

---

# 1 Innleiing

Statskonsult inviterte til ein konferanse 9. og 10. april 1997 om departementa si evalueringsverksemd. Målgruppa var tilsette som arbeider med evalueringar, i forvaltninga og i forskingsmiljø. Føremålet med konferansen var å bidra til bevisstgjerjing om, og felles forståing av, ulike sider ved evaluering som styringsverktøy.

Bakgrunnen for å ta opp dette emnet på brei basis no, er at ein gjennom det nye økonomireglementet for staten set krav om meir systematisk bruk av evalueringar. Vi ynskjer å markere at dette er ei ny og felles oppgåve for heile statsforvaltninga. Dessutan er det i utgreiingsinstruksen slått fast at ein ved offentlege reformer, regelendringar og andre tiltak skal gjennomføre konsekvensutgreiingar. Samstundes er det gjort institusjonelle endringar. Stortinget har oppretta ein Kontroll- og konstitusjonskomite som og tek opp spørsmål knytt til ressursutnytting. Riksrevisjonen har utvida sitt ansvarsfelt, med m.a. fokus på forvaltning av tilskot og på samanhangen mellom ressursbruk og oppnådde resultat.

Vi ser det som sær s viktig at departementa sin praksis på dette området utviklar seg på ein tenleg måte for forvaltninga som heilskap. Ein føresetnad er då at departementa har ei realistisk og mest mogleg samstemt oppfatning av korleis evalueringar kan brukast, og i tida framover bør bli brukt, spesielt i budsjett-sammanheng. Konferansen var tenkt som eit første steg i retning av å etablere ei slik felles forståing.

For å ha eit felles utgangspunkt i denne prosessen, har Statskonsult bede Norsk institutt for studier av forskning og utdanning (NIFU) om å kartleggje departementa sin noverande praksis, jf. kap.5. Kartlegginga syner på den eine sida at det blir gjort god bruk av dei evalueringane som departementa får utført i dag, på den andre sida at praksisen er svært fragmentert. Det er og ei relativt svak kopling mellom evalueringane og budsjettarbeidet i departementa. Det ser ut til at overordna og gjennomgåande omsyn i forvaltninga i liten grad blir teke vare på og at få av evalueringane som vert utførte i dag, er av det slaget som økonomireglementet legg opp til. Slik sett står det mykje att før evaluering som verkemiddel til informasjon og tilbakemelding om departementa sitt arbeid, er tatt i bruk på systematisk måte. Ser ein udekt evalueringsbehov og dagens praksis i sammanheng, blir den verkeleg store utfordringa å halde evalueringsaktiviteten på et rimeleg nivå.

Hovudspørsmåla det vart sett søkjelys på gjennom plenumsforedrag og i gruppediskusjoner, var: Kva inneber krava om evaluering i økonomireglementet for staten? Kva bør vere emne for evaluering og kvifor? Kva er dei viktigaste elementa i planlegging og gjennomføring av evalueringar (planlegging og utforming av mandat, fagleg tilnærming/krav til evaluatorkompetanse,

---

organisering og rollefordeling, spreining og bruk av resultata)?

Gjennom å gje ut denne rapporten ynskjer vi å formidle kunnskap og røynsler som deltakarar på dette området har systematisert. Materialet kan kome til nytte etterkvart som det vert gjort erfaringar med å setje økonomireglementet i verk, og det kan vere relevant for departementa sitt arbeid m.a. med å førebu resultatrapporteringa. Vi tek sikte på å invitere til nye møte om meir avgrensa spørsmål når dei nye, felles rammeføresetnadene er meir på plass. For Statskonsult vil rapporten vere ei kjelde til informasjon både om utfordringar og moglege tiltak. Vi vil bruke materialet som utgangspunkt for å planleggje korleis konferansen bør følgjast opp.

---

## 2 Åpning av konferansen

*Departementsråd Brit Denstad,  
Planleggings- og samordningsdepartementet*

For 10 års siden var **evaluering** et lite brukt begrep her i landet. Det ble nesten utelukkende koplet til vitenskapelig arbeid (forskning) og enkeltpersoners innsats, altså ikke til en vurdering av institusjoner. Det ble nok satt igang en del forsøk i offentlig sektor som (i alle fall i prinsippet) skulle evalueres, men hvor praksis var nokså tilfeldig. Evalueringsbegrepet hadde ingen sterk forankring i offentlig virksomhet.

Siden den tid har pendelen svingt i motsatt retning. I dag kan du knapt lese et offentlig dokument, delta i konferanser, eller – for den saks skyld – skimme avisene, uten at evalueringsbegrepet dukker opp. Det er blitt ett av nøkkelbegrepene også i vår offentlige verden. Enten er det noe som skal evalueres, er evaluert eller burde vært evaluert. Det er – nær sagt – ikke den virksomhet eller det prosjekt som ikke skal under lupen. Evalueringsansvaret er nå også formelt forankret i statens **nye økonomireglement** – det er satt i system.

Og det er utmerket. Fokus for interessen er flyttet fra først og fremst å få satt tiltak igang, til å engasjere seg i gjennomføringen og i resultatene – til å ta ansvar for hvordan fellesskapets ressurser brukes og hva som kommer ut av innsatsen. Det er spørsmål om ambisjonene blir omsatt i praksis, i tråd med de mål og resultatkrav som er fastsatt, både i løpende virksomhet og i avgrensede reformtiltak. Det skal bli spennende å høre hvilke resultater som kan legges fram fra **kartleggingen** av departementenes evalueringsevirsomhet.

**Evaluering** av løpende virksomhet har nær sammenheng med arbeidet med **kvalitetssikring, kvalitetsutvikling og total kvalitet**. Også dette er begreper som setter fokus på prosess og resultater, og som vi strever med å gi innhold i det daglige arbeidet. En del av utfordringen vi står overfor i offentlig sektor må være å bruke evaluering systematisk, som ledd i arbeidet med å heve kvaliteten på vår egen virksomhet og de tjenester vi har ansvar for. Poenget er altså ikke bare å rettferdiggjøre ressursbruk, men å lære for å gjøre jobben på en bedre måte – finne bedre ordninger for å løse oppgavene. Her dreier det seg både om **læring** og **kontroll**.

Men fortsatt er det vel slik at vi ikke er helt presise i forståelsen av hva begrepet innebærer, utover at det forutsetter en systematisk vurdering. **Evaluering** kan være så mangt. Det dekker varierende faglig ambisjonsnivå og bruk av ulike metoder. Slik skal det også være, under forutsetning av at valg av problemstillinger, metode og opplegg er **bevisst tilpasset formålet** med

---

evalueringen. En løpende intern evaluering av hvordan et departement fungerer, som økonomireglementet forutsetter at vi engasjerer oss i, krever en annen systematikk og andre metoder enn en ekstern evaluering av en omfattende, kompleks og kanskje omstridt reform, enten den nå er forankret i helsesektoren, samferdselssektoren eller utdanningssektoren.

Det er noen **forutsetninger** som må være på plass hvis en omfattende evalueringsvirksomhet skal tjene sitt formål.

**For det første** må det være satt **klare mål** for virksomheten som skal evalueres, både overordnede mål og **realiserbare resultatmål**. Faglig avanserte evalueringsopplegg har liten verdi hvis resultatmålene er diffuse eller i praksis ikke kan identifiseres. Det er f.eks. ikke enkelt å vurdere resultatet av et norsk Øst-Europaprogram hvis resultatmålet er «å bidra til å sikre demokrati i Øst-Europa». Det er som kjent mange forhold som påvirker en demokratiseringsprosess – over lang tid. For å si noe konkret om resultater av en innsats må det være en mulig, direkte kopling mellom aktivitet og virkning. Det måtte ev. settes igang et større internasjonalt forskningsprogram for å kunne si noe om hvilke forhold som påvirker demokratiseringsprosessen i en geopolitisk region. Og det ville vel vil sprengte evalueringsbegrepets ramme.

**En annen forutsetning** er knyttet til kvalitet. Evalueringsprosjekter må holde solide **faglige mål**. I det ligger implisitt en forutsetning om faglig uavhengighet. Høye kvalitetskrav er spesielt viktig når det skal trekkes slutninger på grunnlag av arbeidet om behov for endringer i praksis. Evalueringer må ha legitimitet. Det finnes eksempler på evalueringer av **tvilsom kvalitet**. Det hindrer ikke at de kan få konsekvenser. Her ligger risiko for feilvurderinger med skadevirkninger det kan ta lang tid å rette opp.

**For det tredje** må også evalueringstiltak vurderes i en kost/nytte sammenheng. Valg av evalueringsmodell må vurderes både i forhold til formål og virksomhetens betydning. Evaluering er **krevede** i seg selv, både når det gjelder bruk av faglig kompetanse, krav til kvalitet, tid og penger. Mange evalueringer er blitt for «tunge» og dyre i forhold til formålet, andre for overfladiske i forhold til de spørsmål prosjektet skal belyse. Evalueringsinnsats krever **bevisst prioritering**, både i departementene og i de forsknings- og fagmiljøer som engasjerer seg i denne type faglig arbeid.

**Ytterligere en forutsetning** for en meningsfylt evalueringsinnsats har sammenheng med håndteringen av forholdet mellom evalueringsprosjektet og den praksis som skal vurderes. Slike prosjekter krever en spesiell **aktsomhet**. De forutsetter innsikt i og respekt for praksisfeltet, og tilstrekkelig avstand til å sikre en kritisk, faglig basert vurdering – en krevende balansegang.

---

**En siste forutsetning** jeg skal nevne har å gjøre med selve hovedpoenget: bruk av den innsikt og de resultater evalueringen bringer fram. Det må finnes kompetanse også på «brukersiden» – i departementer og etater – til å vurdere hvordan resultatene skal tolkes og brukes. Men det er også viktig at politisk og administrativ ledelse har styrke til å håndtere både offentlighet omkring resultatene og den kunnskap resultatene måtte gi. Det er jo ikke alltid den er like behagelig. Men i et demokratisk styrt land som vårt er også denne type kunnskap et selvfølgelig og viktig bidrag til en åpen og kunnskapsbasert samfunnsdebatt.

