aktuelle forskningsenhetens kommentarer. Basert på dette utformer de et eget svar («position document»), der de også legger frem hvordan evalueringen vil følges opp. Til slutt blir både evalueringen og styrets «position document» publisert.

I tillegg er det flere universiteter som gjennomfører en midtveiseevalueringer slik at de er forberedt på den eksterne evalueringen.¹⁸

2.5 Sammenligning og sentrale dimensjoner i vurderingen av forskningsevalueringer

I det følgende trekker vi fram noen dimensjoner som kan være nyttige for å sammenligne forskjellige evalueringssystemer rettet mot forskning. Vi baserer oss til dels på vår egen gjennomgang, og viser også til eksisterende internasjonale sammenligninger og forsøk på å utvikle typologier for evalueringssystemer. Vi fokuserer spesielt på å tydeliggjøre hvilke dimensjoner som er sentrale for å forstå forskningsevalueringer, fremfor en detaljert sammenligning av detaljene i de forskjellige lands systemer. Dette er en følge av at evalueringssystemene er særdeles kompliserte, og vi anser det som mest hensiktsmessig for utarbeidelsen av et nytt rammeverk å tydeliggjøre hvilke overordnede avveininger og vurdering som forskjellige evalueringssystemer innebærer. I forskningslitteraturen ser vi også tydelig at det er lite enighet om hvordan forskjellige lands systemer bør beskrives og sammenlignes, og det er relativt få entydige aspekter ved de forskjellige lands systemer som gjør det mulig å sammenligne landene på en måte som gir ukontroversielle og nyttige rangeringer av sentrale aspekter ved systemene. Alle landene baserer seg for eksempel til en viss grad på fagfellevurderinger og indikatorer, og det er detaljene i hvordan dette gjøres som skiller landene fra hverandre (Geuna & martin 2003).

¹⁸<https://www.staff.universiteitleiden.nl/research/quality-and-integrity/research-visitations/research-visitations/service-units/leiden-university-libraries>

Oschner (2020) peker på en rekke tidligere forsøk på å lage typologier og gjennomføre internasjonale sammenligninger av evalueringssystemer. Blant disse er Coryn et al. (2007) Geuna & Martin (2001, 2003), Hicks (2012), Jonkers & Zacharewicz (2016), von Tunzelman & Mbula (2003) og Whitley & Gläser (2007).

Mange av disse begynner imidlertid å bli relativt gamle, ikke alle omhandler alle landene vi her undersøker, og mange er også eksplisitt rettet mot hvordan evaluering knyttes til *finansiering*. Vi skal som nevnt ikke vurdere forhold knyttet til finansiering. Whitney & Gläser (2007), Coryn et al. (2007) og Geuna & Martin (2003) er imidlertid av interesse i denne sammenheng. Sistnevnte skiller mellom *ex ante* og *ex post* evalueringer, der *ex ante* evalueringer er rettet mot å forutsi og forbedre resultater, mens *ex post* er bakoverskuende og mer resultatorienterte (Geuna & Martin, 2003). Formative og summative evalueringer har vi allerede diskutert i eget notat om typologisering av evalueringer, og dette diskuteres også i noe mer detalj nedenfor.

Coryn et al. (2007) utvikler et rammeverk for internasjonal sammenligning av offentlig finansiert forskning. De inkluderer 16 land, og alle de fire landene vi undersøker er inkludert. De gjennomfører kategoriseringer på flere dimensjoner, og vi ser her kort på dimensjonen som ikke omhandler finansieringssystemer. De følger her Campbell (2002) i skillet mellom to hovedtilnærminger til evalueringer: En «systematisk og konsistent» (Type A) og en annen preget av pluralisme og «situasjonsbestemt variasjon knyttet til konseptualisering, metoder og anvendelser» (type B) (Coryn et al., 2007, s. 443). Her klassifiseres Nederland og Storbritannia som type A og Finland og Sverige som type B. Som vi har vist, og diskuterer videre nedenfor, er det noe usikkert om kategoriseringen av Nederland som et type A-system er treffende, ettersom det er lagt opp til gode muligheter for både variasjon og institusjonell medbestemmelse i det nederlandske systemet. Forfatterne går så videre og utvikler et system for å vurdere *kvaliteten* på hvert lands evalueringssystem, der hvert land får karakter fra A (eksellent) til F («absence of merit; clearly inadequate; fatal deficiencies»). Flere paneler med eksperter vurderer hvert lands system, og scorer dem på dimensjonene *validitet*, *troverdighet*, *nytte*, *kostnads-effektivitet* og *etikk*. Konklusjonen er at Finland og Sverige får karakteren *F*, mens Nederland og Storbritannia får henholdsvis karakteren *C* av et panel fra New Zealand og *D* av et panel fra USA (Coryn et al., 2007). Vi går ikke videre inn i denne klassifiseringen her, og inkluderer den primært for å vise til en av få større forsøk på kvantitative sammenligninger av evalueringssystemer.

Whitney (2007, s. 9) foreslår et skille mellom *sterke* og *svake* evalueringssystemer. Svake evalueringssystemer omtales på følgende måte:

They are typically organised informally by funding agencies and/or a consortia of universities with little standardisation of procedures or criteria. They rarely publish their conclusions, which are usually intended to encourage organisational improvements, and so be formative, rather than to arrive at summative judgements. Whilst these assessments focus on the overall effectiveness and quality of research groups and departments, they do not rank them [...], nor do they have direct financial consequences, although they may have some indirect ones in the medium term. While such RES [research evaluation systems] may encourage scientists and universities to consider their research strategies more explicitly and to focus more on producing research that is recognised internationally as significant, their impact on current arrangements is likely to be incremental rather than radical ...

Sterke systemer omtales slik:

Strong RES, in contrast, institutionalise public assessments of the quality of the research conducted in individual departments and universities by scientific elites on a regular basis according to highly formalised rules and procedures. These assessments are usually ranked on a standard scale and published so that the relative standing of universities and departments can be readily ascertained. In most cases, they are organised around existing disciplines and scientific boundaries. Such peer review-based evaluations directly affect funding decisions [...], and so can have a considerable impact on the management of universities

I et slikt rammeverk er det for eksempel naturlig å plassere Storbritannia i kategorien *sterkt system*, mens de øvrige landene i større grad beveger seg mot *svake systemer* (Simon & Knie, 2013). Simon & Knie (2013) påpeker imidlertid at disse to kategoriene ikke er tilstrekkelig for å gi et godt bilde av forskjellige systemer, og spesielt skillet mellom *formativ* og *summativ* vurdering – eller *prosess-* og *resultat-fokus* – trekkes også frem av disse. De nevner særlig de skandinaviske landene som eksempler på steder der man har fokusert på formative tilbakemeldinger. Videre peker de spesielt på Nederland som et eksempel på bruk av evalueringer som et «management tool», fremfor ren måling og kvalitetskontroll (Simon & Knie, 2013).

I tillegg til at forskningsevaluering i seg selv er et komplisert fagområde har også hvert land som sammenlignes her særegne systemer som er kompliserte. Også ekspertene på evalueringssystemer er ofte uenige i hvordan et system bør fremstilles og vurderes, noe for eksempel Sivertsens (2017) artikkel om hvordan den «norske modellen» misforstås av andre eksperter på forskningsevaluering tydelig viser.

Vi anser det derfor som hensiktsmessig å skissere noen grove dimensjoner som tydeliggjør forskjellene mellom de fire landene, og som samtidig viser hvilke avveininger som må gjøres når evalueringssystem velges. Evalueringer kan ha mange forskjellige formål, og de forskjellige typologiene nevnt ovenfor peker mot bred enighet om at intet praktisk gjennomførbart evalueringssystem er i stand til å dekke *alle* potensielle formål og behov fullt ut.

Dimensjonene vi har valgt å fokusere på er: autonomi/standardisering, særegenheter/universalitet, variasjon/ensretting, forståelse/sammenligning, formativ/summativ og forhold knyttet til periodisering. Noen av dimensjonene diskuteres kort, og poenger fra de forskjellige dimensjonene spiller sammen inn på den overordnede dimensjonen *formativ* vs. *summativ* evaluering.

Autonomi vs. standardisering

Første dimensjon omhandler graden av nasjonalt rammeverk og koordinering sett opp mot graden av institusjonell frihet til å velge og gjennomføre evalueringer. Her ser vi en fordeling på dimensjonen med REF i Storbritannia på den ene siden (sterk nasjonal koordinering), Nederland i midten (nasjonale retningslinjer, men omfattende muligheter for institusjonell medvirkning i valg av tidspunkt, evalueringsenhet og ambisjoner/mål) og Finland i den andre enden (stor institusjonell autonomi).

Finland har også vært drevet av et ønske om å bruke evalueringer til kvalitets*utvikling* fremfor kvalitets*måling* (Geuna & Martin, 2003). Høy grad av autonomi og institusjonelt eierskap og ansvar for evalueringene bidrar til dette, men gjør som i Nederland at mulighetene for å sammenligne enheter blir dårligere. Det har tidligere vært foreslått å innarbeide ytelsesbasert finansiering lik det vi ser i Storbritannia, men dette møtte store protester og ble avvist på slutten av 90-tallet (Geuna & Martin, 2003).

Særegenheter vs. universalitet

Neste dimensjon omhandler i hvilken grad evalueringene tar hensyn til institusjonenes egne premisser/ambisjoner (i UK blir alle målt på samme skala etter samme kriterier, i Nederland setter institusjonene til en viss grad sine egne mål og sier selv hvem de vil benchmarkes mot). I Finland skal evalueringene knyttes til strategiske mål ved de enkelte institusjonene.

Sivertsen (2017) påpeker at bruk av standardiserte mål og indikatorer muliggjør sammenligning, men at de samtidig gjør det vanskelig å bruke evalueringer formativt. Indikatorer er, sier han «tilbakeskuende», mens evaluering på *organisasjonsnivå* muliggjør formativ evaluering (Sivertsen, 2017, s. 2).

Varierte vs. uniforme undersøkelsesenheter

Et viktig tema i utforming av evalueringssystemer er valg av undersøkelsesenhet. Dette henger tett sammen med de øvrige dimensjonene, men det er allikevel verdt å nevne dette som et eget punkt, siden det påvirker mulighetsrommet for bruk av evalueringer.

På dette punktet er det ingen av landene som har helt faste regler der de samme enhetene på alle institusjoner blir evaluert på en konsistent måte over tid. I REF i UK deles evalueringene opp etter units of assessment (som tilsvarer fagområder). Hver institusjon kan levere inn ett bidrag per unit of assessment, og får dermed en overordnet vurdering på det aktuelle fagfeltet. Dette gjør at institusjonene som har blitt evaluert i samme unit of assessment kan sammenlignes. Det er likevel viktig å presisere at det kan være store variasjoner i hva institusjonene har sendt inn til vurdering ettersom de selv kan bestemme hvilke ansatte, resultater og/eller forskergrupper som skal inkluderes. I den andre enden av spekteret finner vi Nederland, hvor institusjonene selv avgjør hvilke enheter (for eksempel fakultet, institutt eller forskergruppe) de ønsker skal bli evaluert fra gang til gang.

Inngående forståelse vs. sammenligning

Oppsummerer vi dimensjonene 1-3 kan vi se at det er mulig å peke på en overordnet dimensjon som omhandler graden av inngående forståelse sett opp mot mulighetene til å sammenligne undersøkelsesenheter på nasjonalt nivå.

Ochsner et al. (2020) skriver for eksempel om hvordan samfunnsfag og humaniora (SSH, social sciences and humanities) skiller seg betydelig fra naturfagene, og i rapporten om bruk av fagfelleevaluering i evalueringer går de langt i å si at de to faggruppene *bør* evalueres på forskjellige måter.

I et kapittel i den nevnte rapporten beskriver Ochsner (2020) en typologi bestående av 5 idealtyper, der *database og bibliometri* og *SSH-tilpasning* lar ham plassere et utvalg land i et landskap med to dimensjoner. Her er Finland, Norge og Nederland inkludert, og Nederland rangeres som *mest* SSH-tilpasset, deretter følger Norge og til sist Finland. Det er imidlertid relativt små forskjeller mellom landene på denne dimensjonen. Ut fra vår gjennomgang av de fire landene vil det være naturlig å anta at Storbritannia ville bli vurdert som *mindre* SSH-tilpasset enn de øvrige.

En fordel med sammenligning er ifølge Simon & Knie (2013) at det bidrar til å skape *konkurranse* mellom forskere og institusjoner. Dette fremkommer mest tydelig i Storbritannia, på grunn av mulighetene for å rangere og sammenligne (Simon & Knie, 2013). I tilknytning til dette er det også flere som peker på at forskere raskt responderer på endrede insentivsystemer, og at f.eks. vektlegging av siteringer fører til «citation gaming», og strategisk tilnærming til siteringsvirksomhet (Baccini, De Nicola & Petrovich,

2019). Dette kan igjen føre til «bytte» av siteringer mellom forskere i et land, og dermed til mer «innadvendt» siteringsaktivitet, noe Baccini et al., 2019 undersøker over tid i flere land (bl.a. Storbritannia, Sverige og Nederland).

Hovedavveiningen som må gjøres i denne sammenheng er om man ønsker et evalueringssystemer som legger opp til at særegenhetene til disipliner, institusjoner, osv. tas høyde for, slik at det blir mulig for evaluator å *forstå* evalueringssubjektet inngående, fremfor å beskrive det ut fra en mer generell og overfladisk tilnærming. Sistnevnte gjør det i større grad mulig å sammenligne undersøkelsesobjektene, mens en tilnærming basert på inngående forståelse vil kunne stå i veien for det samme.

Formativ mot summativ

Tett knyttet til den foregående dimensjonen er skillet mellom *formativ* og *summativ* evaluering. Geuna & Martin (2003) peker på Storbritannia som det landet med tydeligst vekt på summativ evaluering. Dette på grunn av at en viss grad av standardisering gjør sammenligning mulig, men flere fremhever også at fokuset på *individuelle forskere fremfor institusjoner* forhindrer Storbritanniens evalueringer fra å bli brukt formativt (Simon & Knie, 2013; Sivertsen, 2017). Det er også i Storbritannia denne formen for evalueringer er mest relevant for finansiering av forskning, mens evalueringresultatene også er en viktig komponent for finansiering av forskning i Finland. I Sverige og Nederland virker det ikke som om det er noen umiddelbare konsekvenser for finansiering av forskning om evalueringene ikke viser tilfredsstillende resultat.

På den annen side blir Nederland fremhevet som et land med sterk vektlegging av det formative aspektet ved evalueringer (Geuna & Martin, 2003). Dette skyldes ifølge Geuna & Martin (2003) at både forskningsmiljøene selv og politikere vektlegger behovet for strategiutvikling og uformelle «nedenfra-opp»-evalueringer. En konsekvens av dette er at undersøkelsesenheter i Nederland vil variere fra disiplin til disiplin og program til program, noe som igjen gjør sammenligninger vanskelig. Simon & Knie (2013) konkluderer på samme måte at fleksibilitet og institusjonell autonomi bidrar til å fremme det formative elementet i Nederlands evalueringssystem. Det at institusjonene og enhetene selv kan sette sine egne ambisjoner og mål, samt velge hvem de vil benchmarkes mot, til rettelegger for formativ bruk av evalueringene i kvalitetsutvikling. Finlands høye grad av institusjonell autonomi gjør også at det formative elementet fremstår som sterkt her (Geuna & Martin, 2003).

Et annet viktig aspekt som trekkes frem er behovet for *interaksjon* mellom evaluanden og evaluator dersom evalueringene skal gi grunnlag for organisasjonsendring og utvikling (Simon & Knie, 2013). I Nederland har man *høy* grad av interaksjon, mens man i Storbritannia har *liten* grad av dette (Simon & Knie, 2013). Med interaksjon følger også stor grad av åpenhet og transparens, og Simon & Knie (2013) finner at dette har svært positive konsekvenser for de formative elementene av evalueringene.

Syklus mot enhetlig gjennomføring

Avslutningsvis ser vi kort på hvordan, og når, evalueringer skal gjennomføres. Selv om innretningen og omfanget av evalueringene er forskjellig i alle landene, har alle som felles mål at kvalitet i forskning skal evalueres innenfor en fastsatt tidsperiode. Storbritannia har på den ene siden en svært fast syklus der en stor nasjonal forskningsevaluering foretas hvert 6. år. Nederland har også en syklus på 6 år, men her kan institusjonene selv bestemme hvilke enheter som skal evalueres når innenfor hver syklus. Finland har på sin side ikke en fast syklus, men at evalueringer skal gjennomføres «jevnlige». Jevnlige evalueringer har i praksis vist seg å være mellom hvert sjettede og hvert

åttende år. Til sist skiller Sverige seg litt ut i denne dimensjonen ettersom deres evalueringssystem er svært mangfoldig og inneholder mange ulike typer evalueringer og oppsummeringer av forskningsaktiviteten i landet.

I Finland og Sverige blir det gjennomført regelmessige summative evalueringer av forskning. I Finland blir det annethvert år gjennomført en oppsummering av forskningsproduksjoner, inkludert bibliometriske analyser, men også om sammensetningen av ansatte (dvs. hvor i karriereløpet forskere er), levedyktigheten til forskningsmiljøene og finansiering av forskningen. Også Sverige oppsummerer forskningen annethvert år i et forskningsbarometer.

3 Kvalitet i utdanning

I dette kapittelet beskriver vi innretningen av evalueringen av kvalitet i høyere utdanning i Storbritannia, Sverige, Finland og Nederland. Som i kapittel 2, presenterer vi innretningen av evalueringene som kvalitetsvirkemiddel hver for seg, før vi diskuterer de dimensjonene som fremstår som mest sentrale for å vurdere evalueringssystemer for utdanningskvalitet.

3.1 Storbritannia

I Storbritannia er det The Quality Assurance Agency for Higher Education (QAA) som er ansvarlige for eksterne evalueringer av høyere utdanning. Systemet baserer seg ikke på en nasjonal akkreditering av studieprogrammer, men det er et krav om at utdanningsinstitusjonene har egne kvalitetssikringssystemer som sikrer at studieprogrammene følger en angitt standard.¹⁹ Denne kvalitetsstandard, «UK Quality Code for Higher Education», utvikles av QAA i samarbeid med sektoren. Ifølge QAA skal standarden på denne måten være et kollegialt produkt, der sektoren har innflytelse, og dermed et visst eierskap, over retningslinjene.²⁰ Standarden er basert på tre elementer. For det første presenteres noen forventninger som høyere utdanningsinstitusjoner forventes å oppfylle i sitt eget kvalitetsarbeid. For det andre inneholder standarden eksempler på *best practice* som kan bidra til å oppnå disse resultatene. Dette inkluderer både kjernepraksiser («core practices») som alle høyere utdanningsinstitusjoner skal demonstrere og vanlige praksiser («common practices») som bør følges av institusjonene i tråd med institusjonenes egne mål, kontekst og studentenes behov. Til sist inneholder standarden råd og retningslinjer som vil hjelpe institusjoner med å utvikle og opprettholde effektiv kvalitetssikring.²¹ Blant annet har QAA utviklet råd og retningslinjer for institusjonenes eget overvåkings- og evalueringsarbeid.²² Her nevnes det blant annet at institusjonene selv bør gjennomføre kvalitetssikring basert på overvåking og evaluering av studieprogram. Overvåking skjer i form av kontinuerlig gjennomgang av læringsprosessen til studentene og om studentene oppnår den kompetansen som på forhånd er definert. I tillegg bør institusjonene gjennomføre egevaluering av studieprogrammene. Det er derimot ikke spesifisert hvor ofte slike evalueringer bør gjennomføres.

QAA har ikke en myndighetsrolle, og kan heller ikke sanksjonere utdanningsinstitusjoner som ikke tilfredsstill disse retningslinjene. Høyere utdanning ligger under de nasjonale myndighetene i Storbritannia, og det er derfor noen nasjonale forskjeller i hvordan kvalitet i høyere utdanning vurderes og hvordan kvalitetsstandard brukes i styringen. For eksempel er forventningene og kjernepraksisene i standarden obligatoriske for alle høyere utdanningsinstitusjoner i Storbritannia, mens de vanlige praksisene er frivillige i England, men obligatoriske for institusjoner i Skottland, Nord-Irland og Wales.

Selv om QAA ikke har en formell myndighetsrolle, leverer de retningsgivende dokumenter og anbefalinger til de nasjonale finansierings- og reguleringsorganene. I England er dette Office for Students (OfS), som er ansvarlig for å regulere og finansiere høyere utdanning, samt å vurdere hvorvidt tilbydere av høyere utdanning leverer tilfredsstillende kvalitet. Her brukes kvalitetsstandard som et utgangspunkt, og vurderingene gjøres i

¹⁹ Visse programmer er imidlertid akkreditert, og det er forskjellige akkrediteringsorganer involvert i dette (Stensaker & Harvey, 2006).

²⁰ <https://www.qaa.ac.uk/about-us/what-we-do/our-work>.

²¹ «The revised UK Quality Code for higher Education», s. 2. Hentet fra: https://www.qaa.ac.uk/docs/qaa/quality-code/revised-uk-quality-code-for-higher-education.pdf?sfvrsn=4c19f781_8.

²² <https://www.qaa.ac.uk/en/quality-code/advice-and-guidance/monitoring-and-evaluation>

samarbeid med QAA. For en oversikt over og kritikk av utdanningsevaluering i Storbritannia frem til 2005 viser vi til Harvey (2005).

Når det gjelder eksterne evalueringer, omtaler vi her England som et eksempel, da dette varierer mellom landene. Her er innretningen slik at utdanningsinstitusjoner må vise at de har en såkalt «Access and participation plan». Dette er et dokument som sier noe om sammensetningen av studenter, men også kvalitet i utdanningen.²³ Hvis et studiested ønsker å starte et nytt studieprogram og dermed utstede en akademisk grad, blir eksterne eksperter hyret inn for å vurdere i hvilken grad kriteriene i planen blir tilfredsstillende fulgt. Det er OfS som er ansvarlig for at studiestedene følger kravene som stilles i «access and participation plan».

3.2 Sverige

I Sverige er det Universitetskanslersämbetet (UKÄ) som har ansvaret for å vurdere kvaliteten på høyere utdanning. UKÄ ble opprettet i 2013 og overtok da dette ansvaret fra Högskoleverket.²⁴ Siden 2013 har UKÄ, i samarbeid med universitetene og høyskolene, utviklet et nasjonalt kvalitetssikringssystem på oppdrag fra regjeringen. Dette systemet ble ferdig utviklet og testet i 2016, og ble tatt i bruk fra 2017. Et sentralt prinsipp i systemet er at lærestedene og UKÄ har et felles ansvar for å kvalitetssikre høyere utdanning i Sverige. Systemet er basert på den svenske Høyskoleloven og Høyskoleforordningen, samt den europeiske standarden Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG).

Det nasjonale kvalitetssikringssystemet består av en modell med fire typer evalueringer som alle skal gjennomføres i en evalueringssyklus på seks år (nåværende syklus går fra 2017-2022).²⁵ Disse fire evalueringene er 1) evalueringer av institusjonenes kvalitetssikringsarbeid, 2) vurderinger av søknader fra institusjoner som ønsker å utstede akademiske grader, 3) utdanningsevalueringer og 4) tematiske vurderinger. Evalueringene er basert på fagfelleevalueringer, og UKÄ oppnevner eksterne evalueringspaneler bestående av eksperter, studentrepresentanter og representanter fra arbeidslivet.

Når det gjelder institusjonenes egne kvalitetssikringssystem, er formålet med evalueringen å undersøke i hvilken grad institusjonenes interne kvalitetssikringsprosesser bidrar til å forsterke og opprettholde kvaliteten på emner og studieprogrammer. Evalueringene følger prinsippet om at kvalitetssikringen er institusjonenes eget ansvar, og de skal dermed ikke bare kontrollere institusjonene, men også hjelpe dem med å forbedre sine egne kvalitetssystemer. Denne formen for evaluering baserer seg på egevaluering, en rapport om studieforholdene, intervjuer med representanter for institusjonen, samt site visits. Basert på en samlet vurdering vil institusjonen enten få godkjent uten bemerkninger, godkjent med bemerkninger, eller ikke godkjent. Hvis godkjent med bemerkninger, må bemerkningene utbedres. Evalueringskomiteen følger opp alle institusjoner gjennom dialogmøter i ettertid. Hvis en institusjon blir godkjent med bemerkninger, har den to år på seg til å rette opp manglene. Dersom et system ikke blir godkjent, skal ikke-tilfredsstillende områdene evalueres på nytt etter en periode som bestemmes av UKÄ i samarbeid med lærestedet.²⁶

Evaluering av søknader fra institusjoner som ønsker å utstede akademiske grader baserer seg på læringsutbyttet til studentene, altså om institusjonen legger godt nok til

²³ Se vedlegg A1 på side 81: https://www.officeforstudents.org.uk/media/1406/ofs2018_01.pdf

²⁴ <https://www.uka.se/om-oss/var-verksamhet/historik.html>.

²⁵ <https://www.uka.se/kvalitet--examenstillstand/sa-granskas-hogre-utbildning.html>.

²⁶ <https://www.uka.se/kvalitet--examenstillstand/granskning-av-larosatenas-kvalitetssakringsarbete.html>.

rette for at studentene skal oppnå de nødvendige resultatene som følger av graden det søkes om. Evalueringen gjøres på grunnlag av søknaden, samt intervjuer med representanter for studieprogrammet. I tillegg blir det gjennomført intervjuer av ledere for utdanningsprogrammet. Panelet skriver deretter en rapport der de enten anbefaler å godkjenne eller avvise søknaden. UKÄ kan ta beslutninger på slike søknader fra statlige universiteter og høyskoler (med noen unntak), mens søknader som gjelder andre institusjoner sendes til regjeringen.

UKÄ foretar også evalueringer av et utvalg utdanninger rundt om i Sverige. I perioden 2017-2020 skal for eksempel førskolelærer-, yrkeslærer og sykepleierutdanningene evalueres. Her vurderer panelet hvorvidt studentene får gode forutsetninger for å nå eksamensmålene. Her blir studieprogrammene enten godkjent eller ikke godkjent, og evalueringsrapporten skal også brukes som tilbakemelding for studiestedet. Ved eventuelle bemerkninger har studiestedet ett år til å utbedre disse, og hvis ikke disse bemerkningene er tilfredsstillende forbedret, kan det i ytterste konsekvens resultere i at studieprogrammet ikke lenger kan utstede akademiske grader. Dette innebærer at tilsyn har konsekvenser som gjør at skillet mellom tilsyn og akkreditering blir relativt lite.

Det blir også gjennomført tematiske evalueringer, som skal bidra med viktig kunnskap og nasjonale sammenligninger av de ulike institusjonenes arbeid. «Tema» i denne sammenhengen betyr ikke fagområder, men ulike områder der alle læresteder har et felles ansvar og som på ulike måter påvirker kvaliteten i høyere utdanning. Den første slike tematiske evalueringen ble gjennomført i 2017 og omhandlet lærestedenes arbeid med bærekraftig utvikling. I 2020 evalueres institusjonenes arbeid med inkluderende rekruttering. Slike evalueringer følger ikke noen bestemt manual for hvordan evalueringen skal gjennomføres, og evalueringsmetodene bestemmes avhengig av hvilket tema som skal evalueres.

UKÄ utvikler en plan for hver 6-årssyklus der både ressursbruk og hva som anses som relevante temaer tillegges vekt. Dette betyr at hvilke og hvor mange evalueringer som gjennomføres kan variere fra år til år. I hver evalueringsperiode skal imidlertid samtlige universiteters og høyskolars interne kvalitetssikringssystemer ha blitt evaluert. Hyppigheten og omfanget av de andre evalueringstypene vil derimot variere. For eksempel vil kun et utvalg utdanningsprogram bli evaluert i hver evalueringsperiode, og dette vil primært være utdanninger som leder til autorisasjon, som for eksempel lege-, veterinær- og fysioterapeututdannelse.²⁷ Tematiske evalueringer vil også variere ut ifra hvilke temaer UKÄ og regjeringen ønsker å undersøke.

3.3 Finland

I Finland er det Kansallinen koulutuksen arviointikeskus (Finnish Education Evaluation Centre, FINEEC) som er ansvarlig for utdanningsevalueringer. FINEEC er etablert som en selvstendig ekspertorganisasjon for eksterne evalueringer av utdanning, og opererer innenfor området til det finske Utdanningsdepartementet.²⁸ FINEEC er ansvarlig for kvalitet i utdanning på alle nivåer - fra barnehage til høyere utdanning.

Innen høyere utdanning har FINEEC tre former for evalueringer: 1) kvalitetsrevisjoner av høyere utdanningsinstitusjoner, 2) tematiske evalueringer og 3) evaluering av ingeniørprogrammer. Sistnevnte er en form for akkreditering der ingeniørprogram kan oppnå det

²⁷ <https://www.uka.se/publikationer--beslut/publikationer--beslut/rapporter/rapporter/2016-09-30-nationellt-system-for-kvalitetssakring-av-hogre-utbildning.html>.

²⁸ Act on the Finnish Education Evaluation Centre (1295/2013), Section 1. Mandate. Hentet fra: <https://finlex.fi/fi/laki/kaannokset/2013/en20131295.pdf>.

internasjonale kvalitetsstempelen EUR-ACE. De tematiske evalueringene omfatter både evalueringer av utdanningsprogram innen visse fagområder (som for eksempel utdanningsprogram innen juss eller maritim sektor) og evalueringer av andre typer tema. Eksempler på dette er en evaluering av integrering av innvandrere i høyere utdanning og en evaluering av overganger i studieløp og hvor godt studieløpet fungerer fra søknad og inntak og gjennom de ulike gradene.²⁹ Vi vil her legge størst vekt på den tredje typen evaluering, som FINEEC selv kaller «kvalitetsrevisjoner» (quality audits) av høyere utdanningsinstitusjoner.³⁰

Kvalitetsrevisjonene gjennomføres i sykluser på 6 år, og FINEEC er nå i gang med sin tredje runde med revisjoner av institusjoner i høyere utdanning, som skal gjøres mellom 2018 og 2024.³¹ Selve revisjonen går ut på at en ekstern vurderingsgruppe går gjennom utdanningsinstitusjonenes egne systemer for kvalitetskontroll av forskning og utdanning. I henhold til retningslinjene kan evalueringen gjennomføres av både nasjonale og internasjonale eksperter. Dersom evalueringen gjennomføres av internasjonale eksperter, skal en deltaker fra Finland med ekspertise på utdanningssystemet delta i revisjonen.

Formålet med FINEECs revisjon av høyere utdanning oppsummeres i følgende fire punkter. FINEEC skal

- evaluere hvorvidt utdanningsinstitusjonenes kvalitetsarbeid lever opp til de europeiske standardene og retningslinjene for kvalitetssikring i høyere utdanning (ESG),
- vurdere hvorvidt kvalitetssystemet gir institusjonene relevant informasjon som de kan bruke i eget strategi- og utviklingsarbeid, og om dette resulterer i effektive utviklingsaktiviteter,
- oppmuntre til internasjonalisering, eksperimentering og kreative miljøer i utdanningsinstitusjonene, og
- akkumulere åpen og transparent informasjon om finske utdanningsinstitusjoners kvalitetsarbeid.

Fokuset i revisjonen er å opprettholde og sørge for kontinuerlig forbedring av kvaliteten til utdanningsinstitusjonene. Dette gjøres ved å hjelpe institusjonene med å gjenkjenne sine egne styrker og forbedringsområder. I revisjonen skal det vurderes om utdanningsinstitusjonene:

1. *skaper kompetanse*, gjennom planlegging, implementering, evaluering og forbedring av utdanningen,
2. fremmer påvirkning og fornyelse, gjennom blant annet sosialt engasjement, impact og fornyelse i organisasjonskulturen, og
3. *fremmer kvalitet og velvære*, gjennom et godt kvalitetssystem, strategisk ledelse og kompetanseutvikling

I tillegg kommer et fjerde punkt der utdanningsinstitusjonen selv kan velge et vurderingsområde som er viktig for dem og som de ønsker å utvikle seg på.

Punkt 1-3 graderes på en skala utmerket-god-ikke tilfredsstillende, der «god» er minimumskravet for å bestå revisjonen, mens punkt 4 ikke er relevant for hvorvidt en institusjon består revisjonen eller ikke. Hvis institusjonene ikke består evalueringene, får

²⁹ <https://karvi.fi/en/higher-education/thematic-and-system-evaluations/>.

³⁰ Fremstillingen av FINEECs kvalitetsrevisjoner er basert på FINEECs nettsider, <https://karvi.fi/en/higher-education/audits-higher-education-institutions-2018-2024/>, samt retningslinjene for revisjonsrunden 2018-2024, hentet fra: https://karvi.fi/app/uploads/2019/09/FINEEC_Audit-manual-for-higher-education-institutions_2019-2024_FINAL.pdf.