### **Avslutning**

Det er et hovedpoeng at systematisk bruk av evalueringer i det offentlige arbeid er et **potensielt** sterkt og viktig redskap for å styrke kvalitet og øke effektivitet. Det er vår oppgave å sørge for at det brukes med forstand. Det må ikke bli et instrument for detaljstyring og overdreven dokumentasjon, men for innsikt, resultatorientering og grunnlag for å korrigere kursen. Vi må ha et kritisk blikk på hva vi **vil ha** ut av evalueringene og på hva vi faktisk **får** ut av dem.

NIFUs kartlegging av departementenes evalueringspraksis og Statskonsults konferanse må være et meget velkomment bidrag til vårt arbeid med å få et bevisst og profesjonelt forhold til evalueringer og bruken av den kunnskap de gir. Det er med betydelige forventninger jeg kan åpne konferansen.

---

### 3 Evaluering i nytt økonomireglement for staten

*Spesialrådgiver Øyvind Engh, Statskonsult*

#### Nytt økonomireglement for staten

- Tidligere reformtiltak samles
- Hovedregel om resultatstyring
- Spesiell plikt til å følge opp *tilskuddsforvaltning*
- Krav til styring av underliggende forvaltningsorganer, *etatsstyring*
- Nye systemløsninger
- «Gjennom dette systemet skal det gis opplysninger om de mål som er satt for statlige tiltak. Dette vil gjøre det mulig å stille institusjonene friere ved gjennomføring av tiltakene... En slik desentralisering av ansvar for valg av alternativ for å løse oppgaver vil gi de enkelte enheter økte incentiver og muligheter for effektiv ressursutnyttelse. Samtidig kan de politiske myndigheter styre utviklingen uten en unødvendig grad av detaljstyring... En forutsetning for en slik delegering er imidlertid at Stortinget og regjeringen får seg forelagt rapporter som kan gi grunnlag for å kontrollere bl a i hvilken grad den statlige virksomheten har ført til ønskede resultater». St meld nr 37 (1973–74) s 11 og 15.

#### Haga-utvalget 1984

- Resultatorientert styring av virksomheter fra overordnet departement
- Resultatorientert styring innad i virksomheter
- Endringer i budsjettssystemet
- Disponering av personellressurser
- Press, stimulerings- og konkurransevirkemidler

#### Den nye staten – 1987

- Mange forsøks-, utviklings- og utredningsforslag
- Et konkret forslag: Virksomhetsplan for alle innen utgangen av 1990 –*det ble mye plan og lite oppfølging!*
- Evaluering ikke nevnt

#### 90-tallet

- 1990–91: Forpliktende rammebudsjettering
- 1992: Regjeringens resultatstyringsvedtak
- 1993–95: Arbeid med fagprp, tildelingsbrev, styringsdialog, etatsstyring
- 1994: Riksrevisjonens Dok. nr 3:1 – Om tilskuddsforvaltning
- 1995–96: Ny komité i Stortinget, ny avdeling i Riksrevisjonen

---

## **1996–97 Økonomireglementet**

- Et omfattende regelverk, etater og tilskudd, styring og kontroll
- Felles innføring
- Samme fundament – mer realisme – bedre systemløsninger
- Store utfordringer til virksomhetstilpassing og til å se ulike styringsformer i sammenheng
- Ikrafttredelse 10.1.97, krav om tilpasning innen 1.7.98

## **Hovedregel om resultatstyring**

- Departementet skal fastsette mål og resultatkrav, samt rapportere på en hensiktsmessig måte
- Departementet skal påse at det for etatene blir utarbeidet metoder og resultatindikatorer for å vurdere om etaten viser resultater i samsvar med det som er fastsatt
- Virksomhetsplaner skal utarbeides med utgangspunkt i tildelingsbrev og rapportering skje i henhold til fastsatte mål
- Spesiell plikt for departementene til å formulere målsetting for tilskuddsbevilgninger
- Departementet må med mellomrom gjennomføre evalueringer – oppnås de *fastsatte* målsettinger med virksomheten eller tilskuddsordningen, særlig vekt på de *forutsatte* samfunnsmessige virkninger
- Ledelsesansvaret understrekes

## **Evalueringskravet utdypes**

- Gjelder for virksomheter, tilskudd til private organisasjoner m.m., tilskudd til kommuner og fylkeskommuner og til garantiordninger
- Initieres ut fra politisk viktighet, økonomiske størrelser, resultatutvikling, antatt effektiviseringspotensiale
- Jo mindre presise årlige resultater, jo viktigere med evaluering

---

## **Bruk av evalueringer**

- Evalueringer skal gjennomføres slik at resultatet senest foreligger i januar – kobling til budsjettprosessen
- Resultater skal presenteres for Stortinget i St prp nr 1
- Evalueringer skal være tilgjengelige for Riksrevisjonen
- Hvordan benyttes evalueringer i budsjettprosessen?

## **Eksempel:**

### **Statskonsults gjennomgang av tilskuddsordninger**

- Ordningene evalueres i forhold til målsettingene
  - operasjonaliserte og etterprøvbare mål
  - konsistens ifht tildelings- og oppfølgingskriterier
- Evaluering av etablerte kontrollordninger
  - forvaltning: organisasjon, saksbehandling og kontroll
- Effekter

### **Bør alle evalueringer knyttes til økonomiregelverket?**

- Vekt på effekter og måloppnåelse
- Hva med prosessevalueringer?
- Ansvar for å utforme sektorpolitikken viktigste beveggrunn for initiering av evalueringer
  - selv om: det er viktig at vurderinger av årlig resultat og produksjon i virksomheten utdypes med et langsiktig perspektiv som går ut over sektorgrenser



---

## 4 System og praksis i Danmark

*Underdirektør Kaj Kjærsgaard, Økonomistyrelsen, Danmark*

### Begrepet evaluering

- Systematisk vurdering af offentlige indsatsers design, praksis og effekt

### Forskjellige evalueringsformer

- Evaluering i forhold til
  - målopfyldelse
  - proces
  - effekt
  - interessant

### Har vi en evalueringspolitik i Danmark?

- Nej – ingen generelle forvaltningspolitiske bestemmelser
  - ingen systematisk anvendelse
  - ingen forpligtende regler
- Ja – udbredt evalueringspraksis på nogle områder
  - megen analyse og udredning i den statslige forvaltning
  - udbredt inden for visse sektorer

### Evaluering anvendes i stigende omfang

- Årsager
  - ændret kompetenceprofil i forvaltningen
  - øget udbud af evalueringskompetence og af anvendt forskning
  - mer fleksible, programorienterede forvaltningsformer
  - stigende konfliktniveau og øget politisering af forvaltningen

### Evalueringspraksis opprioriteres

- Øget mål- og resultatstyring har medført
  - kontraktstyring
  - virksomhedsregnskab
  - controllerfunktion
  - ledelsesinformationsystemer
- Krav om effektive institutioner via ad hoc undersøgelser, der evaluerer effekten af større aktiviteter
  - budgetanalyser
  - turnusundersøgelser
  - benchmarking
  - brugerundersøgelser

- 
- Model for forvaltning af tilskudd. Her indgår evalueringsmodul

### **Hvem gennemfører evaluering?**

- Organisationen selv
  - stabsfunktion (fx controller eller analyseenhed)
  - sektorforskningsinstitut
- Finansministeriet (departementet)
  - budgetanalyser
- Økonomistyrelsen
  - organisationsanalyser
  - økonomistyringsanalyser
- Rigsrevisionen
  - forvaltningsrevision
- Selvstændig evalueringenheder i statsforvaltningen
  - Danida
  - Evalueringscentret for videregående uddannelser
- Andre
  - private konsulentfirmaer (via udbud)
  - forskningsverden

### **Bruges evalueringer til noget/forbedrer de den offentlige sektor?**

#### **Ja – de både bruges og misbruges**

- Succes:
  - budgetanalyser – frembringer et effektivt beslutningsgrundlag
  - ledelsesinformationssystemer med resultat- og effektindikatorer
- Mindre succes:
  - effektevalueringer – legitimerer de politiske beslutninger (det politiske konfliktniveau øges)
  - dårlig og forhastet tilrettelæggelse af evalueringer

### **Hvordan kan evaluering som styringsredskab forbedres?**

- Indbygges i de interne styringsprocesser
- Etablering af systematiske eksterne evalueringer
- Opbygning af intern evalueringseksperitice (bestillerkompetence)
- Professionalisering av evaluatorene
- Gennemtænkt og planlagt tilrettelæggelse af evalueringen

---

## 5 Departementenes evalueringspraksis – en kartlegging

*Seniorutreder Karl Erik Brofoss  
Norsk institutt for studier av forskning og utdanning (NIFU)*

### Sammendrag

Departementene har i perioden 1994–96 gjennomført 206 evalueringer til en kostnad på omlag 180 millioner kroner. Evalueringer har således fått et betydelig omfang i sentraladministrasjonen generelt, men det er store variasjoner mellom departementene, både med hensyn til antall evalueringer og den ressursmessige innsatsen som legges ned i denne virksomheten.

Evalueringene er knyttet til ulike virksomheter i de forskjellige departementene, men de fleste er orientert mot effektevalueringer knyttet til det kontrollansvar departementene har i forbindelse med iverksetting av offentlige tiltak. Det tyngste bruksområde for evalueringene er i forbindelse med omorganisering og reformvirksomhet. Det er imidlertid en relativ svak kopling mellom evalueringsevirsomhet og departementenes budsjettarbeid.

I hovedsak er det fagavdelingene som er initiativtakere til evalueringene. Det er kun i forbindelse med igangsettelse av evalueringer av reformer at den politiske ledelse markerer seg, men også her er det fagavdelingene som er de mest aktive. Bare unntaksvis er evalueringene satt i gang uten at det først er identifisert hvem som skal være brukere av resultatene.

Hovedtyngden av evalueringene utføres innen instituttsektoren, hvor seks institutter står for halvparten av evalueringene. Det er ikke tilfeldig hvor de ulike instituttene henter sine evalueringoppdrag. De institutter som har sitt virkeområde generelt innen et departements ansvarsområde er også de som får de fleste oppdragene – som regel uten forutgående anbudsrunder.