³¹ <https://karvi.fi/en/higher-education/audits-higher-education-institutions-2018-2024/>

det ingen konsekvenser utover at det blir gjennomført en ny evaluering der punktene som ikke var tilfredsstillende i den første evalueringen vurderes.

En slik innretning innebærer at evalueringene ikke er en kontroll, men mer en form for tilbakemelding fra en ekstern ekspertgruppe. I tillegg til de faste evalueringskriteriene kan også institusjonene selv påvirke evalueringen til en viss grad, både ved å velge et evalueringsområde og gjennom en prosess som omtales som *benchlearning*. Her velger institusjonen et område der de setter seg et mål og velger en partner for benchlearning. Målet er å lære av en annen organisasjon gjennom interaksjon, erfaringsutveksling og gjensidig utvikling.

Evalueringen foregår utelukkende digitalt, det vil si at det ikke legges opp til besøk hos de enkelte institusjonene. Det gjennomføres imidlertid en form for oppfølgingsseminar der det blir gitt tilbakemeldinger på hva som er punkter for potensielle forbedringer. I tillegg er oppfølgingsseminaret lagt opp som en arena der de enkelte institusjonene deler erfaringer fra kvalitetsarbeidet.

Målet med denne formen for evaluering er at de skal legge et premiss for kontinuerlig utvikling i sektoren. Bestått revisjon (eller evaluering) gir den aktuelle institusjonen et såkalt kvalitetsmerke som er gjeldende i seks år. Innen utgangen av denne seksårsperioden skal institusjonen ha blitt evaluert på nytt. Dersom en institusjon blir vurdert til å ha svært høy kvalitet, kan de, etter en egen individuell vurdering, også motta et kvalitetsstempel for eksellens.

3.4 Nederland

I Nederland blir kvalitet i høyere utdanning ivaretatt gjennom et akkrediteringssystem og periodiske revisjoner av høyere utdanningsinstitusjoner. Det er Nederlands-Vlaamse Accreditatieorganisatie (NVAO) som har ansvaret for å overvåke og undersøke hvorvidt utdanningsinstitusjoner overholder fastsatte lover og regler samt å vurdere om kvaliteten i utdanningen holder et tilfredsstillende nivå.³²

Akkrediteringssystemet ble introdusert i 2002. Fra 2002-10 ble alle eksisterende utdanningsprogram vurdert og eventuelt akkreditert. Deretter ble den andre fasen av akkrediteringssystemet iverksatt i 2011, og det var her institusjonsrevisjonene («institutional audits») først ble introdusert. Fra 2017 ble et nytt rammeverk introdusert som baserer seg mer på tillit og institusjonenes eget ansvar.³³ Dette tillitsperspektivet ble lovfestet i 2018, og den nåværende runden med revisjoner skal i stor grad fokusere på institusjonenes egne kvalitetssikringssystemer og kvalitetskultur.

Både når det gjelder akkreditering og institusjonsrevisjoner, tar NVAO beslutninger basert på uavhengige ekspertvurderinger foretatt av eksterne evalueringspaneler. Panelene kan være nasjonale eller internasjonale, og skal inkludere minst én student. Panelene baserer seg på selvevalueringer og andre dokumenter fra institusjonen, samt site visits (én site visit for vurdering av utdanningsprogrammer og to i institusjonsrevisjoner). Til sist skriver panelet en rådgivende rapport der de gir utdanningsprogrammet eller institusjonen enten en positiv (eventuelt med forbehold) eller en negativ vurdering. Institusjonene får anledning til å kommentere på vurderingen før NVAO tar den endelige beslutningen.

Det er ikke lovpålagt i Nederland å ha et internt kvalitetssikringssystem. Derfor får det ingen store konsekvenser for institusjonen dersom de ikke består institusjonsrevisjonen.

³²https://www.nvao.net/files/attachments/.139/Assessment_Framework_for_the_Higher_Education_Accreditation_System_of_the_Netherlands_2018.pdf

³³<https://www.nvao.net/en/quality-assurance-system-of-the-netherlands>.

Konsekvensene er større når det kommer til akkreditering, ettersom studieprogrammer kan legges ned dersom de ikke består denne vurderingen. Hvis evalueringene avdekker at en utdanning ikke er tilfredsstillende, eller dersom NVAO har mottatt klager på kvaliteten, kan det bli gjennomført ekstra evalueringer eller utredninger. I tilfeller der studier ikke allerede er godkjent, vil evalueringer bli brukt som en del av beslutningsgrunnlaget i akkrediteringssaker. Evalueringspanelene redegjør også i sin rapport for styrkene og forbedringspunktene til det aktuelle utdanningsprogrammet, men dette skal ikke tas hensyn til i beslutningen om akkreditering.

Alle utdanningsprogram i Nederland skal som utgangspunkt vurderes hvert 6. år for å beholde akkrediteringen. Akkrediteringssystemet og institusjonsrevisjonene spiller sammen på den måten at utdanningsprogrammer på institusjoner som har bestått en revisjon gjennomgår en mindre omfattende vurdering enn programmer på institusjoner som ikke har bestått. Dersom institusjonen har bestått en revisjon, vurderes primært faktorer som læringsmål, læringsmiljø, studentvurderinger (som eksamensformer osv.) og oppnådd læringsutbytte. Dersom institusjonen ikke har bestått en revisjon skal i tillegg faktorer som pensum, inntakskrav, ansatte, fasiliteter og systemer for kvalitetssikring vurderes.

Akkreditering av et studieprogram er en nødvendig forutsetning for at utdanningsinstitusjonene kan utstede bachelor- eller mastergrader, og det er i tillegg en nødvendig forutsetning for at studenter kan søke om studielån. Studier som ikke blir godkjent i akkrediteringssystemet, vil heller ikke motta finansiering fra nasjonale myndigheter.

Det nederlandske evalueringssystemet er basert på kriterier både i den nederlandske loven om høyere utdanning og forskning og de europeiske standardene og retningslinjene for kvalitetssikring i høyere utdanning (ESG).

3.5 Sammenligning og sentrale dimensjoner i vurderingen av utdanningsevalueringer

Som i sammenligningen av forskningsevalueringssystemer baserer vi oss her på den forgående gjennomgangen samt forskning på utdanningsevaluering. Det finnes færre eksisterende kategoriseringer med tilordninger av land i forskningen på utdanningsevalueringer, men flere forsøk på å trekke frem de viktigste *dimensjonene* i slike evalueringer er av interesse. I forskningen på dette feltet brukes for øvrig begrepet *evaluering* i noe mindre grad. Noen omtaler «quality evaluation» (Barnett, 1994; Harvey, 2005), mens mange bruker begrepet *kvalitetssikring* («quality assurance») om den type aktiviteter vi her omtaler (Beerkens, 2018; Stensaker et al., 2011). Dette er også et resultat av at for eksempel *tilsyns-* og *akkrediteringsvirksomhet* ikke nødvendigvis faller inn under definisjonen *evaluering*. Vi baserer oss imidlertid her på en bred tilnærming til kvalitetsarbeid i høyere utdanning, og inkluderer for eksempel tilsynsaktivitet, som vi har redegjort for ovenfor. Stensaker et al. (2011) påpeker også at det er noe utydelige skiller mellom de forskjellige formene for evaluering, da for eksempel tilsyn kan bære preg av *akkreditering* der konsekvensene av negative resultater er særlig sterke. Dette gjaldt som vi så blant annet i forbindelse med tilsyn i Sverige.

Barnett peker på et sett med sentrale dimensjoner for å forstå utdanningsevalueringer, og dette er noe forenklet *formål*, *summativ/formativ*, *kvalitetskontroll/kvalitetssikring* og *perspektiv*. Formål handler i stor grad om forskjellige måter å vurdere UH-sektorens funksjon og mandat, samt om man tillater institusjoner å ha noe *ulike* funksjoner og formål. Summativ og formativ evaluering har vi omtalt i detalj i vårt notat om typologisering, og dette omtales også videre nedenfor. Spenningen mellom *kvalitetskontroll* og *kvalitetssikring* knyttes også delvis til forskjellen mellom en summativ og formativ vurdering,

mens *perspektiv* handler om hvorvidt man ser evalueringer primært fra *studentenes* (klientenes) ståsted, eller fagmiljøene og institusjonene (produsentenes) ståsted (Barnett, 1994; Harvey, 2005).

Harvey og Newton (2004) har også utviklet en oversikt over sentrale dimensjoner og aspekter ved utdanningsevaluering, og ser spesielt på *formål, fokus, evalueringsobjekt, begrunnelser og metoder*. De viser at evalueringer i stor grad medfører et syn på kvalitet som er basert på *compliance* og ansvarliggjøring, og foreslår at evalueringsrammeverk i større grad flytter fokus til *transformativ læring*, som vi kommer tilbake til nedenfor.

Vi omtaler i det følgende sentrale dimensjoner ved utdanningsevalueringer og hvilke avveininger som tydeliggjøres av disse dimensjonene. Vi har her valgt å fokusere på følgende dimensjoner: *formålet med høyere utdanning, nivå og evalueringsobjekt, det unike/universelle, omfang og kostnader, kvalitativ/kvantitativ metode, metoder og tilnærming, kontroll og makt og formativ/summativ*. I løpet av denne gjennomgangen omtales også landene som er beskrevet ovenfor.

Formålet med høyere utdanning

For å kunne si noe om *kvaliteten* i høyere utdanning, er det nødvendig å definere *formålet* med høyere utdanning. Ashwin (2020) påpeker at gyldige indikatorer på kvalitet «reflekterer formålet med høyere utdanning». Uten enighet om hva dette formålet egentlig er, er vi imidlertid like langt. Ashwin (2020), Harvey & Newton (2004) og Harvey (2005) foreslår alle en forståelse av formål som er basert på *transformativ læring*, og endringsprosesser som foregår i studentene. Dette peker sterkt mot et behov for en dyp forståelse av det lokale og ulike kontekster, noe vi kommer tilbake til nedenfor. Dette går også igjen hos andre, for eksempel hos Beerkens (2018, s. 273), som skriver at studentenes *læring* «formodentlig er hovedmålet for kvalitetspolitikk». Blant våre land fremstår Nederland som det land som sterkest vektlegger fleksibilitet og et fokus på de forhold som her er diskutert.

Det er tydelig at evalueringer har effekter på institusjonene som evalueres. Flere omtaler økt bruk av evaluering med fokus på kvalitetssikringsprosesser som et middel for å skape mer autonome *institusjoner* som holdes til ansvar (og ikke detaljstyres) (Harvey, 2005; Stensaker et al., 2011; Engebretsen et al., 2012). Dette knyttes igjen gjerne til mer omfattende reformbølger som New Public Management, som har ført til et ønske om mindre sentralisering i forvaltningen og statlige virksomheter (Harvey, 2005). I tilknytning til dette fenomenet trekker Harvey (2005) også frem syn på formålet med offentlig sektor som fokuserer på hvordan UH-sektoren bidrar til samfunnets verdiskapning og økonomiske utvikling. En slik instrumentell tilnærming til UH-sektoren diskuteres videre nedenfor.

Samtidig viser data knyttet til studenters *prestasjoner* at evalueringer også har konsekvenser her. Beerkens (2018) viser til at man har målt mindre dropout, høyere gjennomføringsgrad, bedre avgangskarakterer og at studentene gjør det bedre i arbeidslivet i systemer der studieresultater er en del av kvalitetsevalueringen. Som vi kommer tilbake til er dette imidlertid indikatorer basert på *resultater*, mens andre igjen ønsker å fokusere på transformasjon og *prosess i personer*, noe det er langt vanskeligere både å måle og sammenligne. Harvey (2005) viser også hvordan denne tilnærmingen i Storbritannia kan knyttes til strategier og reformer fra sentralt hold, med fokus på effektivisering, innovasjon og markedsretting av UH-sektoren.

Kvalitet, nivå og evalueringsobjekt

Ashwin (2020) peker på at det er et grunnleggende problem med evaluering at hva som er *kvalitet* endres når man går fra *individ*, til *institusjon*, og så til *systemnivå*. Vi går ikke her dypt inn i definisjonen av kvalitet, men i diskusjonen av øvrige dimensjoner (formativ/summativ, kvalitativ/kvantitativ) blir det også tydelig at en underliggende og grunnleggende utfordring er å enes om hva kvalitet er, eller i det minste *skal forstås som*. Samtidig er de tre nevnte nivåene sentrale i utviklingen av evalueringssystemer, og disse kan også ha svært forskjellige oppfatninger av både formål med UH-sektoren og hva kvalitet er (Barnett, 1994).

Nivået som oftest brukes for sammenligning er *emne-* eller *programnivå*, da institusjonsnivå gjerne blir for overordnet og medfører at viktig innsikt mistes idet resultatene fra gode og dårlige programmer aggregeres (Ashwin, 2020).

Evalueringssystemene for utdanning er imidlertid svært sammensatte og kompliserte, og vi ser en kombinasjon av tilnærminger i de forskjellige landene. Alle landene vi har undersøkt her har en form for institusjonsevaluering som ser på institusjonens kvalitets-sikringssystemer som helhet. Disse evalueringene gjøres imidlertid ofte i kombinasjon med mer spesifikke evalueringer av fagområder eller temaer.

Som nevnt har for eksempel Nederland et system som kombinerer evalueringer av institusjonene som helhet og akkreditering av hvert enkelt studieprogram. Institusjoner som består institusjonsevalueringen, gjennomgår imidlertid mindre omfattende vurderinger når det kommer til akkreditering. Finland og Nederland har også systemer som kombinerer ulike typer evalueringer av ulike nivåer (institusjon, program og/eller tema). Storbritannia er det landet som i størst grad fokuserer på det institusjonelle nivået, selv om dette også kombineres med akkreditering av studieprogram.

Det særegne og det universelle

Som omtalt i forbindelse med forskningsevalueringer vil det også her måtte gjøres avveininger knyttet til vektleggingen av det *spesielle* eller det *universelle*. Ønsket om å *forstå* vil i noen tilfeller medføre en kostnad i form av begrensede sammenligningsmuligheter, siden en tilnærming basert på inngående kunnskap, der man tar hensyn til det særegne ved evaluanden, innebærer en aksept for at institusjonenes (eller programmenes) premisser er noe ulike. En slik tilnærming samsvarer med Ashwin's (2020) syn på utdanning som noe *lokalt*, som er rettet mot konkrete og spesielle personer i en spesiell kontekst. Hva som er kvalitet, og hvordan man oppnår det, vil ut fra et slikt perspektiv forandres med situasjon og kontekst (Ashwin, 2020).

I Storbritannia ser vi generelt at de har et relativt sterkt fokus på å sikre sammenlignbarhet mellom institusjoner, der alle institusjoner blir evaluert etter samme standard (Harvey, 2005). Som vi kommer tilbake til nedenfor har dette klare fordeler, men også tydelig ulemper – det legger kort sagt klare føringer på hva evalueringene kan brukes til, og hvem de er til for. Samtidig er det viktig å påpeke at systemene for utdannings-evaluering er særdeles sammensatte, og man har for eksempel også i Storbritannia muligheten for å gjøre *dypdykk* i spesielle områder etter avtaler med institusjoner (Harvey, 2005).

I Finland har de, i tillegg til felles standarder for evaluering av institusjonene, også åpnet opp for litt mer fokus på det lokale og særegne. Dette gjøres ved at institusjonene selv kan velge et evalueringsområde. Dette området får ikke konsekvenser for om institusjonene består evalueringen eller ikke, men det kan gi institusjonene en tilbakemelding på et område de selv anser som viktig.

Omfang og kostnader

Det er også en kjensgjerning at evaluering er *kostbart*, noe som fører til en nødvendig avveining mellom kost og nytte. Evaluatør må vurdere om et så komplett sett med data som mulig skal innsamles, til store kostnader, eller om man baserer seg på noe grovere indikatorer eller søken etter det Ashwin (2020) omtaler som the «silver bullet».

Ved omfattende og kostbare evalueringer oppstår spørsmålet om alternativkostnader (Harvey, 2005). Evalueringer er *ett* kvalitetsvirkemiddel, men med begrensede ressurser må beslutningstakere vurdere om ressursene som brukes på evalueringer fører til større kvalitetsforbedringer enn om ressursene hadde blitt brukt på andre tiltak.

Storbritannia omtales tidvis som «the British quality juggernaut», og som på forsknings-evalueringssiden er også utdanningsevalueringene svært omfattende og kostbare (Harvey, 2005). Dette medfører ikke bare direkte kostnader til eksterne evaluatører, for eksempel, men også en omfattende byrde for de som arbeider i sektoren (Harvey, 2005).

Kvalitative og kvantitative tilnærminger

Stensaker (2020) omtaler *måleproblemet* som en av utfordringene knyttet til evaluering av utdanningskvalitet. Indikatorer er gjerne knyttet til *prosesser* (for eksempel tilbakemeldinger og dialog) som vi ikke har *prosessedata* på, sier han, og vi baserer oss i stedet på informasjon om inntaks- og uttakskvalitet, for eksempel. Kort fortalt bestemmes ofte kvantitative mål og indikatorer på bakgrunn av *tilgjengelighet*, og ikke nødvendigvis det man egentlig ønsker å måle eller si noe om.

Et annet problem med bruk av indikatorer er knyttet til «Goodheart's lov», som omhandler strategiske hensyn institusjoner tar for å tilpasse sin virksomhet til de målene man styrer etter. Strathern (1997) skriver: «When a measure becomes a target, it ceases to be a good measure». Denne formuleringen beskriver en forveksling som har enda lettere for å oppstå på norsk, i og med at ordet «mål» brukes både i betydningen «target» og «measure». Om vi nyanserer ved å bruke ordene «målsetning» og «målestokk», kan loven oversettes slik: når en *målestokk* blir det man *styrer etter* (målsetningen), er den ikke lenger noen god *målestokk* på hvor godt det styres.

I tillegg vil en kvantitativ tilnærming ofte medføre en presisjon som ikke gjenspeiler virkeligheten. Ifølge Ashwin (2020) arbeider man med «slegger, ikke laser» innen denne type evaluering, og selv om en indikator tilsynelatende gir svært god presisjon, er dette altså allikevel forbundet med stor usikkerhet, og til dels lav validitet.

Beerkens (2018) trekker frem Storbritannia som et land som i stor grad baserer seg på kvantitative mål og «performance metrics». Veien utenom disse utfordringene vil da innebære mer bruk av kvalitative metoder og for eksempel fagfelleevalueringer og site visits.

Kvalitetssikring og akkreditering

Det finnes mange forskjellige metoder til kvalitetsevalueringer (Barnett, 1994). I tillegg finnes flere forskjellige grunnleggende tilnærminger, som akkreditering, tilsyn og mer tradisjonelle evalueringer lik de vi har diskutert på forskningssiden (Beerkens, 2018). Vi går ikke i dybden på forskjellen mellom metodene og tilnærmingene her, men vil påpeke at det ofte vil være behov for å *velge* tilnærming, da det kan være problematisk å kombinere.

En eklektisk tilnærming til metoder innebærer at beslutningstakere vil bruke *mange* metoder for å høste fordelene fra alle disse metodene. Dette, sier Barnett (1994), er en overfladisk tilnærming. Valg av tilnærming og metode er grunnleggende sett bestemt av hvordan vi forstår *kvalitet*, og forskjellige forståelser av kvalitet kan ikke uten videre

anses som komplementære – i noen tilfeller vil de være konkurrerende og gjensidig utelukkende (Barnett, 1994). Videre, dersom man forsøker å unngå en forutgående diskusjon om hva kvalitet er, og tenker at man kan kombinere metoder og på den måte dekke alle aspekter av kvalitet, står man i fare for å implisitt definere kvalitet ut fra et tilfeldig valg av metoder og de resultatene som følger av disse.

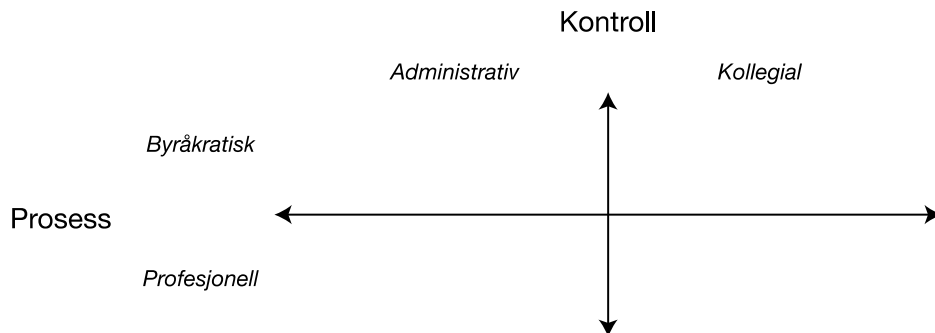
En sentral forskjell i tilnærming er hvorvidt man forsøker å gradere og presist vurdere evaluanden, eller om man baserer seg på minimumskrav. Sistnevnte tilnærming forbindes ofte med *akkreditering* som igjen ofte kritiseres nettopp for å fokusere på minimumskrav (og ikke kvalitetsforbedring) og at det nærmest nødvendigvis baseres på smale og (for) spesifikke kriterier uten å vurdere helhetlig (Stensaker & Harvey, 2006). Stensaker og Harvey (2006) ser nærmere på tre relevante akkrediteringssystemer, hvorav ett er i Nederland og to i Storbritannia. De omtaler i Nederland en utvikling fra *kvalitets-sikring* til *akkreditering* på programnivå, og akkrediteringen er basert på eksisterende kvalitetsvurderingssystemer. Her står institusjonen også fritt til å velge ekstern evaluator (blant godkjente organisasjoner), og systemet kjennetegnes av relativt høy grad av *fleksibilitet* noe som gjør det mulig å tilpasse vurderingene til de spesifikke behov (Stensaker & Harvey, 2006). De britiske systemene omtales som mindre fleksible, inspeksjon fremstår som viktigere enn fagfelleevaluering, og fokus er i større grad *compliance* enn kvalitetsutvikling (Stensaker & Harvey, 2006). Det er imidlertid viktig å bemerke at det finnes flere akkrediteringssystemer i hvert land, og disse kan arte seg svært forskjellig.

Hvem kontrollerer evalueringene?

Et annet aspekt som kan være av betydning er hvem som *har kontroll* over evalueringene (Barnett, 1994). Universiteter er *spesielle* hevdes det ofte, spesielt siden de er organisasjoner for kunnskapsutvikling, der formålet i stor grad er å fremskaffe ny kunnskap (Barnett, 1994). Ny kunnskap kan per definisjon ikke forstås eller kjennes på forhånd, og det er dermed umulig for en topp- og detaljstyrt organisasjon å styre strategisk mot dette ukjente. Derfor har fagmiljøer tradisjonelt hatt mye makt, og organisasjonen har vært relativt desentraliserte (Barnett, 1994).

Som Stensaker et al. (2011) og Beerkens (2018) viser har imidlertid nyere evalueringssystemer ført til en forskyvning av makt fra faglig til administrativ ledelse og fra desentralisert til sentralisert ledelse. I tillegg har det vært en utvikling fra institusjonell frihet til nasjonal kontroll, og Harvey (2005) knytter dette i Storbritannia til en manglende *tillit* mellom styresmakter og UH-sektoren. Storbritannia har imidlertid gjort noen endringer de siste årene blant annet ved at QAA har samarbeidet med sektoren og institusjonene om å utvikle en Quality Code for å sikre felles eierskap. Den samme endringen kan man se i Nederland der akkrediteringssystemet i sin nåværende fase har blitt endret til å fokusere mer på gjensidig tillit og ansvarliggjøring av institusjonene.

Til tross for at mange trekker frem negative aspekter ved Storbritannias evalueringssystem, spesielt knyttet til bruk av kvantitative metoder og fokus på det universelle, sier Barnett (1994) at dette systemet i stor grad «eies» og kontrolleres av akademikere selv. På den annen side, sier han, kan man ha evalueringssystemer som i langt større grad er styrt og bestemt av byråkrater på nasjonalt nivå (Barnett, 1994). *Kontroll* og *prosess/form* kan dermed skilles (Barnett, 1994). Statlig *kontrollerte* evalueringer kan utføres av fagfeller. Med en slik tilnærming kan evalueringssystemer plasseres i følgende landskap:



Figur 1: Vurdering av kontroll i evalueringssystemer (Barnett, 1994).

Et slikt rammeverk lar oss se mulighetene for administrativt kontrollerte systemer med sterke faglige innslag i prosess, og systemer med akademisk kontroll som benytter mer byråkratiske prosedyrer (Barnett, 1994).

Kollegial kontroll (akademisk kontroll) over evalueringer er ifølge Barnett (1994) basert på blant annet *eget ansvar, egevaluering, interne normer, ikke-eksplisitte vurderingskriterier, sensitivitet til kontekst, lukkethet og lite transparens*. I tillegg preges denne type kontroll av evalueringer med lav reliabilitet og høy validitet. Dette i motsetning til *byråkratisk* kontrollerte evalueringer, som blant annet er *system- og regelbasert, universelle, muliggjør sammenligning, basert på eksterne normer og høy reliabilitet og lav validitet* (Barnett, 1994). Til tross for at det finnes få *rene* systemer trekker Barnett her opp en rekke faktorer som kan være viktig i valget av grunnleggende tilnærming til kontroll over utdanningsevaluering. Det er også mulig å benytte dette analyseverktøyet *innad* i institusjoner (ikke bare på nasjonalt nivå), noe vi ikke går inn på her.

Storbritannia trekkes gjerne frem som et eksempel på et land med administrativ kontroll og profesjonell prosess. Her foretas evalueringene av evalueringsorganet QAA, som baserer seg på løpende indikatorer og institusjonenes egevalueringer. Alle de tre andre landene baserer seg på fagfelleevalueringer, både i institusjonsevalueringer og akkreditering. I Finland tar evalueringene i tillegg hensyn til institusjonenes unike ønsker og kontekst.

Formativ og summativ evaluering

Også i litteraturen om utdanningsevaluering er det mange som peker på den grunnleggende forskjellen mellom *summativ* og *formativ* evaluering (Barnett, 1994).

Barnett (1994) trekker denne distinksjonen videre, og knytter den summative tilnærmingen til *overvåkning* og den formative til *opplysning*. Dette er, sier han, en helt sentral dimensjon som kan brukes for å vurdere ethvert evalueringssystem. Stensaker et al. (2014) skiller på lignende måte mellom *kontroll* og *opplysning*.

Overvåkning knyttes dermed gjerne til behov for kontroll, og *accountability*. *Compliance* er også et begrep som ofte brukes i denne sammenheng (Harvey, 2005). Barnett (1994) kobler dette igjen til økende fokus på markedstenkning, der for eksempel kundetilfredshet og lignende begreper fra næringslivet får økende innpass også i det offentlige gjennom reformbølger som New Public Management. Dette gjenspeiles også i Harvey (2005), som igjen peker på hvordan økt fokus på *accountability* og *compliance* kan *true* institusjonell *autonomi* og *akademisk frihet*. Det finnes altså vidt forskjellige oppfatninger også av hva som sikrer autonomi og akademisk frihet. I en mer inngående studie av *makt* og *akkreditering* viser Engebretsen et al. (2012) hvordan forskjellige former for makt

er uløselig knyttet til evalueringssystemer. De skiller mellom to former for makt: overvåkning (basert på Foucaults *panopticon*) og «styrbarhet» (Foucaults *governmentality*).

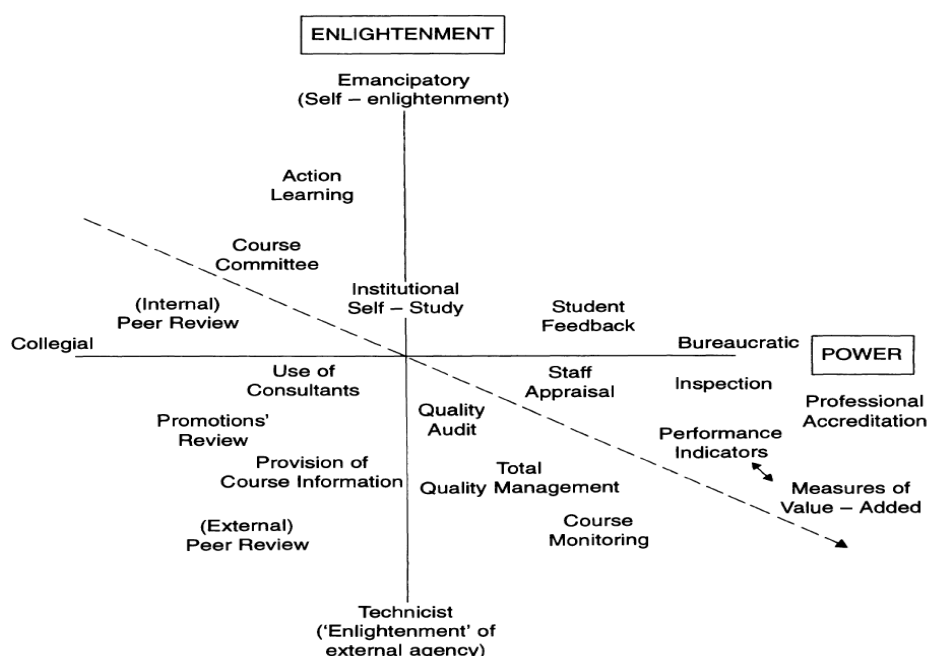
En *dobbel instrumentalisme* kan sees i evalueringer basert på kontroll og bruk av kvantitative mål og indikatorer, sier Barnett (1994). For det første innebærer det en *teknisk* tilnærming til vurdering av kvalitet gjennom kvantitative mål, og for det andre brukes evalueringene instrumentelt som et *styringsverktøy* med det formål å få UH-sektoren til å ligne mer på næringslivet.

Formativ evaluering, på den annen side, knyttes til behovet for å *forstå* (ikke *måle*), og Barnett (1994), peker på hermeneutikk og dialog, og tar oss dermed over til kvalitative metoder.

Kvalitetskontroll kan forstås som en summativ tilnærming, mens kvalitetsforbedring i større grad indikerer et formativt element (Barnett, 1994). Det er imidlertid uklart i hvilken grad Barnetts bruk av disse begrepene gjenspeiler dagens begrepsbruk. Engebretsen et al. (2012) påpeker også at både begrepet *kvalitet* og det som omtales som *kvalitetssikring* har vært gjenstand for endring over tid. Kvalitet, slik det intuitivt forstås i hverdagspråket, «tømmes for innhold» og erstattes med et nytt begrep som er tilpasset den logikken og de metoder som anvendes i evalueringssystemene (Engebretsen et al., 2012, s. 409).

Det er imidlertid mulig å skille mellom systemer som er «dømmende, retrospektive og formelle» (kvalitetskontroll) og de som er «ikke-dømmende, fremtidsrettede og relativt uformelle» (kvalitetsforbedring) (Barnett, 1994). Summative evalueringer som skal ansvarliggjøre er i stor grad rettet mot *eksterne*, mens formative evalueringer i større grad rettes mot selve evalueringobjektet. Et sentralt poeng for Barnett (1994) er at evaluering ikke kan, eller i det minste svært *vanskelig* kan, være en fullgod kombinasjon av begge tilnærminger.

Om denne dimensjonen sammenstilles med den foregående om *kontroll* og makt over evaluering, kan man også plassere evalueringer i følgende landskap der «opplysning» er én av aksene, og der opplysning av *evaluand* vs. *eksterne parter* er motpolene:



Figur 2: Utdanningsevaluering, fra Barnett (1994).

Storbritannia har det evalueringssystemet som ligger nærmest et summativt ideal, og forskningen viser også at det er lite belegg for at systemet fører til forbedringer i studentenes læringserfaringer (Harvey & Newton, 2004). Samtidig fremheves det svenske systemet som en kontrast til dette, der de i større grad har en tilnærming basert på «myk» støtte, og evalueringssystemet fungerer som en katalysator som har en rekke indirekte effekter, men lite fokus på direkte målbare resultater (Harvey & Newton, 2004). Ifølge Stensaker et al. (2011), er en nøkkelfunksjon ved evalueringer å «stimulere til endring og forbedring av undervisning og læring».

Om internasjonal sammenligning og internasjonalt samarbeid

Stensaker (2020) og Beerkens (2018) peker begge på en trend mot *evidensbasert* utdanningsevaluering. Det finnes etter hvert mye forskning på evaluering, og dette fører til bedre muligheter for å basere evalueringer på sikrere kunnskap enn man tidligere har hatt.

Evidensbasert evaluering fremstilles også tidvis som at evaluering er en sikker og universell vitenskap, og Stensaker (2020) anbefaler for eksempel mer samarbeid mellom forskjellige UH-systemer for å utvikle evalueringssystemer. Vi så imidlertid at blant andre Ashwin (2020) sterkt vektlegger det *lokale* aspektet ved undervisning. Beerkens (2018) påpeker også at det finnes flere eksempler på hvordan evalueringssystemer har blitt kopiert fra en kontekst til en annen med negative effekter. To viktige hensyn som må tas når man sammenligner slike systemer er om UH-systemet er *lite* eller *stort*, og om systemet har *mye* eller *lite* ressurser (Beerkens, 2018).