Departementene har en tett oppfølging av evalueringene ved at de oppretter styringsgrupper som er aktive særlig i oppstartingen av prosjektene.

Departementenes hovedinntrykk av evalueringene er svært positivt. Det store flertallet av evalueringene bedømmes til å ha et godt datagrunnlag, de konklusjoner som trekkes har et solid empirisk fundament og anbefalingene følger av konklusjonene.

De fleste konklusjoner og anbefalinger i evalueringsrapportene brukes. Først og fremst brukes de til å forbedre eksisterende tiltak sammen med det å få nye idéer i en sak.

---

Samlet sett vurderer departementene evalueringene som svært nyttige innspill ved at de gir økt innsikt gjennom at de foretar en samlet vurdering – noe den løpende saksbehandlingen sjelden gir anledning til.

Departementene fremhever at hvis evalueringer skal være nyttige i forbindelse med det nye økonomireglementet må en rekke forutsetninger oppfylles: For det første må tiltakenes målsettinger være klart definert. For det andre må målene være slik formulert at det er mulig å avgjøre om målene faktisk er nådd. For det tredje må evalueringsmandatene springe ut av de tiltak som skal evalueres og den beslutningssammenheng de skal inngå i. For det fjerde må evalueringene dokumentere konkrete effekter av de tiltak som vurderes. For det femte må konklusjoner og anbefalinger følge av godt dokumenterte analyser. For det sjette må anbefalingene være realistiske i forhold til tiltakets rammebetingelser.

## 5.1 Innledning

Norsk institutt for studier av forskning og utdanning, NIFU, har på oppdrag av Statskonsult gjennomført en kartlegging av departementenes evalueringspraksis. Kartlegging omfatter de evalueringer departementene har gjennomført i perioden 1994–1996 slik de fremkommer av departementenes rapportering til NIFU. I alt har NIFU registrert 206 evalueringer i den aktuelle perioden. Registreringen omfatter imidlertid ikke alle evalueringer, da enkelte avdelinger i enkelte departementer av ulike grunner ikke har rapportert tilbake til NIFU. Vi er imidlertid overbevist om at vi har fanget inn godt over 90 % av alle evalueringene. Den følgende presentasjonen burde derfor gi et dekkende bilde av departementenes evalueringsvirksomhet.

Presentasjonen er delt i to hovedbolker. I den første, avsnittene 4.2 og 4.3, foretar vi en beskrivelse av departementenes evalueringspraksis med utgangspunkt i alle de 206 evalueringer som er identifisert. Her legges det vekt på hvor store ressurser som brukes totalt og fordelt på departementene, videre hva slags type evalueringer som er utført (om det er måloppfyllelsesanalyser, prosessanalyser eller effektanalyser), og hvilken tilknytning evalueringene har til departementenes øvrige virksomhet som strategiutvikling, budsjettering, etc.

I den andre hovedbolken, avsnittene 4.4 til og med 4.8, går vi nærmere inn på enkelte av evalueringene. Av de 206 evalueringer som er registrert, har vi trukket tilfeldig 100 stykker som er blitt gjenstand for en nærmere oppfølging gjennom en spørreskjemaundersøkelse til de «prosjektansvarlige» i departementene. NIFU har fått data om 90 av evalueringene. Vi er således av den oppfatning at datagrunnlaget gir et dekkende bilde av departementenes evalueringsvirksomhet. I denne delen av presentasjonen er vi ikke opptatt av det enkelte departement, men hvordan departementene samlet håndterer evalueringsvirksomheten.

---

I presentasjonen vil vekten bli lagt på evalueringenes organisatoriske forankring i departementene: hvem som tar initiativet, hvilke formål evalueringene skal ha, hvorvidt evalueringsobjektene er tematisk følsomme, hvem som formulerer mandatene, og om det er laget planer for en systematisk oppfølging av evalueringresultatene når disse foreligger.

Videre vil vi beskrive departementenes deltakelse i selve gjennomføringen av evalueringene: har man etablert styringsgrupper, deltar man i utformingen av evalueringsopplegget, har man kontakt underveis etc. I den avsluttende delen vil vi gjøre rede for den bruk departementene har gjort av de evalueringer de har bestilt: omfanget av bruken, har man brukt konklusjonene, eventuelt anbefalingene, til hva har man brukt dem? Dette leder frem til en presentasjon av departementsrepresentantenes vurdering av brukbarheten av evalueringer som et beslutningsgrunnlag i departementenes virksomhet og hvilke forutsetninger som da må være tilstede.

NIFU vil understreke at denne undersøkelsen *ikke* er en evaluering av departementenes evalueringspraksis, men en beskrivelse av denne basert på departementsrepresentantenes egne oppfatninger av de sider ved evalueringsevirsomheten som er presentert i avsnittene 4 til og med 8.

## **5.2 Departementenes evalueringspraksis – et overblikk**

### **5.2.1 Omfang**

Av innledningen fremgår det at departementene har gjennomført 206 evalueringer i perioden. Til dette har det medgått omlag 180 millioner kroner. Vi har imidlertid kun informasjon om 189 evalueringer slik at tallet nok er en god del høyere, sannsynligvis godt over 200 millioner kroner. Departementene bruker således ikke ubetydelige summer på denne virksomheten. Tabell 1 på neste side gir et inntrykk av hvordan omfanget fordeler seg på det enkelte departement.

Tabell 1: Budsjett fordelt etter departement for “populasjon” (registreringsskjema) og for utvalg (spørreskjemaundersøkelsen).

DEÅ	POPULASJON			UTVALG		
	Antall Missing i () *	B u d s j e t t		Antall Missing i () *	B u d s j e t t	
		Kr	Gj.snitt		Kr	Gj.snitt
UD	17	16.317.700	959.865	8	7.737.100	967.138
NoE	12	6.940.000	578.333	6	4.420.000	736.667
AD	7	2.190.000	312.857	4	1.660.000	415.000
BFD	7	19.075.000	2.725.000	3	3.771.500	1.257.167
KUF	29	73.992.000	2.551.448	14	40.463.000	2.890.214
LD	7	940.000	134.286	3	240.000	80.000
SD	10	4.961.600	496.160	5	2.629.600	525.920
KD	6 (-4)	990.000	165.000	2 (-2)	350.000	175.000
KAD	52 (-3)	32.269.800	620.573	21 (-1)	17.195.000	818.810
MD	13	3.430.000	263.846	4	530.000	132.500
FD	5	2.114.200	422.840	2	1.132.600	566.300
FID	2 (-2)	700.000	350.000	1 (-1)	400.000	400.000
SHD	17 (-1)	11.308.200	665.188	10	7.091.700	709.170
JD	5 (-7)	3.278.000	655.600	2 (-1)	1.143.000	571.500
<b>SUM</b>	189 (-17)	178.506.500	944.479	85 (-5)	88.763.500	1044.276

\* Tallene i parentes i kolonnen utgjør antall evalueringer hvor vi ikke har informasjon om budsjettall.

Enkelte departementer peker seg klart ut med hensyn både til antall evalueringer og penger som brukes til denne virksomheten. Det økonomisk tyngste engasjementet står KUF for med 73 millioner kroner, fulgt av KAD med 32 millioner kroner deretter følger BFD og UD. Det er grunn til å understreke at SHD sannsynligvis er grovt underrepresentert i denne tabellen i og med at vi bare har informasjon fra en avdeling. Tabellen viser også at evalueringene gjennomgående er relativt små. Over 60 prosent av dem koster under 600.000 kroner. På den annen side finner vi enkelte meget omfattende evalueringer som koster fra 10 millioner til nesten 20 millioner kroner.

De innledende konklusjoner er således at evalueringer har fått et betydelig omfang i sentraladministrasjonen generelt, men at det er store variasjoner mellom departementene både med hensyn til antall gjennomførte evalueringer og den ressursmessige innsatsen som legges ned i denne virksomheten.

## 5.2.2 Evalueringstyper

Begrepet evaluering er ikke entydig. Det legges ofte ulikt meningsinnhold i begrepet og det finnes mange ulike typer evalueringer avhengig av hva som står i fokus for oppmerksomheten. I denne omgangen vil vi la begrepsdiskusjonen ligge og konsentrere oss om hvordan departementenes evalueringspraksis ser ut i forhold til ulike evalueringstyper.

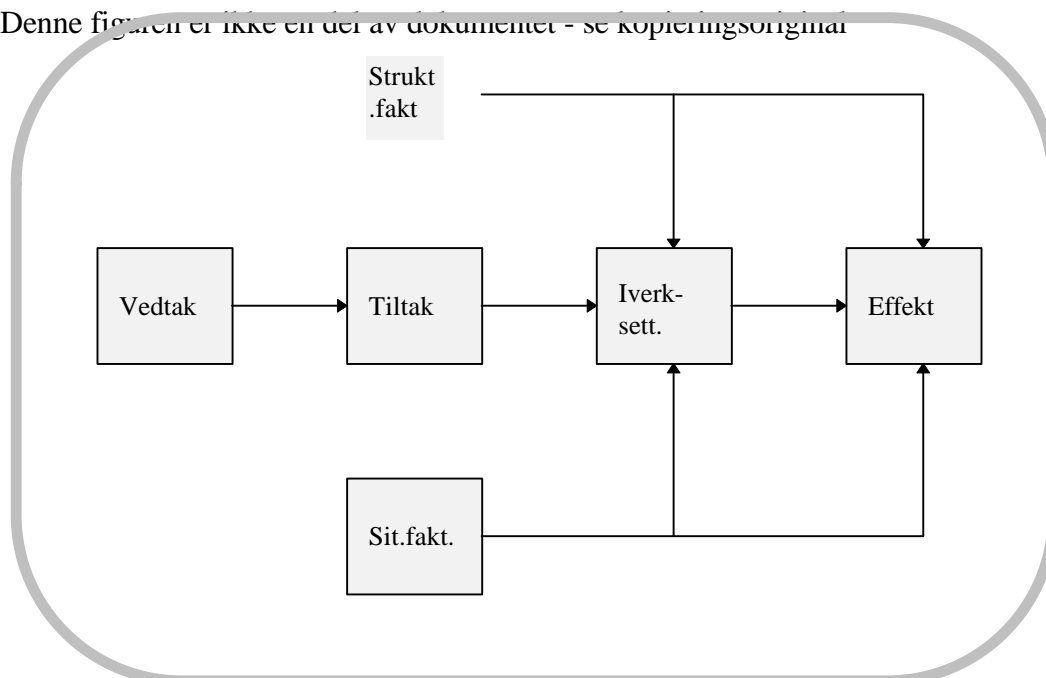
Modellen i figur 1 inneholder 6 hovedrelasjoner. Disse gir opphav til 6 hovedtyper av evalueringer slik det er fremstilt i tabell 2 under.