4 Referanser

- Ashwin, P. (2020, 4. mai). The challenges facing system-wide evaluations of educational quality. Foredrag under webinarer "Nokut webinar on Evaluation of Higher Education".
- Baccini, A., De Nicolao, G., & Petrovich, E. (2019). Citation gaming induced by bibliometric evaluation: A country-level comparative analysis. *PLoS One*, *14*(9), e0221212.
- Barnett, R. (1994). Power, enlightenment and quality evaluation. *European journal of education*, *29*(2), 165-179.
- Beerkens, M. (2018). Evidence-based policy and higher education quality assurance: progress, pitfalls and promise. *European Journal of Higher Education*, *8*(3), 272-287.
- Campbell, D. F. (2002). Conceptual framework for the evaluation of university research in Europe. *Center for International Science and Technology Policy (CISTP), The George Washington University, draft paper relativo al Panel "Evaluating university research, joint ventures and technology transfer*, *9*.
- Coryn, C. L. S., Hattie, J. A., Scriven, M., & Hartmann, D. J. (2007). Models and mechanisms for evaluating government-funded research: An international comparison. *American Journal of Evaluation*, *28*(4), 437–457.
- Engebretsen, E., Heggen, K., & Eilertsen, H. A. (2012). Accreditation and power: A discourse analysis of a new regime of governance in higher education. *Scandinavian Journal of Educational Research*, *56*(4), 401-417.
- Geuna, A., & Martin, B. R. (2001). University research evaluation and funding: An international comparison. SPRU Electronic Working Paper Series (Vol. 71). SPRU. <http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp71/sewp71.pdf>
- Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, *41*(4), 277–304.
- Harvey, L. (2005). A history and critique of quality evaluation in the UK. *Quality Assurance in Education*.
- Harvey, L., & Newton, J. (2004). Transforming quality evaluation. *Quality in higher education*, *10*(2), 149-165.
- Hicks, D. (2012). Performance-based university research funding systems, *Research Policy* *41*(2), 251–261.
- Jonkers, K., & Zacharewicz, T. (2016). Research Performance Based Funding Systems: a Comparative Assessment (No. EUR 27837 EN). European Commission.
- Ochsner, M., Kancewicz-Hoffman, N., Hołowiecki, M., & Holm, J. (Eds.). (2020). Overview of peer review practices in the SSH. ENRESSH Report. European Network of Research Evaluation in the Social Sciences and Humanities. <https://dx.doi.org/10.6084/m9.figshare.12032589>
- Ochsner, M. (2020). Place, role, form and significance of peer review in National Research Evaluation Systems. In M. Ochsner, N. Kancewicz-Hoffman, M. Hołowiecki, J. Holm (Eds.), Overview of peer review practices in the SSH. ENRESSH Report (pp. 10–14). European Network of Research Evaluation in the Social Sciences and Humanities. <https://dx.doi.org/10.6084/m9.figshare.12032589>

- Simon, D., & Knie, A. (2013). Can evaluation contribute to the organizational development of academic institutions? An international comparison. *Evaluation*, 19(4), 402-418.
- Sivertsen, G. (2017). Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications*, 3(1), 1-6.
- Stensaker (2020, 4. mai). A national evaluation framework for higher education and research – some reflections. Foredrag under webinarer “Nokut webinar on Evaluation of Higher Education”.
- Stensaker, B., & Harvey, L. (2006). Old wine in new bottles? A comparison of public and private accreditation schemes in higher education. *Higher Education Policy*, 19(1), 65-85.
- Stensaker, B., Langfeldt, L., Harvey, L., Huisman, J., & Westerheijden, D. (2011). An in-depth study on the impact of external quality assurance. *Assessment & Evaluation in Higher Education*, 36(4), 465-478.
- Strathern, M. (1997). ‘Improving ratings’: audit in the British University system. *European review*, 5(3), 305-321.
- von Tunzelmann, N., & Mbula, E. K. (2003). Changes in research assessment practices in other countries since 1999: Final report. <http://www.rareview.ac.uk/reports/Prac/ChangingPractices.pdf>
- Whitley, R., & Gläser, J. (Eds.). (2007). The changing governance of the sciences. The advent of research evaluation systems. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-6746-4>
- Whitley, R. (2007). Changing governance of the public sciences. I Whitley, R., & Gläser, J. (Eds). *The changing governance of the sciences. The advent of research evaluation systems*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-6746-4>



Vedlegg 2: Typologisk kartlegging av evalueringer

Notat: Prosjekt *Kartlegging av evalueringer*

Innhold

INNHold	1
1 INNLEDNING	1
2 TYPOLOGISERING – TEORI	2
2.1 Kombinasjon av ulike tilnærminger og formål	3
2.2 Fire basistyper	4
2.3 Blandede kategorier	6
3 FORKLARING AV TYPOLOGIEN	9
3.1 NFR-kategorisering	9
3.2 Formål (mandat)	9
3.3 Metoder brukt i evalueringer	10
3.4 Oppfølging av evalueringer	11
4 OPPSUMMERING	12
LITTERATURLISTE	14

1 Innledning

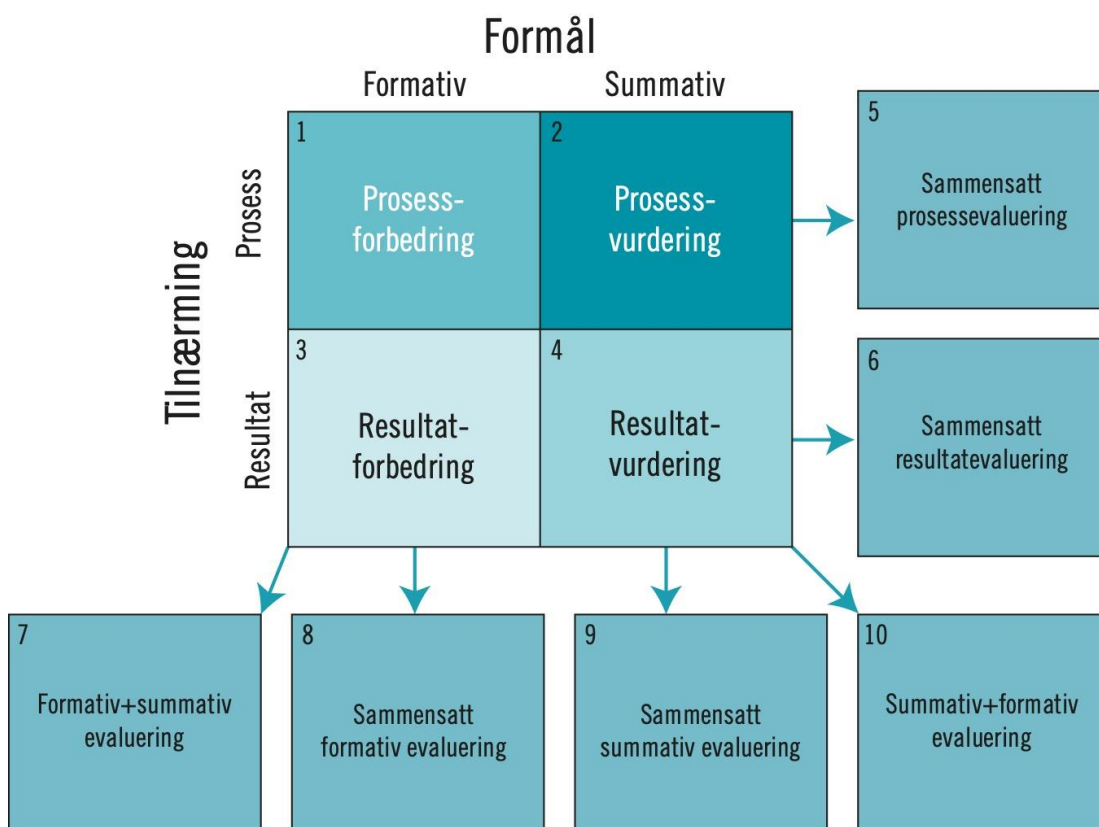
Som et ledd i prosjektet *Kartlegging av evalueringer* gjennomfører vi en typologisering av gjennomførte evalueringer av kvalitet i forskning og høyere utdanning. En typologi er et organisert system av *typer*, og de lar oss utvikle og videreutvikle begreper, identifisere underliggende dimensjoner, samt oppnå et uttalt mål i utlysningen: kategorisere og sortere (Collier et al., 2012). Vi gjennomfører en typologisering med mål om å a) utvikle et hensiktsmessig begrepsapparat, og b) identifisere sentrale dimensjoner ved gjennomførte evalueringer. Vi fokuserer på evalueringer på nasjonalt nivå gjennomført av NOKUT og NFR de siste 10 årene, men vi anser det også som hensiktsmessig å gå ut over denne rammen i noen tilfeller.

Vi legger vekt på å utvikle et begrepsapparat som egner seg for å kategorisere evalueringer av både forskning og høyere utdanning. Vi har tatt utgangspunkt i Chens (1996) typologi for programevaluering, og i dette notatet presenterer vi vår tentative kategorisering av nasjonale evalueringer av forskning og høyere utdanning. Vi kommenterer også kort forskningsrådets kategorier, og inkluderer disse i oversikten over gjennomførte evalueringer. I tillegg til å la oss kategorisere gjennomførte evalueringer på en hensiktsmessig måte, gir typologiene et nyttig begrepsapparat og en oversikt over hvilke forhold ved evalueringer som er sentrale. Vi vil også bemerke at vi ser denne typologiseringen i sammenheng med interessentanalysen vi gjennomfører. Typologiseringen bidrar delvis til å strukturere intervjuene, og funnene fra interessentanalysen vil kunne brukes til å videreutvikle typologiseringen i etterkant.

De fleste norske evalueringer har flere formål, og dette er en av hovedårsakene til at vi tar utgangspunkt i Chens (1996) typologi med *blandede* kategorier. Disse kategoriene er imidlertid ikke fullt ut utviklet av Chen (1996), og vi anbefaler at interessentanalysen følges av en videreutvikling av disse kategoriene og en bedre forståelse av hvordan forskjellige tilnærminger og formål kan *integreres* i evalueringer.

2 Typologisering – teori

Vi utvikler en typologi som både sier noe om hva man forsøker å oppnå med evalueringer (for eksempel institusjonsforbedring, kontroll, informasjon til samfunnet) og hvordan man går frem for å gjøre dette (for eksempel følgevalueringer over tid med fokus på *prosesser* eller tverrsnittundersøkelser fokusert på *mål* og indikatorer). Chen (1996) skiller mellom evaluering av *mål* og *prosess*, og *summativ* og *formativ* evaluering, avhengig av om formålet er å vurdere en tilstand, eller å forbedre den. Dette er de to hoveddimensjonene vi baserer oss på i arbeidet med en typologi, som vises i figur 1:



Figur 1. Chen's typologi for programevaluering. Kategori 1-4 beskrives av Chen (1996) som basisformer for evaluering, mens kategori 5-10 er blandingsformer for evaluering.

Den teoretiske bakgrunnen for typologiseringen er Chen (1996) sin videreutvikling av den klassiske inndelingen mellom formativ og summativ evaluering (Scriven, 1991). Formativ evaluering er en tilbakemelding ment for å forbedre evaluanden (det som evalueres). Summativ evaluering er en evaluering med fokus på *utfall* eller *produksjon*, og kan samtidig gi tilbakemelding på hvorvidt for eksempel et program bør fortsette eller ikke. Et problem med en slik inndeling er at evalueringer ikke nødvendigvis er *enten* formative eller summative. Ett eksempel på dette er et internasjonalt ekspertpanels vurdering av forskningsproduksjon. Dette er i utgangspunktet en summativ evaluering, dersom de ser på sluttproduktet, det vil si produksjonen, til en institusjon. Men, dersom samme ekspertgruppe i tillegg kommer med endringsforslag som kan bidra til å øke produksjon av forskning i høy internasjonal klasse, vil evalueringen også ha formative trekk. Evalueringene beskrevet i denne typologiseringen er i hovedsak en blanding av formativ og summativ form, noe vi kommer tilbake til når vi diskuterer blandede kategorier.

Den andre dimensjonen handler om hvorvidt evaluator fokuserer på *prosess* (*process*) eller *resultater* (*outcome*). Vi velger å kalle denne dimensjonen *tilnærming*, mens Chen (1996) omtaler dette som *program stage*. Dette er et naturlig resultat av at Chen i hovedsak omtaler *programevaluering*, mens vi i hovedsak omtaler evalueringer av kontinuerlige prosesser.

I spørsmål knyttet til hvorvidt evalueringene er *formative* eller *summative* er det nødvendig å identifisere hvilket perspektiv vi inntar. Vi velger her å fokusere på *evaluandens* perspektiv, og ikke perspektivet til evaluator eller bestiller av evalueringene. En evaluering av et universitet kan være summativ sett fra institusjonens side, men den kan også være et element i en *formativ* evaluering av nasjonal forskningspolitikk.

I vårt arbeid med å kategorisere evalueringene, har vi i utgangspunktet basert oss på mandatet til evalueringskomiteen, i tillegg til oppsummering/sammendrag og konklusjon. Her har vi funnet informasjon om både hva som er formålet med evalueringen og hvilke elementer ved forskning eller utdanning som faktisk er evaluert. I evalueringene gjennomført av NOKUT er formålet med evalueringen og hva resultatene skal brukes til ofte beskrevet enten i konklusjonen, i et eget kapittel med videre anbefalinger eller i egne institusjonsrapporter. Alle evalueringer inkludert i vedlegg 2a og 2b er først grovkategorisert, før vi finkategoriserte gjennom identifisering av likhetstrekk mellom evalueringer.

2.1 Kombinasjon av ulike tilnærminger og formål

Skillet mellom formative og summative evalueringer er sjelden helt tydelig i praksis, og i de fleste evalueringer finner vi elementer av begge formål. Kombinasjonen av de to formålene er imidlertid ikke uproblematisk, da evaluanden til en viss grad må antas å forholde seg strategisk til evalueringene som gjennomføres. Dersom en evaluering oppfattes som summativ kan det tenkes at evaluanden er mer tilbakeholden med hva slags informasjon som deles. Dette blir spesielt problematisk dersom evalueringen i stor grad avhenger av evaluandens egen innrapportering av fagmiljøer og forskergrupper som skal inngå i evalueringen. Videre er dette også problematisk dersom det er tydelige elementer av egevaluering som ligger til grunn for vurderingene som gjøres.

Her er det imidlertid verdt å vurdere om det er evalueringens formål eller den potensielle anvendelsen av evalueringene som veier tyngst. En summativ vurdering kan være rettet mot evaluandens behov for informasjon om egne resultater, sin posisjon målt opp mot sammenlignbare institusjoner, e.l. Dersom en slik evaluering ikke offentliggjøres, vil den allikevel kunne oppfattes som både nyttig og ufarlig. På den annen side vil også formative evalueringer som peker på forbedringspunkter nødvendigvis også kunne brukes til å rette kritikk og oppfølgingstiltak mot en institusjon i etterkant. Det er altså ikke gitt at det er evalueringens formål som skaper problemer knyttet til åpenhet og tilgang til gode og realistiske data fra evaluanden; hvordan evalueringene kan brukes av andre aktører fremstår som sentralt.

Denne problemstillingen diskuteres også i øvrige sammenhenger der man skiller mellom formativ og summativ vurdering (for eksempel innenfor pedagogikk og læring), og også i forskningen på evalueringer (da gjerne knyttet til begrepene ansvarliggjøring og læring). Dette er evalueringens *double formål*, som omtales i forbindelse med bistandsevalueringer (Reinertsen, Bjørkdahl & McNeill, 2017).

Denne type problemstillinger vil bli gjenstand for undersøkelse i interessentanalysen som følger denne typologiseringen.