Tabell 2: Evalueringstyper

EX-POST ANALYSE	
Evalueringsspørsmål	Analysetype
1. Hvilke tiltak for å nå målet?	Tiltaksanalyse
2. Hva ble resultatet?	Effektanalyse
3. Hvilke faktorer bidro til resultatet?	Årsaksanalyse
4. Hvorfor ble ting som de ble?	Iverksettingsanalyse
5. Samsvarer resultat og mål?	Måloppfyllelse
6. Var tiltakene effektive?	Effektivitetsanalyse

### Figur 1: Evalueringstyper

Denne figuren er ikke en del av dokumentet - se kopieringsoriginal



I denne undersøkelsen har vi inkludert følgende evalueringstyper. Tallene viser hvor mange prosent av evalueringene som faller i hver kategori. En evaluering kan inneholde flere elementer:

Målevaluering	12,1
Prosessevaluering	28,2
Effekt/konsekvensevalueringer	39,8
Virkemiddelsevaluering <sup>1</sup>	40,3
Brukerundersøkelser	23,3

Fordelingen viser klart at hovedtyngden av evalueringene er knyttet opp til virkemiddelsanalyser og effektevalueringer. Dette er som en skulle vente gitt departementenes oppgaver som iverksettere av offentlig politikk. I tabell 3 har vi gått nærmere inn på variasjonene mellom departementene med hensyn til hva slags evalueringer som gjennomføres.

*Tabell 3: Departementenes evalueringsprofiler. Evalueringstyper fordelt på departementer. Prosent. En og samme evaluering kan omfatte flere evalueringstyper. Prosentsummene kan derfor overstige 100.*

	måloppnåelse	prosess	effekt/konsekvens	virkemiddel	brugerundersøkelse	N
UD	47	18	41	47	18	17
NOE	8	0	8	50	33	12
AD	14	14	71	29	14	7
BFD	0	0	29	100	0	7
KUF	14	48	35	41	7	29
LD	40	40	100	40	40	5
SD	40	40	0	20	10	10
KD	0	0	50	20	40	10
KAD	6	26	57	44	15	54
MD	0	54	46	62	23	13
FD	0	50	50	0	0	2
FID	0	0	50	50	0	4
SHD	10	70	40	30	80	10
JD	9	46	27	46	64	11

<sup>1</sup> Både virkemiddelsevalueringer og brukerundersøkelser er problematiske i denne sammenhengen. Av figur 1 ser vi klart at dette ikke er egentlig evalueringstyper i ren forstand. Virkemiddelsevalueringer omfatter som regel en effekt- og en måloppnåelsesanalyse. Brukerundersøkelser vil som regel være en «spørreskjemaundersøkelse» som leder frem til en effektanalyse. Med andre ord dette er mer karakterisering av en datainnsamlingsmetode enn en karakterisering av evalueringstype. Når disse er tatt med, er det fordi de ofte brukes som begreper for evalueringstyper.



---

Tabellen viser noen interessante forskjeller i evalueringsprofiler mellom departementene. Tre departementer skiller seg klart ut når det gjelder omfanget av målanalyser: UD, LD og SD. Hvorfor det er slik, er ikke godt å si. En mulig forklaring kan være at dette er departementer som i stor utstrekning står i offentlighetens lys. UD på grunn av den store interessen knyttet til bistands- politikken, LD og SD på grunn av den sentrale plass landbruks- og samferdsels- politikken inntar på den politiske arena.

Det er videre interessant å registrere at en rekke departementer legger vekt på prosessevalueringer til tross for det relativt beskjedne omfang slike evalueringer har totalt sett. Her er det særlig SHD, MD, FD og JD som peker seg ut. Flere av dem er svært beskjedne med hensyn til antall evalueringer, men når de først gjennomfører en evaluering, så er man svært fokusert på prosessevalueringer. Også dette er rimelig tatt i betraktning at disse departementene har vært og til dels er involvert i forsøk, reformer og reorganiseringsprosesser.

De fleste departementene er orientert mot effektevalueringer. Det er bare SD som ikke plasserer noen av dets evalueringer i denne kategorien. Heller ikke NOE har lagt særlig vekt på slike sider. Dette kan imidlertid skyldes at dette departementet subsummerer slike evalueringer under virkemiddevalueringer. Dette er også klart en mulig forklaring når det gjelder SD. En liknende kopling kan finnes for FD. I det hele bør disse to kategoriene ses under ett.

### **5.2.3 Tilknytning til departementale prosesser**

I foregående avsnitt har vi antydnet at noe av den variasjonen som der observeres, kan muligens skyldes at evalueringer er knyttet til ulike virksomheter i de forskjellige departementene

*Tabell 4: Evalueringenes tilknytning til arbeidsoppgaver fordelt på departementene. Prosent.*

	strat	bud	stort	plan	til skudd	reform	annet
UD	55	0	0	0	0	0	45
AD	14	14	14	0	0	43	14
BFD	14	0	0	0	0	29	57
KUF	17	17	0	13	4	42	8
LD	29	58	0	0	14	0	0
SD	0	0	57	0	0	43	0
KD	0	0	0	0	29	43	29
KAD	23	9	11	0	23	17	17
MD	13	25	0	38	0	0	13
FD	0	20	0	0	0	80	0
FID	100	0	0	0	0	0	0
SHD	0	20	10	70	0	0	0
JD	8	17	0	25	8	25	17

En slik forklaring finner en viss støtte i tabell 4. To departementer skiller seg ut når det gjelder å bruke evalueringer i forbindelse med strategiutvikling: UD og FID. Disse to bruker evalueringene nesten utelukkende i den sammenheng. Riktignok plasseres 45% av UD's evalueringer under "annet", men det kommer sannsynligvis av departementets ansvar for bistandspolitikken med dets særegne karakter i forhold til øvrig departementsvirksomhet.

Den mest interessante observasjonen i forhold til denne konferansens formål, er den relativt svake koplingen mellom evalueringsvirksomhet og budsjettarbeidet. Det er kun LD som skiller seg ut i denne sammenhengen ved at hele 58% av evalueringene er knyttet til budsjettarbeidet. Det er tydelig at evalueringer til nå ikke har vært et aktivt hjelpemiddel i departementenes budsjettarbeid.

Samferdselsdepartementet er det departement som i perioden i størst grad er pålagt å gjennomføre evalueringer fra Stortingets side. Utover SD er det bare AD og SHD som har fått seg pålagt en slik eksersis.

SHD skiller seg klart ut når det gjelder å bruke evalueringer i forbindelse med plan- og FoU-arbeid. Hele 70 % av departementets evalueringsvirksomhet er knyttet opp til dette. Her må man imidlertid huske på at datagrunnlaget er kun hentet fra en avdeling, og det er den som er ansvarlig for departementets forsøksvirksomhet. Men i den avdelingen er det tydelig at evalueringer er et aktivt redskap i oppfølgingen av avdelingens virksomhet. Forøvrig er det bare MD og JD som nytter evalueringer som et planleggingsverktøy i noe omfang.

---

Når det gjelder å bruke evalueringer i forbindelse med forvaltningen av tiltak/tilskudd er det KD og KAD som gjør dette i noe omfang. For de andre departementenes vedkommende er en slik bruk langt mer sporadisk, om det i det hele tatt brukes.

Det tyngste bruksområde for evalueringene i departementet er i forbindelse med omorganisering og reformvirksomhet. Her er det tydelig at evalueringer aktivt brukes i oppfølgingsarbeidet. En rekke departementer er involvert i slik virksomhet. Særlig har FD vært aktive, men AD, KUF, SD og KD ligger ikke langt etter.

### 5.3 Valg av evaluator

Hovedtyngden av evalueringsoppdragene utføres innen instituttsektoren. Hele 120 ganger har forskningsinstitutter vært involvert. Dette er over dobbelt så hyppig som den nest største aktørgruppen, konsulentfirmaer med 56 ganger. Deretter følger universitets- og høyskolesektoren med 28 ganger knapt foran kategorien "annet", som består av en heterogen blanding av enkeltindivider, ikke-identifiserbare grupper og statlige organer. Hvis vi ser nærmere på instituttgruppen, viser det seg at det er seks institutter som fremstår som særlig aktive på evalueringsfronten:

Norsk institutt for by- og regionsforskning (NIBR)	13 evalueringer
Norsk institutt for studier av forskning og utdanning (NIFU)	12 «
Institutt for samfunnsforskning (ISF)	9 evalueringer
Norges byggforskningsinstitutt (NBI)	9 «
Stiftelsen for samfunns- og næringslivsforskning (SNF)	7 «
FAFO - Forskningsstiftelsen	6 «

Disse seks instituttene står for omlag 50 % av all instituttdeltakelsen i departementale evalueringer i perioden. Mer interessant er det at det synes å være et systematisk mønster med hensyn til hvilke institutter som utfører evalueringer for hvilke departementer. NIBR har spredd seg i hovedsak på to departementer KAD og MD, NIFU er fokusert på KUF, mens FAFO, SNF, ISF og NBI er knyttet opp til KAD. Her er det interessant å registrere at det er en ytterligere arbeidsdeling hvor eksempelvis ISF synes å ha en spesielt nær kontakt med arbeidsavdelingen og NBI en tilsvarende kontakt med bolig- og bygnings-avdelingen. Det ser således ut til at departementene har "sine" evaluatorene og at de tunge evalueringsmiljøene har hovedtyngden av departementale evalueringsoppdrag i ett departement. Dette betyr selvsagt ikke at det enkelte departement/ avdeling kun benytter seg at ett institutt eller at det enkelte institutt kun forholder seg til ett departement. Men det betyr at det har utviklet seg et nettverk som det trekkes veksler på når evalueringsoppdrag skal fordeles – noe som i og for seg er en svært rasjonell fremgangsmåte.