Videre gir vår typologi muligheten til å analysere den praktiske gjennomføringen av blandingsevalueringer. Når man gjennomfører blandede evalueringer er det nødvendig å integrere basistypene som kombineres på en systematisk måte, og dette kan gjøres ved strukturell integrering eller sekvensiell integrering. Strukturell integrering omfatter metoder og mekanismer for å forbinde de forskjellige dimensjonene, og Chen (1996) omtaler forskjellige måter dette kan gjøres på, som for eksempel *dimension elaboration* og *causal elaboration*. Førstnevnte innebærer for eksempel at informasjonen man innhenter i en summativ vurdering utbroderes i tilstrekkelig detalj til at den samtidig kan brukes til formativ vurdering. Causal elaboration innebærer at man søker å avdekke kausalmekanismer, og forhold mellom avhengige, mellomliggende og uavhengige variabler. På den måten får man en dypere forståelse av *hvorfor* et program har, eller ikke har, effekt, og dermed også bedre muligheter til å knytte *prosess* til resultater, og *summative* til formative formål. For å oppnå dette anvendes ofte *process tracing* som metode, kombinert med et fokus på *sosiale mekanismer* (Collier, 2011; Bennett & Checkel, 2015; Steel, 2004; Hedström & Ylikoski, 2010). Sekvensiell integrering innebærer at de forskjellige elementene gjennomføres etter tur.

2.2 Fire basistyper

Vi omtaler de fire grunnkombinasjonene av de to mulige verdiene på hver hoveddimensjon som *basistyper* i vår typologi.

Kategori 1 – Prosessforbedring

Evaluering med fokus på prosessforbedring er definert av Chen (1996) som «evaluering ment for å gi informasjon for programforbedring gjennom å analysere styrker og svakheter med implementeringen av programmet».

Eksempel 1: Evaluation of Norwegian education research (NFR, 2018c)

Evalueringen av norsk utdanningsforskning er delt inn i kvalitet og kapasitet i forskning, strategisk fokus og samarbeid, og relevans og samfunnsnytte. Evalueringen kommer med anbefalinger til både NFR og institusjonene som er evaluert. Anbefalingene til institusjonene omhandler bedre samarbeid, både nasjonalt og internasjonalt, og mer fokus på forskning av høy kvalitet.

Eksempel 2: Educational quality in economics, political science and sociology in Norway (NOKUT, 2018)

Her evalueres studieprogrammene, inkludert hvordan de er organisert internt, hvordan pensum er satt sammen og hvordan undervisningen foregår, med det formål å finne ut av hva som fungerer og hva som kan forbedres. Her fikk evalueringspanelene instruks om å fokusere på implementeringen av utdanningsprogrammene, og ikke studentenes resultater eller gjennomføring, og evalueringene kan dermed kategoriseres som prosessforbedrende.

Kategori 2 – Prosessvurdering

Evaluering av prosess er ikke begrenset til formative evalueringer i form av prosessforbedring, men kan også utføres som en prosessvurdering. Chen (1996) nevner *kvalitetskontroll* som et eksempel på prosessvurdering, og det er også nærliggende å anse tilsynsvirksomhet i form av en evaluering av om kvalitetssystemer tilfredsstillende oppgitte kriterier som en *prosessvurdering*. Formålet med en prosessvurdering vil være å undersøke om for eksempel kontinuerlige prosesser, som institusjonens kvalitetssystemer, er *hensiktsmessige*, tilfredsstillende oppgitte kriterier, og er implementert på en hensiktsmessig måte.

Eksempel 1: Evaluering av filosofi og vitenskapshistorie (NFR, 2010a)

Så vidt vi kan bedømme, er det ikke gjennomført noen evalueringer som er rene prosessvurderinger. Det er likevel flere evalueringer med elementer av prosessvurdering. En av disse er evaluering av filosofi og vitenskapshistorie fra 2010. I evalueringen blir det blant annet pekt på at det i liten grad er samarbeid på tvers av de norske instituttene, samt at det er liten grad av mobilitet. Dette kan forstås som en vurdering av prosessene forbundet med forskningsproduksjon, men panelene kommer ikke med noen klare anbefaling om *hvordan* disse prosessene skal forbedres. Hvis mobilitet i seg selv er et ønsket utfall, og ikke et ledd i å forbedre forskningskvalitet, vil måling av mobilitet være en resultatvurdering.

Kategori 3 – Resultatforbedring

Tilsvarende distinksjon gjelder når vi ser på resultat. Evalueringer av resultat kan både ha formål om resultatforbedring og resultatvurdering. Resultatforbedringer har som mål å undersøke hvilke elementer i et program som bidrar til, eller eventuelt hemmer, gode resultater. Med slike evalueringer søker evaluatoren å avdekke hva som er viktig for å oppnå et godt resultat. For eksempel kan dette være å evaluere hva som fungerer og ikke fungerer ved en institusjon for å oppnå et mål om et visst antall publikasjoner på høyt internasjonalt nivå.

Eksempel 1: Evaluering av forskning innenfor de matematiske fagene i et internasjonalt perspektiv (NFR, 2012)

I NFR sin fagevaluering fra 2012, var ett av målene å identifisere forskningsfelt med behov for å bli faglig forbedret slik at kompetansen på området ble opprettholdt på et akseptabelt nivå. Utvalget kommer med anbefalinger som peker på endringer som må gjøres for å opprettholde et akseptabelt nivå på forskningen gjennomført blant de forskjellige institusjonene. Ett eksempel er at utvalget påpeker at noen matematiske institutt i liten grad samarbeider med andre institutt, og at dette bør forbedres for å øke potensialet for publikasjoner i anerkjente internasjonale tidsskrifter.

Kategori 4 – Resultatvurdering

Resultatvurderinger tilsvarer i stor grad den klassiske summative evalueringen, hvor målet er å evaluere om et program har oppnådd målene sine. Dette kan for eksempel være en evaluering som utelukkende vurderer antall forskningsartikler publisert i journaler på et visst nivå.

Eksempel 1: Evaluering av sosiologisk forskning i Norge (NFR, 2010c)

NFRs fagevalueringer fra 2010 fokuserte mye på resultatvurdering av publisert forskning i et internasjonalt perspektiv. Formålet med evalueringen var blant annet å vurdere forskningskvaliteten og relevansen av forskningen samt publiseringsaktiviteten. Evalueringen fant at noen forskere og forskningsgrupper publiserte på høyt internasjonalt nivå, men at dette varierte både mellom institusjoner og mellom forskere på de forskjellige institusjonene.

Eksempel 2: Studiebarometeret (NOKUT, årlig)

NOKUTs årlige Studiebarometer er summativ av natur og et godt eksempel på en resultatvurdering. Her skal studentene (som brukere) vurdere studieprogrammet sitt og i hvilken grad det lever opp til deres krav og forventninger. Selv om Studiebarometeret ofte brukes videre av NOKUT eller institusjonene til utviklingsarbeid, eller som en del av datagrunnlaget i NFRs evalueringer, er selve undersøkelsen i seg selv summativ. Det er ikke en evaluering i streng forstand, ettersom det kun er en spørreundersøkelse uten analyse i etterkant. Likevel er det viktig for mange av institusjonene, og siden de oppfatter Studiebarometeret som en del av NOKUTs helhetlige evalueringsarbeid, velger vi å bruke det som et eksempel.

2.3 Blandede kategorier

De fire basistypene er et godt utgangspunkt for kategorisering, men vi ønsker å utvikle en noe mer finmasket typologi enn dette, siden evalueringer av den typen vi her undersøker ofte er omfattende, sammensatte og har flere formål. De *blandede* kategoriene er spesielt relevante for utviklingen av en typologi for evalueringer av kvalitet i forskning og utdanning. Chen (1996) påpeker at disse er spesielt hensiktsmessige når a) det er gjentatte evalueringer med et ønske om å gi hyppige tilbakemeldinger til interessenter, b) når evalueringene skal tilfredsstillende behovene til forskjellige interessenter med forskjellige behov, og c) når interessentene har behov for *omfattende* informasjon om evaluanden. Alle disse forholdene gjør seg sterkt gjeldende i evalueringer av kvalitet i forskning og høyere utdanning i Norge. SAMEVAL har for eksempel som mål å vurdere tilstand, utforme anbefalinger til institusjonene, gi et grunnlag for videreutvikling av forskningspolitikk, samt forbedre grunnlaget for forskningsrådets finansieringsordninger (NFR, 2018).

Kategori 5 – Sammensatt prosessevaluering

Chens inndeling viser at prosessevaluering kan være en kombinasjon av prosessforbedring og prosessvurdering i form av en sammensatt prosessvurdering. Dette er altså en evaluering av implementeringen av et program der man både vurderer om implementeringen har vært vellykket eller ikke, og kommer med forslag til forbedringer. Denne type kombinasjon er ofte basert på *dimension elaboration*, ved at man for eksempel summativt vurderer prosessene i tilstrekkelig dybde til at man også kan bruke dette til å komme med anbefalinger om forbedringer.

Eksempel 1: NOKUTs tilsynsrapporter

Et eksempel på slike sammensatte prosessevalueringer kan være NOKUTs evalueringer av institusjonenes egne kvalitetssikringssystemer. Her skal de sakkyndige komiteene først evaluere om institusjonene tilfredsstillende kravene til kvalitetssikring av egne studietilbud. Dersom dette ikke er tilfellet, kommer NOKUT med anbefalinger og institusjonene må gjøre forbedringer før NOKUT så gjør en ny evaluering. Dette er altså en kombinasjon av en prosessvurdering og prosessforbedring. På samme måte som Studiebarometeret, regner ikke NOKUT tilsynsaktiviteten sin som en form for evaluering. Det er imidlertid en del av det helhetlige bildet for institusjonene, som ofte forbereder seg til tilsyn på mye av den samme måten som de forbereder seg til andre typer evalueringer.

Kategori 6 – Sammensatt resultatevaluering

Man kan også foreta evalueringer som både har som formål å forbedre og vurdere resultatene av et program, altså sammensatte resultatevalueringer. Her ser man på resultatene av et program, og evaluerer både om resultatene er gode nok og hva som eventuelt kan bidra til å forbedre resultatene.

Eksempel 1: Evaluering av primærnæringsinstituttene (NFR, 2018d)

Evalueringen bærer preg av å være rettet mot resultat. På den ene siden anbefales tiltak som skal forbedre resultatene, målt i både forskningsproduksjon og samfunnsrelevans. Utvalget anbefaler blant annet at samfunnsrelevansen skal bestå av å «bidra med langsiktig behov for fremtidig omstilling av norsk økonomi med utgangspunkt i primærnæringssektoren». Det vil si plassere egen kompetanse inn i tenkningen rundt store samfunnsutfordringer. I tillegg kommer utvalget med anbefalinger om å opprettholde og styrke den vitenskapelige publiseringen. Evalueringen er også en resultatvurdering gjennom diskusjon av samfunnseffekter, for eksempel interaksjon med forvaltning, næringsliv og politikk. Det vises blant annet til forskning på fiskehelse som har bidratt til å løse miljøproblemer, bedre fiskevelferd og lagt videre grunnlag for verdiskaping i havbruksnæringssektoren. I tillegg er instituttene publisering og sitering vurdert.

Kategori 7 – Formativ + summativ evaluering

Evalueringer kan bestå av både formative og summative elementer. Chen (1996) gir ikke en utfyllende forklaring på blandingstypene av evaluering, men vi tolker kategori 7 dithen at evalueringen har en hovedtyngde på formativ evaluering, men også innslag av summativ evaluering.

Kategori 8 – Sammensatt formativ evaluering

En sammensatt formativ evaluering fokuserer både på prosessforbedring og resultatforbedring, i tillegg en kombinasjon av disse.

Eksempel 1: Fagevalueringer: allmennlærer-, ingeniør- og førskolelærerutdanningen (NOKUT, 2006, 2008, 2010)

NOKUTs tidligere fagevalueringer, for eksempel evalueringene av allmennlærerutdanningen, ingeniørutdanningen og førskolelærerutdanningen i henholdsvis 2006, 2008 og 2010, er eksempler på sammensatte formative evalueringer. Her så panelene både på organiseringen av utdanningsprogrammene og på studentenes resultater og grad av gjennomføring for å finne ut hva som fungerte og hva som burde forbedres.

Kategori 9 – Sammensatt summativ evaluering

En sammensatt summativ evaluering inneholder vurdering av både prosess og resultat, i tillegg til en kombinasjon av disse. Her evaluerer man altså både om implementeringen har vært vellykket og om man har oppnådd ønskede resultater.

Eksempel 1: Panelevalueringer under SAMEVAL (NFR, 2018b)

Rapportene fra de seks fagpanelene under SAMEVAL består i stor grad av en kombinasjon av prosessvurdering og resultatvurdering. Blant annet beskriver utvalgte forskningsgrupper under forskjellige institutt i UH-sektoren og evaluerer deres samlede forskningsresultater. I tillegg er det lagt stor vekt på bibliometriske analyser og siteringsanalyser for å beskrive enkelte institutter eller institusjoner sin samlede forskningsproduksjon.

Kategori 10 – Summativ + formativ evaluering

Denne kategorien, i likhet med kategori 7, inneholder både summative og formative elementer. Her har imidlertid evalueringen en hovedtyngde på summativ evaluering, men den inneholder også formative elementer.

Eksempel 1: Evaluation of Social Sciences in Norway (NFR, 2018a)

SAMEVAL er et eksempel på en evaluering som faller under denne kategorien. Hovedrapporten oppsummerer funnene fra de forskjellige panelene og gir en helhetlig samlet vurdering. Det er lagt stor vekt på resultatvurdering i form av bibliometriske analyser og siteringsanalyse. Videre vurderes sammensetningen av forskergrupper, forskerutdanning (PhD-utdanning), og undervisning, altså prosessvurdering. Hovedrapporten gir flere tilrådninger til institusjonene om sammensetning på lokalt nivå, både til institutt- og til UH-sektoren (prosessforbedring). Basert på tilrådingene mener evalueringskomiteen at både volumet og kvaliteten på forskningen kan forbedres, internasjonaliseringen forbedres, og mobiliteten blant forskere økes (resultatforbedring).

En del av SAMEVAL skal; «*gi innsikt, råd og anbefalinger for institusjonene evaluert som kan bli brukt til å øke forskningsstandarden*». Eksempel på råd og anbefalinger kan være endring i sammensetningen av forskergrupper, faglig eller metodisk fokus eller rekrutteringsstrategi, det vil si prosessforbedring. Disse må sees i sammenheng med råd og anbefalinger for resultatforbedring, det vil si antall publikasjoner og kvalitet på publikasjoner, i tillegg til samfunnsrelevans og samfunnsnytte.

3 Forklaring av typologien

I vår typologisering av gjennomførte evalueringer av kvalitet i forskning og høyere utdanning, har vi i tillegg til Chen-kategoriene også inkludert noen andre faktorer. For det første har vi inkludert NFRs egne evalueringskategorier. I tillegg har vi sett på formålet/mandatet i de ulike evalueringene, samt hvilke metoder som er brukt. Alle disse momentene er inkludert i de vedlagte tabellene (vedlegg 2a og 2b), der vi har kategorisert alle evalueringer gjennomført av NOKUT og NFR de 10 siste årene. I dette underkapittelet forklarer vi de ulike delene av tabellen og drøfter noen funn og utviklingstrekk.

3.1 NFR-kategorisering

NFR deler sine evalueringer inn i sju forskjellige kategorier; *fagevalueringer*, *tematiske evalueringer*, *instituttevalueringer*, *virkemiddelevueringer*, *senterevalueringer*, *effekt-evalueringer*, og *forskningsbaserte evalueringssoppdrag*.¹

I denne typologiseringen baserer vi oss på fagevalueringer, tematiske evalueringer, og instituttevalueringer, ettersom vi oppfatter at det er disse evalueringene som mest direkte retter seg mot forskningskvalitet. NFR har ikke eksplisitt kategorisert noen evalueringer under effektevalueringer, men beskriver formålet med effektevalueringer som å øke «*forståelsen av forholdet mellom satsningen og dens effekter på forskning og samfunnet for øvrig, og en skiller mellom resultater (output), virkninger (outcome) og effekter (impact)*». NFR skriver videre at dette kan inngå som en del av de andre evalueringene, og vi har ikke funnet noen evalueringer som eksplisitt omtales som en effektevaluering. NFR gjennomfører også virkemiddelevueringer, som kan være følge-, underveis-, og effektevalueringer, avhengig av formålet med evalueringen. Virkemidler kan for eksempel være Senter for Fremragende Forskning (SFF)². Det henvises i stor grad til fag- eller temaevalueringer i denne formen for evalueringer. I analyse av Forskningsrådets programmer for velferd og arbeidsliv fra 2000-2014 (NIFU, 2019) er det for eksempel blant annet vurdert hvordan forskningen spres, brukes og kommer til nytte blant brukere og i samfunnet for øvrig. Disse vurderingene er basert på evalueringene SAMEVAL, HUMEVAL, og evaluering av evalueringen av de samfunnsvitenskapelige instituttene.