Dette nettverksmønsteret bekreftes indirekte ved at omlag to tredjedeler av evalueringsoppdragene som inngår i vårt utvalg på 90 evalueringer, fordeles direkte til oppdragstaker og ikke gjennom en anbudsrunde.

---

Hvis nettverksantakelsen er riktig, skulle vi forvente at departementene legger større vekt på at de potensielle oppdragstakerne har kunnskap og kompetanse innen den sektor evalueringen skal gjøres og at de har erfaringer med den type aktiviteter, enn at det legges vekt på prosjektforslagene utforming, evaluatorenes dokumenterte faglige dyktighet, pris og tidsbruk. Figur 2 gir et inntak til denne problemstillingen.

## **Figur 2: Valg av evaluatorene – kriterier**

Denne figuren er ikke en del av dokumentet - se kopieringsoriginal

Figurens hovedtendenser går klart i den forventede retningen idet det legges størst vekt på evaluatorenes kunnskap og erfaring. Minst vekt legges det på pris og tidsbruk. Dette mønsteret forsterkes hvis vi skiller mellom de som gir oppdragene direkte og de som bruker anbudsmetoden. Her viser det seg at ved anbudsmetoden legger 39% av respondentene vekt på prosjektforslagene utforming, i motsetning til ved den direkte metoden hvor kun 1,7% tillegger dette vekt. På samme måte finner vi en forskjell når det gjelder pris og tidsbruk, hvor en langt større andel vektlegger dette ved tilbud enn ved den direkte metoden. Den direkte metoden vektlegger derimot langt sterkere kunnskap og erfaring.

Det er således liten grunn til å tvile på at nettverk spiller en viktig rolle ved departementenes evalueringspraksis. Nettverksmetoden har den fordel at den trekker veksler på den substansielle innsikt forskningsinstituttene innen sektoren sitter inne med når en evaluering skal gjennomføres. Ulempen kan selvfølgelig over tid bli at instituttene til de grader integreres i slike nettverk at de ikke lenger oppfattes å stå uavhengig av oppdragsgiver. Foreløpig er det imidlertid ingen ting som tyder på at en slik situasjon har oppstått. Til det er det for mange aktører involvert. Men man skal være oppmerksom på de farer en eventuell "monopolisering" kan innebære.

---

## 5.4 Evalueringenes organisatoriske forankring

En rekke studier har vist at det er viktig at evalueringer forankres i den organisatoriske enhet som skal bruke resultatene. Studiene viser entydig at det er helt avgjørende for fremtidig bruk at initiativet til evalueringene springer ut av bruker-enheten selv, at formålet med evalueringene er avklart i forkant av evalueringene og at mandatutformingen ikke overlates til andre enn de fremtidige brukerne. Likeledes er det viktig at brukerne av evalueringsresultatene er identifisert i forkant og at en formidlingsplan er utarbeidet.

### 5.4.1 Initiativ

Som man skulle vente, er det fagavdelingene som i hovedsak er initiativtakerne til evalueringene. Fagavdelingene og fagkontorene er initiativtaker i nær 70 % av tilfellene. Det er imidlertid interessant å registrere at den politiske ledelse selv tar initiativet til drøye 10 % av evalueringene. Vi ville forvente at den politiske ledelse ville være særlig aktive initiativtakere hvis saksområdet under diskusjon oppfattes å være politisk følsomt. Så er ikke tilfelle. Det er kun i forbindelse med igangsettelse av evalueringer av reformer at den politiske ledelse markerer seg. Slike reformer oppfattes da også som å være politisk følsomme. Men selv i slike tilfeller viser det seg at det er fagavdelingene som fremtrer som de store initiativtakerne.

I motsetning til det man finner når det gjelder forskningsevalueringer, er det meget sjelden at utenforstående er initiativtaker. Her er det imidlertid et klart unntak. Stortinget har både overfor AD, KAD, SHD og særlig SD gitt pålegg om å gjennomføre evalueringer.

### 5.4.2 Formålet med evalueringene

Det er tydelig at departementene har et bevisst forhold til egen evalueringsportefølje. Vi har kun registrert 5 tilfeller hvor evalueringenes formål ikke eksplisitt er formulert i forkant av eksersisen.

*Tabell 5: Departementenes evalueringsportefølje fordelt etter evalueringenes formål*

	prosent	N
organisasjon	21.1	19
program	13.3	12
tiltak	18.9	17
virkemiddel	10.0	9
tilskudd	12.1	11
reform	16.7	15
regelverk	3.3	3
annet	28.9	26

---

Tabellen viser evalueringene fordelt på objekttyper. Evalueringer av organisasjoner inntar en fremtredende plass. Sammen med evalueringer av tiltak/ tilskottsordninger og reformer er dette de hyppigste evalueringsobjektene.

*Tabell 6: Evalueringenes tilknytning til departementenes kontrollansvar*

	prosent	N
i stor grad	41.2	35
i noen grad	31.8	27
i liten grad	14.1	12
ingen tilknytning	12.9	11
	100	85

Av tabell 6 fremgår det at departementene i stor grad knytter evalueringene til det kontrollansvar de har i forbindelse med iverksetting av offentlige tiltak. Det er kun i et tyvetalls tilfeller at så ikke er tilfelle, og da er evalueringene i hovedsak gjennomført i forbindelse med departementenes FoU/ planleggingsvirksomhet.

At evalueringer inngår som et rutinisert redskap i departementenes virksomhet, fremgår også av det forhold at det overveiende flertall av evalueringene (63,2%) oppfattes å være knyttet til rutinevirksomhet.

### **5.4.3 Identifiserte brukere**

Bare unntaksvis (6 tilfeller) er evalueringene satt i gang uten at initiativtakerne først har identifisert hvem som skal være mottakerne og brukerne av evalueringens resultater. I de aller fleste tilfellene er det sammenfall mellom initiativtakerne til evalueringene og mottakerne/brukerne av evalueringens resultater. Det er således fagavdelingene som i hovedsak tar initiativet og det er også disse som er brukerne.

Tatt i betraktning at det er initiativtakerne som også er brukerne, er det noe overraskende at det for 40% av evalueringenes vedkommende utarbeides konkrete formidlingsplaner. En nærmere undersøkelse viser at dette i hovedsak skjer når flere enn initiativtakeren er involvert på brukersiden og særlig hvis evalueringen i stor grad angår departementets kontrollansvar – man rapporterer oppover.

### **5.4.4 Konklusjon**

Departementene har åpenbart et meget bevisst og ryddig forhold til egen evalueringsevne. Evalueringene er godt organisatorisk forankret i departementene ved at det er fagavdelingene som står ansvarlig for både initieringen av evalueringene,

---

mandatutformingene og som brukere av evalueringens resultater. Det er således åpenbart at man har et meget bevisst forhold til evalueringer som et verktøy. Det er kun i noen få tilfeller at man ikke i forkant har en meget klar oppfatning av hva formålet med den enkelte evaluering skal være. I hovedsak er evalueringene knyttet til departementenes kontrollansvar som politisk-administrative forvaltningsorganer.

## 5.5 Departementenes oppfølging i gjennomføringsfasen

Tidligere studier har vist at oppdragsgivernes utbytte av evalueringene til en stor grad avhenger av at ikke bare mandatene utformes i forhold til tiltenkt bruk, men at det er nødvendig med en nær oppfølging av selve evalueringsarbeidet slik at man påser at evalueringssopplegget samsvarer med de intensjoner oppdragsgiveren hadde i utgangspunktet.

Nesten alle evalueringene har en formell kopling til oppdragsgiver enten ved at det opprettes en styringsgruppe (13,5 %), en referansegruppe (24,7 %), eller utpeker en kontaktperson (33,7 %). I en rekke tilfeller har man også kombinert eksempelvis det å opprette en egen referansegruppe med å utpeke en særskilt kontaktperson med ansvar for å følge opp evalueringen. Bare i 5 % av evalueringene har man ikke noen formell kontakt. En ting er at man oppretter forbindelser mellom oppdragsgiver og utfører, noe helt annet kan være om kontakten utnyttes. Departementene rapporterer at i nær to tredjedeler av evalueringene har man hatt en løpende kontakt med evaluatorene, og i nesten samtlige av disse vurderer departementene kontakten som meget god eller god.

Departementene har åpenbart vært svært aktive i selve oppstartingen av evalueringene.

*Tabell 7: Departementenes deltakelse i gjennomføringen av evalueringene. Prosent (antall observasjoner)*

	utforming av opplegget	deltok underveis i gjennomføringen
i stor grad	40.4 (36)	5.6 (5)
	44.9 (40)	26.7 (23)
deltok ikke	14.6 (13)	67.4 (58)

I omlag 40 % av evalueringene hevder departementene at de i stor grad har deltatt i utforming av selve evalueringssopplegget. I ytterligere 45 % av tilfellene har de deltatt i noen grad. Det er således klart at departementene har vært seg bevisst behovet for en rimelig tett oppfølging i oppstartingsfasen av evalueringssopplegget for å sikre et godt samsvar mellom intensjoner og gjennomføring.

---

Derimot er det de færreste som følger nært opp i selve gjennomføringen av evalueringene.

Omlag to tredjedeler bryr seg ikke med gjennomføringen etter at opplegget er fastlagt. Blant de som følger opp også denne fasen er kontakten mer sporadisk.

Når evalueringsarbeidet nærmer seg slutten, blir departementene igjen mer aktive. I omlag tre fjerdedeler av evalueringene leser departementene utkast til rapporten og gir kommentarer til utkastene. Denne virksomheten finner departementene meget nyttig i og med at de får anledning til å rette opp faktiske feil, uten at de blander seg bort i de vurderinger som gjøres i evalueringene.

## 5.6 Vurdering av rapportene

Noen av forutsetningene for at en evalueringsrapport skal brukes er at den kommer til rett tid, at den gir svar på de problemstillinger som er reist i mandatet og at den evalueringsfaglig er til å stole på.