Vi anser derfor at fagevalueringer, tematiske evalueringer og instituttevalueringer er relevante for denne typologiseringen, og utelater de øvrige. Fagevalueringer omfatter forskning innen et definert fag eller disiplin, tematiske evalueringer går på tvers av fagområder, mens instituttevalueringer er evaluering av forskningsinstitutter som mottar basisfinansiering av NFR. Vi har inkludert NFR-kategoriene i typologiseringen, og de vurderer dem som et supplement til Chen-kategoriseringen, ved at de gir informasjon om hva som er evalueringsobjektet.

3.2 Formål (mandat)

I typologiseringen baserer vi oss på formålsbeskrivelsen i mandatene til evalueringene vi vurderer. Dette er stort sett en punktvis liste som oppsummerer de viktigste vurderingskriteriene for utvalget. Det er likevel slik at en typologisering av denne typen ikke kan gi et fullstendig bilde av hele formålet med en evaluering. Det kan være nyanser i formålet, mandatet, eller presisering av mandatet, som ikke blir beskrevet i typologiseringen som kan gi et skjevt bilde av utviklingen i evalueringer.

¹ Se <https://www.forskningsradet.no/Statistikk-og-evalueringer/evalueringer/vare-evalueringer/> for NFR sin oversikt over NFR sin kategorisering av evalueringer

² Finansieringsvirkemiddel for å fremme fremragende forskning.

I evalueringen av fysikk fra 2010 (NFR 2010b) deles mandatet inn i en overordnet vurdering, vurdering på instituttnivå, vurdering av forskningsgrupper, finansiering og organisering, og forskerutdanning. I tillegg presiserer flere evalueringer, som evalueringen av de 12 regionale forskningsinstituttene, at mandatet er *vidt*, slik at det åpner for avgrensninger og presiseringer av hovedvekten i evalueringen.

3.3 Metoder brukt i evalueringer

Med metode menes bakgrunns materialet som evalueringen bygges på. Metodene har i stor grad vært like siden 2010, men det er noen nyanser som likevel bør fremheves. Under beskriver vi noen forskjeller, basert på tre utvalgte evalueringer, gjennomført i hhv. 2010, 2015 og 2018.

Faktarapport

I flere evalueringer, blant annet Evaluering av Sosiologisk forskning fra 2010 (NFR, 2010c), tinget NFR en bakgrunnsrapport/faktarapport utarbeidet av NIFU. Bakgrunnsrapporten inneholder informasjon om ansatte, finansiell situasjon og internasjonal publisering relatert til norsk sosiologisk forskning. Tilsvarende rapport ble produsert for evalueringen av miljøinstituttene (NFR 2015a), som også inneholdt fakta om søknader og tilslag på søknader til NFR. I SAMEVAL (NFR, 2018a) fremgår det ikke tydelig om bakgrunnsrapporter/faktarapporter er inkludert i evalueringen, men det henvises til enkelte bakgrunnsdokumenter som ikke er publisert. Faktarapporter inneholder også tall fra FoU-statistikken og forskningsrådets årsrapporter.

Dokumentstudier

Avhengig av evalueringstype, er det flere former for bakgrunnsinformasjon som også blir inkludert i evalueringene. Dette kan være stortingsmeldinger, tidligere utredninger eller oppfølging av tidligere utredninger, andre relevante utredninger, og informasjon hentet fra institusjoners nettsider.

Egenevaluering

Egenevalueringer består av interne evalueringer fra de forskjellige instituttene og enhetene som er evaluert. I evaluering av Sosiologisk forskning fra 2010, inkluderte egnevalueringen en (kritisk) vurdering av egen situasjon, utfordringer, og planer for fremtiden. I tillegg skulle egnevalueringen inkludere forskningsprofil, produsert forskning, forskningssamarbeid, finansiering, forskerutdanning og rekruttering, og formidling og samfunnskontakt. Tilsvarende punkter ble brukt i evalueringen av miljøinstituttene i 2015, og i SAMEVAL i 2018.

Bibliometrisk analyse

Bibliometrisk analyse oppsummerer og teller antall publiseringer nasjonalt og internasjonalt, gjerne over de foregående fem årene. En slik analyse baserer seg stort sett på antall publiseringer og publiseringspoeng, eller antall publikasjoner fordelt i journaler på nivå 1 eller 2 (som i evalueringen av sosiologisk forskning fra 2010).

Bibliometrisk analyse er i evalueringene stort sett gjennomført av eksterne oppdragsgivere. Den bibliometriske analysen i SAMEVAL ble gjennomført av DAMVAD Analytics, mens NIFU har gjennomført flere, blant annet i forbindelse med Evaluering av Sosiologisk forskning fra 2010. Evalueringen av miljøinstituttene fra 2015 baserte seg ikke på bibliometriske analyser, men på kvalitetsvurdering av tilgjengelige data, tidligere fagevalueringer (der det er relevant) samt brukernes oppfatninger av kvaliteten på insti-

tuttenes forskning. Spesielt i tematiske evalueringer blir det presisert at det er utfordrende å sammenligne publisering på tvers av faggrupper. For eksempel nevnes det i evalueringen av klimaforskning fra 2015 (NFR, 2015b), at det er vesentlige forskjeller i publiseringstradisjoner mellom naturvitenskap, samfunnsvitenskap og humaniora.

Fra evaluering av Sosiologisk forskning i 2010, til SAMEVAL er, så vidt vi kan bedømme, den eneste vesentlige endringen i bibliometriske analyser at analysen inkluderer en «field normalized citation index», som er en vektning av faktiske siteringer til en enkel publikasjon, gitt gjennomsnittlig antall siteringer til alle andre lignende publikasjoner. Sistnevnte er det som kalles forventet antall siteringer. I evalueringen av Sosiologisk forskning i 2010 brukes siteringsdata utelukkende som bakgrunnsdata.

Intervju

I NFR-evalueringer blir det stort sett gjennomført intervjuer med representanter for evaluerte enheter, men noen evalueringer benytter seg også av gruppeintervjuer, supplert med enkeltintervjuer. I evaluering av Sosiologisk forskning fra 2010 og SAMEVAL intervjuet man representanter fra de fleste enhetene som ble evaluert, mens i evalueringen av miljøinstituttene fra 2015, benyttet man seg av gruppeintervjuer. I sistnevnte evaluering ble også informanter i NFR intervjuet, primært for å følge opp hvordan NFR vurderer kontakten med instituttene i forbindelse med søknader til NFR. Der det er relevant, som i evalueringen av miljøinstituttene, blir det også supplert med brukerundersøkelser i bedrifter, og i forvaltningen for øvrig.

I noen evalueringer er det også gjennomført besøk hos de evaluerte institusjonene. Så vidt vi kan se er ikke dette en metode som er utstrakt brukt, noe som trolig sammenfaller med at NFR sine evalueringer er blitt vesentlig mer omfattende. I evalueringen av de tolv regionale forskningsinstituttene fra 2012 ble det lagt opp til en fordeling mellom evalueringskomiteen i hvem som skulle besøke de forskjellige instituttene. Da den første av i alt fire evalueringer innen fordelingsarenaene i Norge ble gjennomført i 2015, ble møtene med de evaluerte instituttene lagt til Oslo. Her ble det gjennomført parallelle møter i forskjellige former, mens det i både HUMEVAL og SAMEVAL ble gjennomført intervju med de evaluerte institusjonene.

3.4 Oppfølging av evalueringer

I etterkant av en stor faglig-, tematisk-, eller instituttevaluering, oppnevnes det som regel et eget utvalg som skal følge opp evalueringene. Disse oppfølgingsevalueringene har to former. Den ene baserer seg på fagplaner, noe som vi ikke anser som direkte relevant for evaluering av forskningskvalitet. Den andre formen er oppfølgingsutvalg, som baserer arbeidet på den internasjonale evalueringen, innspill fra evaluerte miljøer, og faglige møter. Oppfølgingsutvalget skal følgelig basere seg på de viktigste funnene i den internasjonale evalueringen, og basert på dette komme med anbefalinger om oppfølging på lenger sikt. Vi anser dette som en del av hovedevalueringen, og vil ikke inkludere oppfølgingsrapporter som en del av typologiseringen.

4 Oppsummering

Gjennom de siste ti årene, som er den perioden vi i hovedsak har fokusert på, kan man skimte noen utviklingstrekk i evalueringsarbeidet. For det første ser vi et større internasjonalt fokus, både med mer internasjonale ekspertpaneler og mer sammenligning over landegrensene. Videre ser vi også at både NFR og NOKUT jobber mer med å utvikle nye evalueringsmetoder og -modeller. NOKUT utviklet både nye indikatorer på utdanningskvalitet og en ny evalueringsmodell i forbindelse med KOMBEVAL i 2018. I tillegg er NFRs og NOKUTs felles evaluering av samspillet mellom utdanning og forskning i forbindelse med SAMEVAL og KOMBEVAL et innovativt pilotprosjekt, ettersom utdanning og forskning tradisjonelt blir evaluert separat.

Evalueringene til NFR bærer preg av vesentlige endringer de siste 10 årene. Blant annet ble det i 2010-2012 gjennomført evalueringer som kun fokuserte på enkelte fag og som i stor grad var begrenset i omfang (se tabeller i vedlegg 2a). NFR beskriver selv utviklingen de siste årene i SAMEVALs hovedrapport (NFR, 2018). Hovedtrekkene er at de har gått fra mindre fagevalueringer, som etter vårt syn i stor grad har hatt et summativt fokus, til mer helhetlige evalueringer med en blanding av både formative og summative elementer. Med formativ i denne sammenheng mener vi, som nevnt tidligere, at evalueringene har en formativ funksjon for de evaluerte institusjonene, og ikke at de inneholder råd/anbefalinger til ansvarlige departement eller NFR. Vi finner i liten grad at metodene som anvendes i evalueringene til NFR har endret seg i samme grad som omfanget.

Noen substansielle endringer finner vi imidlertid, blant annet at de legger større vekt på bruk av siteringsindekser i bibliometriundersøkelsene de senere årene, sammenlignet med evalueringer i 2010-2015. I sistnevnte periode var det større fokus på hvorvidt forskning ble publisert internasjonalt, og i de siste årene synes det å ha vært et ytterligere fokus på «value-added» og vitenskapelig/akademisk «impact» av forskning som blir publisert i internasjonale tidsskrift. Det vil si i hvilken grad denne forskningen blir sitert eller henvist til i andre internasjonale studier, og i hvilken grad forskningen blir sitert i henhold til det som kalles «forventet sitering» av den type forskning. Noen evalueringer fra 2010 og 2011 baserte seg også på besøk hos de evaluerte institusjonene. Dette ble raskt endret til fellessamlinger, men hverken samlinger eller besøk virker gjennomførbart i store omfattende evalueringer som SAMEVAL og HUMEVAL.

Første helhetlige evaluering som samlet mer enn ett fagområde var *Evaluation of Biology, Medicine and Health Research in Norway*, som bestod av syv paneler fordelt på flere fagområder, inkludert klinisk forskning, psykologi og psykiatri. Tilsvarende overordnede evalueringer har senere blitt gjennomført blant annet i evalueringene *Evaluation - Research in Information and Communication Technology in Norway* og *Evaluation of the Norwegian Social Science Research Institutes*. Et annet moment som har blitt inkludert i evalueringer som HUMEVAL og SAMEVAL er samfunnsnytte og samfunnsrelevans – også omtalt som *societal impact*. Dette er fremhevet som viktige vurderingskriterier i flere evalueringer, blant annet beskrevet under kategori 9, og får stadig større fokus i utlysninger av forskningsmidler både på nasjonalt og internasjonalt nivå.

Som nevnt over, har det også skjedd en utvikling i metoder i NOKUTs evalueringer. NOKUT har gjennomført langt færre evalueringer enn NFR, og disse kan i hovedsak deles opp i to grupper eller faser. De første evalueringene (av ingeniør-, førskolelærer- og allmennlærerutdanningene i 2006, 2008 og 2010) benyttet seg av tradisjonelle evalueringsmetoder basert på institusjonenes selvrapportering, institusjonsbesøk og dokumentgjennomgang. De nyere evalueringene har i større grad fokusert på nye evalueringsmodeller. Her har evalueringspanelene, i tillegg til å evaluere fag og institusjoner, også kommet med tilbakemeldinger på evalueringsmodellen, datagrunnlag og lignende.

Selv om NFR har blitt mer formative i sine evalueringer over denne 10-årsperioden, kan funnene fra denne typologiseringen tyde på at NFRs evalueringer generelt er mer summative, mens NOKUT har et mer formativt fokus. Dette stemmer overens med NOKUTs beskrivelse av eget samfunnsoppdrag som at de skal bidra til å utvikle kvaliteten, bistå utdanningsinstitusjonene i utviklingsarbeidet deres og stimulere til kvalitetsutvikling som sikrer høyt internasjonalt nivå. Dette understrekes også ved at alle fagevalueringer kommer med konkrete anbefalinger til hver institusjon.

Litteraturliste

- Bennett, A., & Checkel, J. T. (Eds.). (2015). *Process tracing*. Cambridge University Press.
- Chen, H.-T. (1996). A comprehensive typology for program evaluation. *Evaluation Practice* 17(2): 121-130.
- Chinta, R., Kebritchi, M., & Ellias, J. (2016). A conceptual framework for evaluating higher education institutions. *International Journal of Educational Management*.
- Collier, D. (2011). Understanding process tracing. *PS: Political Science & Politics*, 44(4), 823-830.
- Collier, D., LaPorte, J., & Seawright, J. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly*, 65(1), 217-232.
- Hansen, H. F., T. Aarrevaara, L. Geschwind og B. Stensaker. (2019). Evaluation Practices and Impact: Overload? I Pinheiro, R., Geschwind, L., Hansen, H. F., & Pulkkinen, K. (red.) *Reforms, Organizational Change and Performance in Higher Education: A Comparative Account from the Nordic Countries*. Palgrave Macmillan.
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual review of sociology*, 36, 49-67.
- Mårtensson, P., Fors, U., Wallin, S. B., Zander, U., & Nilsson, G. H. (2016). Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy*, 45(3), 593-603.
- NFR (2010a). Philosophy and history of ideas in Norway - Evaluation of research 2004-2008. Rapport. Norges Forskningsråd, 2010.
- NFR (2010b). Basic Physics Research in Norway– An evaluation. Rapport. Norges Forskningsråd, 2010
- NFR (2010c). Sociological research in Norway - An evaluation. Rapport. Norges Forskningsråd, 2010
- NFR (2015a). Miljøinstituttene i Norge – Hovedrapport. Rapport. Norges Forskningsråd, 2015.
- NFR (2015b). Norwegian climate research - An evaluation. Rapport. Norges Forskningsråd, 2015.
- NFR (2018a). Evaluation of the Social Sciences in Norway (SAMEVAL – Rapport. Norges Forskningsråd, 2018.
- NFR (2018b). Evaluation of the Social Sciences in Norway (SAMEVAL) – Report from the Panel 3 – Political Science. Rapport. Norges Forskningsråd, 2018.
- NFR (2018c). Evaluation of Norwegian education research. Rapport. Norges Forskningsråd, 2018
- NFR (2018d). Evaluering av primærnæringsinstituttene. Rapport. Norges Forskningsråd, 2018.
- Reinertsen, H., K. Bjørkdahl, and D. McNeill. "Confronting the Contradiction—An Exploration into the Dual Purpose of Accountability and Learning in Aid Evaluation." *EBA (Expertgruppen för biståndsanalys)*, Stockholm (2017).

Scriven, M. (1991). Beyond formative and summative evaluation. In G. W. McLaughlin & D. C. Phillips (Eds.), *Evaluation and education: At quarter century*. Chicago, IL: University of Chicago Press, pp. 19-64.

Steel, D. (2004). Social mechanisms and causal inference. *Philosophy of the social sciences*, 34(1), 55-78.