I motsetning til hva man hevder om forskning generelt så kommer det overveiende flertall av evalueringene i rett tid i forhold til det som er avtalt (84,2 %). Dette utgjør således ikke noe hinder for bruk av resultatene. I de aller fleste tilfellene (95,8 %) er departementene også av den oppfatning at evalueringene gir svar på de problemstillinger som er reist i mandatene. I omlag to tredjedeler er det et godt samsvar og for 28 % av evalueringene er samsvaret meget godt.

Når det gjelder den evalueringsfaglige vurderingen av rapportene, er departementene bedt om å vurdere rapportenes datagrunnlag, i hvilken grad konklusjonene er empirisk godt underbygget, og i hvilken grad eventuelle anbefalinger følger av de konklusjoner som er trukket. Fordelingen av vurderingene er presentert i tabell 8.

*Tabell 8: Departementenes vurdering av rapportene. prosent (antall observasjoner)*

	datagrunnlag	konklusjonenes empiriske fundament	om anbefal. følger av konklusjonene
meget godt	19.8 (16)	21.6 (16)	18.8 (12)
godt	66.7 (50)	67.6 (50)	53.1 (34)
mindre godt	13.6 (11)	10.8 (8)	4.7 (3)
ikke gitt anbefalinger			23.4 (15)

Hovedinntrykket er overveiende svært positivt. Det store flertallet av evalueringene bedømmes til å ha et godt til dels meget godt datagrunnlag, de konklusjoner som



---

trekkes har et solid empirisk fundament og anbefalingene oppfattes å følge av konklusjonene. Når det er sagt, er det likevel åpenbart et stort forbedringspotensiale i det kun omlag 20% av evalueringene vurderes til å ha et svært godt datagrunnlag, at konklusjonene er meget godt underbygget og at anbefalingene i meget stor grad følger av konklusjonene. På den annen side er det kun de færreste som bedømmes som lite tilfredsstillende langs disse dimensjonene.

Konklusjonen må således være at departementene vurderer evalueringene til å holde et godt faglig nivå – de er solide og man kan derfor legge resultatene til grunn i det videre arbeid. Når man samtidig fremholder at de gir svar på de problemstillinger som er reist og at evalueringene kommer til rett tid, er det ingen grunn til ikke å ta resultatene i bruk.

## 5.7 Bruken av evalueringsresultatene

### 5.7.1 Omfang

Når det gjelder bruk av evalueringsresultatene, har vi delt dette i to ved å skille bruk av konklusjonene og bruk av anbefalingene. Når vi tar hensyn til at 19 av evalueringene fortsatt er under arbeid og følgelig ikke har avlevert sluttrapport, så viser det seg at i hele 87 % av evalueringene har man brukt konklusjonene. Hvis vi videre tar hensyn til at det i flere av evalueringene ikke fremmes anbefalinger (21 stykker), er bruksprosenten for anbefalingene 84. Med andre ord er bruksfrekvensen av evalueringsresultatene svært høyt. Det er også slik at hvis man først bruker anbefalingene, så vil man også samtidig bruke konklusjonene mens vi finner enkelte tilfeller hvor konklusjonene brukes, men ikke anbefalingene. Det er således ikke tvil om at evalueringene inngår som et aktivt verktøy og ikke kun som en symbolsk handling hvor evalueringsresultatene ikke følges opp.

### 5.7.2 Brukstyper

I grove trekk kan evalueringer brukes på tre hovedmåter: Instrumentelt ved løsning av konkrete oppgaver; begrepsdannende ved å påvirke vår måte å oppleve verden på; og politisk/strategisk/ symbolsk som ammunisjon/ argumentasjon i utformingen av retningslinjene for samfunnsutviklingen.

Naustdalslid/Reitan (1992)<sup>2</sup> har i boka "Kunnskap og styring" gitt følgende definisjoner av de ulike brukstypene:

**Instrumentell:** Måltrettet bruk av forskningsresultater eller forskningsbasert informasjon til å løse konkrete problemer for en organisasjon eller en beslutningstaker.

---

<sup>2</sup> Naustdalslid, J. og Reitan, M. (1992) *Kunnskap og styring*, NIBR, rapport 1992:15

---

**Begrepsdannende:** Forskningen brukes bevisst eller ubevisst til å forme begreper om verden, våre oppfatninger om årsaker og virkninger og er med på å skape de problemoppfatninger vi omgir oss med.

**Kunnskapsberedskap:** Forskning innen en enhets ansvarsområde uten at den er satt i forbindelse med en konkret sak eller et problem

**Politisk bruk**

**Strategisk:** Forskning blir brukt av en eller flere parter for å oppnå egne målsettinger. Legitimering av egne posisjoner

**Symbolisk:** Forskning initieres uten tanke på å skulle brukes. "Kjekt å ha". For å vise at man er et rasjonelt menneske. Forskning gir status.

I figur 3 presenteres fordelingen på de ulike brukstypene. Som det fremgår av figuren, har vi operasjonalisert brukstypene ved å spørre respondentene om de hadde brukt den enkelte evaluering på følgende måter og presenterte en liste med muligheter som så skulle krysses av. De tre første bruksmåtene er alle knyttet til løsningen av en konkret sak. Vi har derfor latt dette representere instrumentell bruk. Det å bruke evalueringer til å skaffe seg mer kunnskap om eget saksfelt og å vurdere et virkemiddel generelt er uttrykk for en kunnskapsberedskapsbruk, mens det å legitimere pågående virksomhet er vår operasjonalisering av en politisk bruk av evalueringresultatene. I tillegg og som en egen kategori, har vi spurt om evalueringene har vært brukt i sammenheng med departementenes økonomistyringssystem.

**Figur 3: Bruk av evalueringer**

Denne figuren er ikke en del av dokumentet - se kopieringsoriginal.

---

Figuren viser et mangfold av bruksmåter hvor det er tydelig at det å bruke evalueringene til å forbedre et eksisterende tiltak er en viktig bruksmåte sammen med det å få nye idéer i en sak. Det å bruke evalueringer for å legitimere den pågående virksomhet er derimot ikke særlig utbredt, selv om det åpenbart forekommer. Den bruksmåten som peker seg ut ved å nesten ikke forekomme, er bruken av evalueringer i forbindelse med økonomistyringssystemer. Dette burde være en tankevekker og en motiverende faktor for det videre arbeid med å utvikle et økonomistyringssystem hvor evalueringer inngår som et aktivt verktøy. Nå bør det legges til at selv om man i dag i liten grad finner noen koplinger mellom evalueringsevirsomhet og økonomistyring, så indikerer tabellen at evalueringer ikke er et ukjent verktøy i forhold til spørsmål som står helt sentralt i et økonomistyringssystem – eksempelvis det å vurdere alternative tiltak og forbedre eksisterende tiltak.

## **5.8 Evalueringer som et verktøy i økonomistyringen**

### **5.8.1 Nytten av evalueringene**

Godt over 90 % gir uttrykk for at evalueringene har vært av stor verdi som et beslutningsgrunnlag i det aktuelle tilfellet. Begrunnelsene er først og fremst knyttet til at evalueringene gir beslutningene legitimitet, innsikt og at de gir beslutningene et metodisk fundament.

Når det gjelder legitimitet, fremheves særlig at det er uavhengige eksterne aktører som kommer med sine synspunkter. Dette representerer uavhengige faglige vurderinger som oppfattes å ha en større faglig tyngde enn det departementet representerer. Flere departementer fremhever også at de eksterne evalueringene representerer et nyttig "korrekt" i forhold til departementet ved at en uhildet "tredje part" uttaler seg om sakskomplekset.

Evalueringene gir en økt innsikt ved at de i større grad foretar en samlet vurdering, noe den løpende saksbehandlingen sjelden gir anledning til. Særlig fremheves det at evalueringene gir mer systematisk kunnskap om effekter enn det departementene vanligvis fanger inn og hvor observasjonene blir mer tilfeldige. På den måten virker evalueringene også supplerende i forhold til den kunnskap departementet selv sitter inne med. Evalueringenes store styrke er at de ved siden av å beskrive tilstander også kan få frem hvilke tiltak/mekanismer som skaper disse tilstandene. Evalueringene gir således en dypere innsikt og gir et grunnlag for å (videre)utvikle tiltak som har de rette effekter.

Metodespørsmål står sentralt i begrunnelsen for hvorfor evalueringer oppfattes å ha vært nyttige. Solid gjennomførte evalueringer gir både informasjon man kan stole på og som er knyttet til den aktuelle problemstillingen. De representerer således en systematisk erfaringsinnsamling og gir et dokumentert beslutningsgrunnlag. Det er her også innvendingene mot evalueringene dukker opp. Hvis ikke det metodiske apparat

---

er på plass, gjør evalueringene bare vondt verre ved at de representerer en synsing som gir seg ut for å være noe mer.

### **5.8.2 Nødvendige forutsetninger målstyring**

Den forutgående gjennomgangen viser at departementene i sin evalueringsevne følger et mangfold av evalueringstyper og et mangfold av bruksformer. Dette mangfoldet reflekteres også i deres svar på spørsmålet om hvilke forutsetninger som må være tilstede for at en evaluering skal være nyttig for et departement i en målstyringssammenheng. Til tross for dette mangfoldet er det visse mønstre i departementenes svargivning. Disse mønstrene vil vi kort gjøre rede for i de følgende avsnittene.

I alt hadde 71 av de 90 respondentene oppfatninger av hvilke forutsetninger som må være tilstede. Svarene grupperte seg i to dominerende kategorier: forutsetninger knyttet til initieringen av evalueringene og forutsetninger som knytter seg til gjennomføringen av dem, henholdsvis 43 og 46 respondenter. I tillegg er det noen få, åtte stykker, som knytter forutsetningene til departementenes oppfølging av evalueringene.

I initieringsfasen fremheves det at det er avgjørende at departementene vet hva de skal bruke evalueringresultatene til. Det innebærer blant annet at det i forkant skal være avklart i hvilken beslutningskontekst evalueringene skal inngå i, og at det er klart hvem som skal bruke resultatene. Uten en slik forankring vil det lett bli til at evalueringene ikke skreddersys godt nok til målstyringsproblematikken og blir noe som er frikoplet dette.