Vedlegg 2a Typologisering av evalueringer - NFR

Evaluering (Kode + navn)	År	Kategorisering	Mandat	Metode	Kommentar
NFR_2018_PRIM_INST - Evaluering av primærnæringsinstituttene	2018	NFR Chen Institutt 2,3,6	Gjennomgang av instituttene roller, formål og oppgaver, rammebetingelser, samspill og samarbeid, relevans for brukerne samt fremtidige utfordringer og muligheter. Utvalget er spesielt bedt om å vurdere den fremtidige innretningen av basisfinansieringssystemet for de syv instituttene som er evaluert.	Egenevaluering, faktarapport, bibliometrisk undersøkelse, utredning om basisfinansiering og arenasammensetning, møter med instituttene og begrenset brukerundersøkelse.	Fjerde og siste instituttevaluering innenfor de fire fordelingsarenaene for statlig basisfinansiering. Bibliometriundersøkelse gjennomført av Oxford Research i samarbeid med CWTS/Leiden, basert på både internasjonale databaser og tall fra CRISTin. Utredning om basisfinansiering gjennomført av NIFU.
NFR_2018_SVFI NS - Evaluation of the Norwegian Social Science Research Institutes	2018	NFR Chen Institutt 2,3,6	Evaluere relevansen, kvaliteten og ferdigheter, rekruttering og bidrag til utdanning av forskere, internasjonalt samarbeid, rammebetingelsene til instituttene og instituttene struktur og rolle i FoU-utviklingen	Egenevaluering, faktarapport, rapport om samfunnsrelevans, Bibliometrisk undersøkelse, møter med instituttene og begrenset brukerundersøkelse.	Tredje instituttevaluering innenfor de fire fordelingsarenaene for statlig basisfinansiering.
NFR_2018_SAM EVAL - Evaluering av samfunnsvitenskapelig forskning i Norge (SAMEVAL)	2018	NFR Chen Tematisk 10	Sammenfatte forskningskvalitet og relevans, strukturelle utfordringer og hvordan institusjonene og NFR har fulgt opp tidligere evalueringer. Skal gi innspill til fremtidig politikk på forskningsområdet	Statistikk, bibliometrisk undersøkelse, dokumentstudier, intervju, selvevalueringer	Delt inn i seks paneler: Geografi, samfunnsøkonomi, statsvitenskap, sosiologi, sosialantropologi og økonomi- og administrative fag. Første NFR-evaluering som spesifikt fremhever samfunnsnytt og samfunnsrelevans av forskningen som en del av evalueringen. Følger ellers modellen satt i HUMEVAL. Bibliometrisk undersøkelse brukt i varierende grad, utenom i panelet som evaluerte samfunnsøkonomi, som følge av dårlig kvalitet i andre fagfelt enn samfunnsøkonomi.

Evaluering	År	Kategorisering	Mandat	Metode	Kommentar
NFR_2018_SAM EVAL_PN - Seks paneler; geografi, samfunns- økonomi, statsvitenskap, sosiologi, sosiologantropologi og økonomi- og administrative fag.	2018	NFR Chen Fag 5,6,8,9	Kritisk gjennomgang av forskningskvaliteten i internasjonal kontekst, gi anbefalinger for fremtidig forskning, og gi innsikt, råd og anbefalinger til hvordan institusjonene kan bedre forskningsstandarder.	Statistikk, bibliometrisk undersøkelse, dokumentstudier, intervju, selvevalueringer, studiebarometer (NOKUT).	SAMEVAL og HUMEVAL ser på sammenkoblingen mellom forskning og undervisningsaktivitet og tverrfaglig forskning. I motsetning til HUMEVAL, identifiserer ikke SAMEVAL forskningsgrupper som har potensial for publisering på høyt internasjonalt nivå.
NFR_2018_UTD EVAL - Evaluation of Norwegian education research	2018	NFR Chen Fag 1,5,6	Identifisere og analysere sentrale utviklingstrekk i forskningsfeltet. Avdekke styrker og svakheter i forskningens kvalitet og kapasitet, strategisk fokus og samhandling og relevans og samfunnsbidrag.	Intervju, dokumentstudier, spørreskjema, statistikk, egenevaluering, bibliometrisk undersøkelse.	Evaluerer kvalitet og kapasitet i forskningen. Strategisk fokus og samarbeid. Relevans og samfunnsmessig påvirkning basert på egnevaluering.
NFR_2017_POL - Evaluation - Norwegian Polar Research	2017	NFR Chen Tematisk 1,4	Evaluerer forskningskvalitet og forskningens påvirkning, nasjonal, internasjonalt samarbeid, relevans, og gi forslag til fremtiden på temaet.	Spørreskjema, dokumentstudier, bibliometrisk undersøkelse, observasjon, dialogmøter, statistikk, egnevaluering.	Bibliometrisk undersøkelse gjennomført av NIFU. Relevans vurdert etter siteringer. Panelet problematiserer også i hvilken grad siteringer er et godt mål på forskningens påvirkning.
NFR_2017_HUM - Evaluation of the Humanities in Norway (HUMEVAL)	2017	NFR Chen Fag 10	Første evaluering på 30 år for feltet under ett. Kritisk gjennomgang av forskningskvaliteten i internasjonal kontekst, gi anbefalinger for fremtidig forskning, og gi innsikt, råd og anbefalinger til institusjonene evaluert hvordan institusjonene kan bedre forskningsstandarder	Statistikk, intervju, selvrapportering, dokumentstudier, bibliometrisk undersøkelse (inkl. siteringer).	Delt inn i åtte fagpaneler. Evaluerer forskningsområder og forskningsgrupper, samt samfunnsrelevans. Sistnevnte basert på «Impact statements» fra UK Research Excellence Framework.

Evaluering	År	Kategorisering	Mandat	Metode	Kommentar
NFR_2017_HUM_PN - Fagevalueringer under HUMEVAL-åtte panel; estetiske studier, nordiske språk og lingvistikk, nordisk og komparativ litteratur, moderne og klassiske språk, litteratur og områdestudier, arkeologi, historie og kulturstudier, filosofi, religion og teologi, og mediastudier.	2017	NFR Chen	Ekvivalent av det beskrevet i NFR_2017_HUM	Intervju, faktarapport, spørreskjema, dokumentstudier, egnevalueringer, bibliometrisk undersøkelse. Gjennomført av evalueringskomite, NIFU og Oxford Research.	SAMEVAL og HUMEVAL ser på sammenkoblingen mellom forskning og undervisningsaktivitet og tverrfaglig forskning. I motsetning til SAMEVAL, identifiserer HUMEVAL forskningsgrupper som har potensial for publisering på høyt internasjonalt nivå. Ellers er mandatet likt som SAMEVAL.
NFR_2016_INGI_NST - Evaluation of Norwegian Technical Industrial Research Institutes	2016	NFR Chen	Bidra til ny kunnskap om styrker og forbedring blant 14 teknisk-industrielle forskningsinstitutt med basisbevilgning. Vurderer om forskningsinstituttene bidrar med kunnskap til offentlige og næringslivet, hvordan rollen til instituttene blir oppfylt i instituttsektoren, vurdering av internasjonalt samarbeid, og vurdering av den finansielle situasjonen.	Egenevaluering, faktarapport, spørreundersøkelse, dokumentstudier av forskningens relevans og bibliometrisk undersøkelse.	Andre evaluering innenfor de fire fordelingsareanene i instituttsektoren.

Evaluering	År	Kategorisering	Mandat	Metode	Kommentar
NFR_2015_ING - Basic and long-term research within Engineering Science in Norway	2015	NFR Chen Fag 2,4	Kritisk gjennomgang av forskning i et internasjonalt perspektiv, internasjonalt tilbakemelding på fremtidige utfordringer og hvordan disse skal møtes, og gi NFR grunnlag for forskningspolitiske råd til regjeringen. Identifisere forskergrupper som har oppnådd høy internasjonalt standard, områder som må forsterkes, rekruttering og oppfølging av tidligere evalueringer.	Intervju, statistikk, egenvurdering fra forskningsgruppene, dokumentstudier basert på evaluering av forskningsrelevans, samarbeid nasjonalt og internasjonalt og bibliometrisk undersøkelser.	10 år siden forrige evaluering. Eget kriterium for evaluering av samfunnsmessig og faglig (value-added) påvirkning.
NFR_2015_MILJ NST - Miljøinstituttene i Norge	2015	NFR Chen Institutt 2,3,6	Evalueringen skal belyse formål og oppgaver, rammebetingelser, rolle i forskningssystemet, samspillet mellom instituttene og andre FoU-aktører nasjonalt og internasjonalt, relevans, og fremtidige utfordringer og muligheter. Evalueringen skal gi råd for utforming av nasjonal instituttpolitikk og gi grunnlag for utforming av NFR sin politikk.	Intervju, selvrapportering, dokumentstudier inkludert faktarapport og evaluering av instituttene relevans, spørreskjema, statistikk.	Første evaluering innen de fire fordelingsarenaene i instituttsektoren. De åtte instituttene vurderes enkeltvis og samlet. Instituttene skal bidra med forskning av høy kvalitet og relevans for næringsliv, forvaltning og samfunnet.
NFR_2012_REGI NS - Evaluering av de tolv regionalforskning s-instituttene	2012	NFR Chen Institutt 1,2,4	Vurdere instituttene kompetanse, kvalitet og relevans, instituttene rammebetingelser og oppdragssituasjon, instituttene samarbeidsevne og instituttene strategier for fremtiden.	Faktarapport utarbeidet av NIFU, egnevaluering, instituttbesøk, brukerundersøkelse, dokumentstudier.	Rapporten bidrar til vurderingen av et nytt basisfinansieringssystem for forskningsinstituttene. Rapporten skal bidra til å styrke kunnskapsgrunnlaget for instituttene egne strategiske utvikling og for å utvikle en målrettet og effektiv instituttpolitikk.
NFR_2012_KLIM A - Evaluation - Norwegian Climate Research	2012	NFR Chen Fag 2,4	Kritisk gjennomgang av norsk klimaforskning i et internasjonalt perspektiv, og gi forslag til politikkutforming.	Intervju, faktarapport, selvevaluering, bibliometrisk undersøkelse, spørreskjema, dokumentstudier, høring.	Mandatet spesifiserer fire temaer som skal evalueres: forskningskvalitet og kapasitet, strategisk fokus og interaksjon, samarbeid med interessenter, og relevans for samfunnsutfordringene.

Evaluering	År	Kategorisering	Mandat	Metode	Kommentar
NFR_2011_SOS ANT - Social and cultural anthropological research in Norway	2011	NFR Chen Fag 1,2,4	Vurdere kvaliteten på forskning i et internasjonalt perspektiv, beskrive styrker og svakheter med forskningen, identifisere områder hvor kvaliteten kan bedres, øke kunnskapen til forskningsgrupper, NFR og departement om fremtidens forskning på temaet.	Selvevaluering, statistikk, intervju og bakgrunnsrapport fra NIFU.	Evaluert de fire sosialantropologiske instituttene i Norge, samt seks andre forskningsgrupper.
NFR_2011_MED - Evaluation of Biology, Medicine and Health Research in Norway	2011	NFR Chen Chen Tematisk 1,2,4 8	Kritisk gjennomgang av styrker og svakheter ved forskningen i et internasjonalt perspektiv med fokus på vitenskapelig kvalitet og relevans. Identifisere områder med sterk vitenskapelig posisjon, og områder som er svake, og vurdering av den institusjonelle strukturen.	Selvevaluering, statistikk, intervju og bakgrunnsrapport fra NIFU.	Ser også på tverrfaglig samarbeid og om institusjonene har fulgt opp tidligere evalueringer, og vurderer samfunnsrelevans av forskningen.
NFR_2010_PSY CH - Basic Physics Research in Norway	2010	NFR Chen Institutt 1,2,4	Kritisk gjennomgang av forskning i et internasjonalt perspektiv, identifisere forskningsgrupper på høyt internasjonalt nivå eller som har potensiale til å og internasjonalt tilbakemelding på hvordan norsk forskning skal møte utfordringene fremover, spesielt tverrfaglige forskningsfelt og samfunnets utfordringer.	Selvevaluering, statistikk, intervju, site-visits på utvalgte institusjoner, «høring» i Oslo, og bakgrunnsrapport fra NIFU.	Evalueringen inkluderer ikke utelukkende forskning finansiert av NFR, men også av Fondet for forskningsavgift på landbruksprodukter og Forskningsmidler over Jordbruksavtalen. Evalueringen basert på 2004-2008.
NFR_2010_PHIL - Philosophy and the history of ideas in Norway: Evaluation of Research 2004- 2008	2010	NFR Chen Fag 1,2,4	Kritisk gjennomgang av forskningens kvalitet og relevans, nasjonal og internasjonalt samarbeid, vurdering av hvilke disipliner eller underdisipliner som står sterkt eller svakt internasjonalt sett, og bidra til å utvikle/forbedre god forskning og gode forskningsmiljø i Norge.	Selvevaluering, statistikk, intervju, bibliometrisk analyse og bakgrunnsrapport om finansiell situasjon, publisering og ansatte.	Evalueringen begrenset til institutt på UiB, UiO, NTNU og UIT.

Evaluering	År	Kategorisering		Mandat	Metode	Kommentar
NFR_2010_SOS SV - Sociological research in Norway	2010	NFR	Fag 1,4	Gi bilde av kvaliteten på forskningen ut fra en internasjonal standard, bidra til læring og egenutvikling i forskningsmiljøene, identifisere tiltak som kan bidra til å fremme kvalitet, og styrke forskningsmiljøenes, NFRs og departementenes kunnskapsgrunnlag for videreutvikling av forskningen.	Dokumentstudier (inkl bakgrunnsrapport om finansiell situasjon, publisering og ansatte), selvevaluering, bibliometrisk undersøkelse, intervjuer.	13 enheter evaluert, hvorav to ble evaluert under ett etter eget ønske. Inkluderer bakgrunnsrapport fra NIFU Step (Nå NIFU) om ansatte, finansiering og publisering.
NFR_2010_ØKL ND - Evaluering av norsk økologisk landbruksforskning (1999-2009)	2010	NFR	Tematisk 1,4	Evaluere omfanget av økologisk forskning, faglig kvalitet, forskningens rolle med sikte på å fremme økologisk mat og potensial for kunnskapsoverføring, kunnskapsoverføring til konvensjonelt landbruk, og fremskaffe grunnlag for NFR som gjør at de kan anbefale innretning av fremtidig forskning.	Statistikk, dokumentstudier, intervju, spørreundersøkelse	Gjennomført i to deler. En for faglig kvalitet og en for vurdering opp mot regjeringens mål om kunnskapsoverføring til konvensjonelt landbruk.

Vedlegg 2b Typologisering av evalueringer – NOKUT

Evaluering	År	Kategorisering	Formålsbeskrivelse	Metode	Kommentar
NOKUT_2018_KOMBEVAL_ECON - Educational quality in economics in Norway	2018	Chen 1	Evaluere og bidra til forbedring av utdanningskvaliteten innen økonomi i Norge, gi direkte tilbakemeldinger til hvert institutt, samt å teste ut en ny modell for slike evalueringer.	Dokumentstudier, statistikk (Studiebarometeret, DBH), selvevalueringer og eksterne rapporter fra eksperter innen pedagogikk.	Panelet kritiserte rammene for evalueringen for å fokusere for mye på prosess, og mente at resultatet er det viktigste målet for utdanningskvalitet.
NOKUT_2018_KOMBEVAL_STV - Educational quality in political science in Norway	2018	Chen 1, 8	Evaluere og bidra til forbedring av utdanningskvaliteten innen statsvitenskap i Norge, gi direkte tilbakemeldinger til hvert institutt, samt å teste ut en ny modell for slike evalueringer.	Dokumentstudier, statistikk (Studiebarometeret, DBH), selvevalueringer og eksterne rapporter fra eksperter innen pedagogikk.	
NOKUT_2018_KOMBEVAL_SOS - Educational quality in sociology in Norway	2018	Chen 1, 8	Evaluere og bidra til forbedring av utdanningskvaliteten innen sosiologi i Norge, gi direkte tilbakemeldinger til hvert institutt, samt å teste ut en ny modell for slike evalueringer.	Dokumentstudier, statistikk (Studiebarometeret, DBH), selvevalueringer og eksterne rapporter fra eksperter innen pedagogikk.	
NOKUT_NFR_2018 - Joint evaluation of research and education - a pilot study	2018	Chen 1, 5	Evaluere hvordan forskning og utdanning påvirker hverandre og hvordan dette påvirker kvalitet, samt å teste en modell for å slike evalueringer.	Data fra SAMEVAL og KOMBEVAL, litteraturstudie, dokumentstudier og data fra Studiebarometeret.	Panelene besto av medlemmer fra både SAMEVAL- og KOMBEVAL-panelene. Litteraturstudiet om samspillet mellom utdanning og forskning ble utført av NIFU.
NOKUT_2010_FØRLÆR - Evaluering av førskolelærerutdanning i Norge 2010	2010	Chen 1, 8	Fremskaffe kunnskapsgrunnlag for videre utvikling av utdanningen, særlig med tanke på arbeidslivsrelevans.	Selvevalueringer, spørreundersøkelser (NOKUT og StudData), dokumentstudier, statistikk og intervjuer/institusjonsbesøk.	Fokuserer i stor grad på prosess, men tar også inn evaluering av resultater mot slutten (antall uteksaminerte, studiepoengproduksjon, antall menn osv.)