Flere av respondentene tar opp spørsmålet om alle sider ved målstyring er like egnet som objekter for evalueringer. De peker på at det er visse forutsetninger som må være til stede. For det første forutsetter det at tiltakenes målsettinger er rimelig klart definert slik at man har en referanseramme å holde tiltakenes effekter opp mot. Ikke bare må de være klart definert, men de må også være operasjonaliserbare slik at det lar seg undersøke hvorvidt målene faktisk er nådd. En slik effektorientering forutsetter også at tiltakene har fått virke over noe tid slik at det er rimelig å forvente at tiltakene har fått tid til å få effekter. Her hevdes det at det syndes i dagens praksis – man er for utålmodige. Målstyring forutsetter således en annen tidshorisont enn den tradisjonelle regelstyringen. Denne innsikten er ikke alltid tilstede.

En rekke respondenter påpeker at den, etter deres vurdering, største svakheten med dagens praksis er at evalueringsmandatene ikke er tilstrekkelig presist utformet i forhold de informasjonskrav et gjennomført målstyringssystem stiller. Det bør derfor innskjerpes at mandatene må springe ut av de tiltak som skal evalueres og den beslutningskontekst evalueringene skal inngå i. Hvis ikke, vil evalueringene lett frikoples målstyringsprosessen. Det er således departementene som har ansvaret for at

---

evalueringene blir et verktøy i et målstyringsperspektiv. Dette er ikke evaluators ansvar.

Når det gjelder mandatenes innhold, vil disse naturligvis variere avhengig av hva slags tiltak som skal evalueres. Men respondentene fremhever at mandatene må bygges opp av konkrete og spesifiserte problemstillinger, først og fremst knyttet til effekter av tiltak og hvorfor et tiltak får de effekter det får. Bare på den måten kan evalueringene bidra til innsikt som kan ligge til grunn for handling. For å sikre at evalueringene blir mest mulig målrettet, påpeker flere av respondentene at det er behov ikke bare for å formulere presise problemstillinger, men at departementene legger arbeid i å utvikle konkret hvilke kriterier som skal ligge til grunn for vurderingene. De er klart av den oppfatning at jo mer som overlates til evaluator med hensyn til å trekke opp evalueringens rammer, jo mindre sannsynlig blir det at evalueringen vil bli et nyttig hjelpemiddel i målstyrings- prosessen.

Helt avgjørende for hvorvidt en evaluering vil bli brukt, er evalueringens faglige kvaliteter. Det første krav er at evalueringen gir svar på de problemstillinger som er gitt i mandatet. Dette er imidlertid ikke tilstrekkelig. Det bilde som tegnes i evalueringen må også stemme overens med brukerens oppfatninger av situasjonen. Hvis ikke bruker kjenner seg igjen, vil evalueringen sannsynligvis bli lagt på hylla. Det betyr ikke at kun konforme evalueringer vil bli brukt, men at det nok vil stilles svært strenge krav til evalueringens metodiske tilnærming og gjennomføring hvis "avvikende" virkelighetsbeskrivelser skal aksepteres. Kravet om metodisk ryddighet er imidlertid ikke forbeholdt evalueringer som utfordrer virkelighetsoppfatninger, men er et gjennomgående krav til all evaluering. Kravet er at evalueringene skal anvende forsvarlige datainnsamlings- og analyse-metoder. Dette er en ufravikelig forutsetning fordi uten et skikkelig metodisk fundament vil de konklusjoner og de anbefalinger som gis, ikke være til å stole på, og kan følgelig ikke ligge til grunn for departementets vurderinger.

Ved siden av kravet om at evalueringene svarer på de problemstillinger som er reist i mandatet, så er kravet om at evalueringene kan dokumentere konkrete effekter av et tiltak, det viktigste. Jo mer spesifisert, jo nyttigere er evalueringen i så måte. Respondentene fremhever dette fordi en slik dokumentasjon gir et godt grunnlag for handling. Det blir, slik en respondent uttrykte det: "en logisk tankerekke fra beskrivelse til forslag."

### **5.8.3 En systematisk evalueringsstrategi**

I den utstrekning evalueringene inngår i en mer systematisk evalueringsstrategi er disse forankret i den enkelte fagavdeling. NIFU har ikke kommet over noe tilfelle hvor et departement som sådan har utviklet en evalueringsstrategi for hele departementet. Dette innebærer at det finnes et mangfold av evalueringsstrategier i forvaltningsapparatet rettet inn mot svært ulike sider ved den offentlige virksomhet avhengig av den enkelte fagavdelings virksomhetsområde. I alt inngår 70% av

---

utvalgets evalueringer i en slik sammenheng. Dette er et imponerende høyt tall og vitner om en svært bevisst innstilling til evalueringer som et verktøy i avdelingenes arbeid.

Hva de ulike evalueringsstrategiene dekker, reflekterer i høy grad hvilket ansvarsområde avdelingene har. Strategiene spenner over alt fra en systematisk oppfølging av virkemiddelbruken innen enkelte avdelinger, via evaluering av et saksfelt eller oppfølging av særskilte tiltak og reformer til en oppfølging av større enkeltsaker slik som Gardemoutbyggingen. Her finnes det ikke noe entydig mønster. Det virker imidlertid som at når et saksområde kommer i Stortingets søkelys, så fører dette til en større bevissthet rundt behovet for å utvikle en systematisk evalueringsstrategi. Utenriksdepartementet kan stå som et eksempel i så måte hvor bistands- og nødhjelpsvirksomheten blir gjort til gjenstand for en systematisk oppfølging ved hjelp av blant annet evalueringer.

De evalueringsstrategier som finnes, er naturlig nok ikke knyttet opp til økonomistyring slik nytt økonomireglement for staten legger opp til. Det er således et åpenbart behov for å utvikle en sådan strategi. Denne strategien bør gå på tvers både av fagavdelinger i det enkelte departement og på tvers av departementer. Formålet med en slik strategi er tosidig. For det første vil det bidra til å utvikle en felles forståelse for muligheter og begrensninger ved evalueringer som et verktøy i et økonomistyringssystem. For det andre vil arbeidet med å utvikle en evalueringsstrategi føre til at kunnskap om og bruk av evalueringer blir forankret i departementsstrukturen på en helt annen måte enn dagens praksis, som i beste fall er forankret i den enkelte avdeling – eller mer sannsynlig i enkeltpersoner.

---

---

## 6 Tema frå gruppediskusjonane

Den første delen av gruppediskusjonane tok for seg kva som bør og ikkje bør evaluerast, og når ein bør og når ein ikkje bør gjennomføra ei evaluering. Eit viktig spørsmål å avklara i denne samanhengen er kva som bør vera *grunnlaget for å gjennomføra ei evaluering*. Deltakarane var i stor grad einige om at utgangspunktet bør vera å få *dekka eit reelt behov for kunnskap*.

Ein god dialog mellom oppdragsgivar og evaluator og ei felles forståing av situasjon og oppgåve, er avgjerande for at ein skal få nytte av ein evalueringsprosess. Det er difor spesielt viktig med ein omfattande dialog under innleiinga av eit evalueringsarbeid. Mange frå evalueringsmiljøa oppfatta bruk av anbod som hemmande i denne samanhengen. Førespurnadene er då strengt definerte med høg grad av kravspesifikasjon. Premissane for evalueringane blir utelukkande gitt frå sektoren si side utan nødvendig kunnskap om metode.

Økonomireglementet for staten stiller krav om formalisert bruk av evalueringar som styringsverktøy. Enkelte understreka at dette ikkje nødvendigvis ville seia forskingsbaserte evalueringar. Rapportering om bruk av midlar er nyttig for departementet, Stortinget og Riksrevisjonen. Slike vurderingar kan gjerast på mange nivå, og mykje kan målast relativt enkelt. Premissane for evalueringar bør leggjast lokalt slik at tilnærminga blir tilpassa dei ulike verksemdene. Institusjonstilpassinga i økonomiregelverket viktig i denne samanhengen.

Tidspunktet for gjennomføring av evalueringar kan vera avgjerande. Verknadene av enkelte tiltak kan målast først relativt lang tid etter at tiltaket blir sett i verk. Det er viktig at det som skal evaluerast er modent for evaluering. Eit utfall av for tidleg evaluering kan vera at ein legg ned tiltak som då ikkje kan visa til dei ønska effektane, men som viser seg å vera gode over tid. Ein bør ikkje framskunda evalueringar for å kunna visa til resultat.

Noko som vart understreka i fleire av gruppene var at prosessevalueringar var like viktige som effektevalueringar. Grunnen til dette er ikkje minst læringselementet som ligg i gjennomføringa av ei evaluering.

Det vart og diskutert i kva grad ein burde bruka interne og eksterne evaluatarar. Ved å bruka interne krefter vil evaluatarane ha betre kunnskap om sektoren/faget, men mindre om evaluering/metode. Der legitimitet er viktig, til dømes innanfor politisk problematiske område eller der kor mange interesser er involverte, er det nødvendig med eksterne evaluatarar. Sjølv med eksterne evaluatarar kan det vera vanskeleg å oppnå legitimitet på grunn av departementet si sterke rolle som oppdragsgivar. I denne samanhengen er det viktig å understreka at evalueringar ikkje må bli kjøp av legitimitet. Evaluatarar må ha lov til å stilla politisk ukorrekte spørsmål.

---

I fleire av gruppene understreka ein behovet for å utarbeida evalueringsstrategiar og for å vidareutvikla evalueringskompetanse i dei ulike departementa. Den generelle evalueringskompetansen er låg, det som fins av kompetanse er ofte knytt til enkeltpersonar eller enkelte grupper.

På bakgrunn av egne erfaringar var det mange som stilte seg spørjande til om bildet av departementa sin evalueringspraksis slik det vart presentert i rapporten frå NIFU, kunne vera rett. Det kunne vera mogeleg at rapporten gav eit bilde av intensjonar og ikkje faktisk atferd.



---

## 7 Vad bör evalueras? Evalueringar i ett resultatstyrningsperspektiv

*Avdelningsdirektör Yngve Boye  
Kommunikationsforskningsberedningen, Sverige*

### Bakgrund

Mål- och resultatstyrning har införts på alla nivåer inom Sveriges statliga sektor och jag skulle tro att utvecklingen är likadan i Norge. Detta inlägg vill sätta in evalueringarna och uppföljningarna i ett mål- och resultatstyrningsperspektiv för att söka besvara frågan vad som bör evalueras. Mina exempel tas från det fält jag känner – finansieringen av sektoriell forskning, utveckling och demonstration, ofta förkortad FUD. I Sverige, likaväl som i Norge, spelar sektoriell FUD en stor roll för samhällets kunskapsförsörjning. Den myndighet, Kommunikationsforskningsberedningen, som jag arbetar på har som uppdrag och övergripande mål kort och gott: *Att stödja sådan FUD som bidrar till att uppnå de mål för trafik-, post- och telepolitiken som statsmakterna beslutat.*

I Sverige startade rapporten ”Kan myndigheter utvärdera sig själva?” en viss debatt. Utan att här gå in på rapportens slutsatser vill jag för min del deklarerera:

- 1) Myndigheter är skyldiga att låta utvärdera olika delar av sin verksamhet och kan också göra det med stor framgång men
- 2) Självfallet räcker det inte med egna evalueringar utan myndighetens uppdragsgivare – regeringen/departementet – har också sitt evalueringsansvar som beställare, för att mål- och resultatstyrningen skall fungera tillfredsställande.

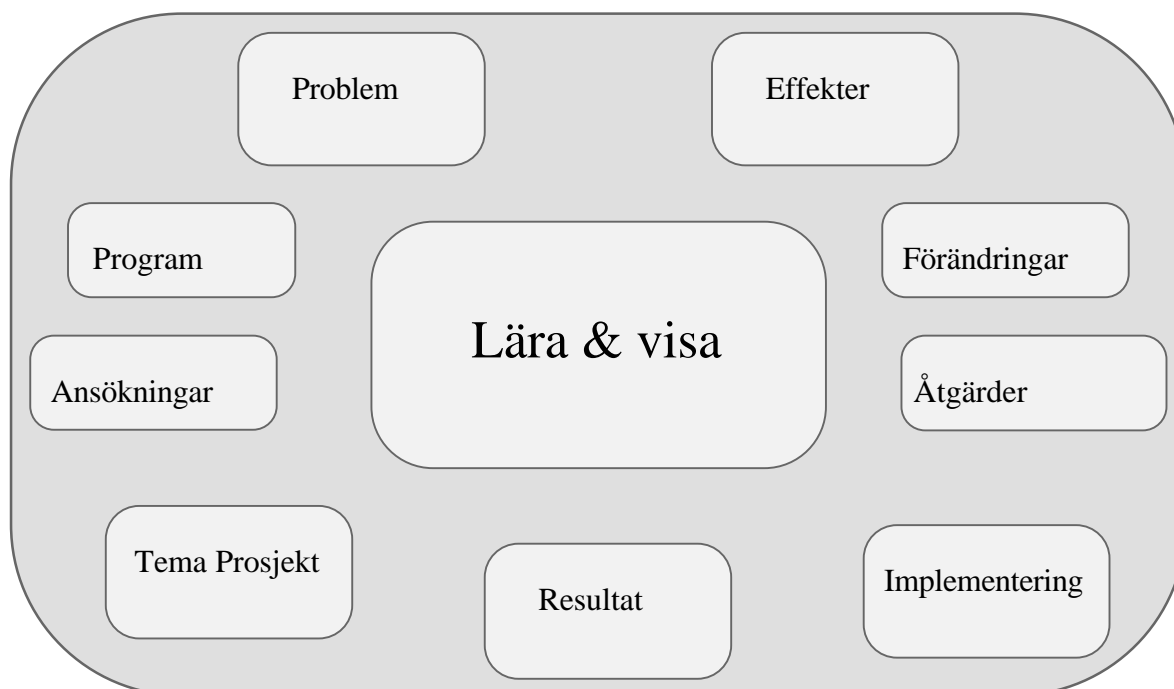
### Tre förutsättningar för en effektiv evalueringsverksamhet

- 1) Dialog med uppdragsgivaren om vad som är målen med olika delar av myndigheternas verksamhet
- 2) Uppföljning kopplad till de mål som successivt skall ges allt tydligare och allt mer operationella formuleringar på grundval av vad som framkommer i dialogen och av egna erfarenheter
- 3) Regelbunden redovisning av resultaten relaterade till de uppställda målen till uppdragsgivaren/departementet i en årsredovisning.

Med stöd av 1– 3 kan en helhet beskrivas som behövs som utgångspunkt för valet av evalueringsobjekt och evalueringsmetod.

---

## Två grundläggande syften med evalueringar: att lära och att visa



Följande bild syftar till att ge en överblick av olika evalueringsbehov som kan gälla för en myndighets verksamhetsgren:

När vi funderar med utgångspunkt från figuren är det viktigt att hålla i minnet vilken roll myndigheten skall spela vid genomförandet av verksamhetsgrenen ifråga. Handlar det om att

- utveckla statsmakternas egna eller vissa andra utpekade aktörers kunskaper eller erfarenheter?
- stödja vissa aktörers verksamhet eller nya delar av den genom finansierings- eller informationsstöd?
- kontrollera efterlevnaden av lag, förordning, bestämmelse?
- eller vad skall verksamhetsgrenen åstadkomma?

Figuren ovan utgår från en myndighets roll som beställare av forskning, utveckling och demonstration. Olika *forskningsprogram* utgör myndighetens *verksamhetsgrenar*.

Varje led i kedjan från problem till effekt kan ha ett behov av evaluering. Men ofta är det mest ändamålsenligt att koncentrera sig på en viss del i processen, kanske den som upplevs som svagast/minst effektiv eller den som man vet minst om.

Min myndighet har ålagts att varje år i samband med sin årsredovisning lämna en plan över de utvärderingar som skall genomföras under året efter det år årsredovisningen avser. Minst 20 % av myndighetens verksamhetsområde skall evalueras. Hittills har våra utvärderingar huvudsakligen varit inriktad på kvalitet och relevans av forskningen vi stöder. Mer och mer vänds dock intresset mot att

---

utvärdera effekterna av den forsknings-, utvecklings- och demonstrations- verksamhet vi beställer och finansierar. Även effektiviteten i meningen cost-effectiveness kommer mer och mer i fokus. De dimensioner vi anser särskilt intressanta framgår av bilaga 1.

## **Sensmoral**

”För att göra de rätta sakerna och göra sakerna rätt” behövs utvärderingar i ett växelspel med kontinuerlig uppföljning av verksamheten ifråga. Alla resurser är begränsade, även evalueringsresurserna. För goda prioriteringar av dem krävs en helhetsbild om den verksamhet det gäller. En sådan kan man skaffa sig genom att ställa och besvara kritiska frågor. Exempel på sådana frågor finns i bilaga 2. Dialogen med uppdragsgivaren måste också hållas levande om problemen, målen och resultaten t ex i en årsredovisning. För exempel se bilaga 3.

Som gammal statstjänsteman får man inte sällan en obehaglig känsla av déjà-vu när man tar del av evalueringars bedömningar och resultat. Jag tror att det beror på att kontinuiteten brister. Man börjar från början i allt för många fall, d v s att ackumuleringen av erfarenheterna och helhetsgreppet är för svaga. Orsakerna till detta är tror jag till stor del att finna i en bristande uppföljning och alltför knapp dialog med uppdragsgivaren.

---

## Bilaga 1

### *Exempel på olika evalueringsperspektiv att anlägga på en sektors kunskapsförsörjning*

<i>Dimension</i>	<i>Nyckelfrågor</i>
<b>Vetenskaplig relevans</b>	Vetenskapligt intresse? Uppmärksamhet på problemen från det internationella vetenskapliga samhällets sida?
<b>Vetenskaplig kvalitet</b>	Tillämpas vedertagen vetenskaplig metodik? Potential för att bidra till verkliga kunskaps-tillskott eller ny mera generell förståelse av viktiga fenomen och samband?
<b>På verkan i den teoretiska eller praktiska verkligheten</b>	Nya insikter, nya instrument och metoder av värde för forskare utvecklare och praktiker? Tillämpbarhet?
<b>Genomförande</b>	Uppläggnings sund? Effektivt genomförande? Ändamålsenlig organisation och styrning. God avvägning av resurser. Måluppfyllelse?
<b>Svenska behov</b>	Behovet och påverkan på den svenska arenan?
<b>Användarnas behov</b>	Finns identifierade användare vars behov tillgodoses? Vem exploaterar resultaten?
<b>Påverkan i undervisning</b>	Öppnas möjligheter för nya uppgifter för forskarstuderanden eller påverkas kursplaner/litteratur på grundläggande akademisk nivå? Revision av läroböcker?
<b>Sammanfattande bedömning</b>	Resultaten värd kostnaderna?

---

## Bilaga 2

### *Exempel på uppföljningsrutin*

#### **1. Verksamhetsgrenens existensberättigande**

- 1.1 Vilka problem motiverar verksamhetsgrenens existens?
- 1.2 Vilka är intressenterna som har nytta av att problemen kommer närmare sin lösning?
- 1.3 Vilken prioritering bör göras av de problem som de av myndigheten finansierade insatserna i främsta rummet bör inriktas på? (Analys av de problemområden som verksamhetsgrenen innehåller med avseende på framgångsmöjligheter och angelägenhet.)
- 1.4 Inriktats insatserna verkligen mot de prioriterade problemen och vilka större avvikelser kan konstateras därvidlag?

#### **2. Målformulering**

- 2.1 Hur ser relationen problem – mål för insatserna ut?
- 2.2 Hur bör huvudmål och delmål för programmet formuleras?
- 2.3 Är målen logiska givet problemen och intressenternas behov? (Check av logiken i programstruktur och mål)

#### **3. Rutiner och kriterier för bedömning av forskningsansökningar och initiering av nya FUD-ansatser**

- 3.1 Hur bör de medel som anslagits till verksamhetsgrenen fördelas på olika åtgärdsområden?
- 3.2 Vilka brister ser vi i dag vad gäller rutiner och kriterier vid bedömning av åtgärderna?
- 3.3 På vilka delområden måste vi initiera insatser för att uppfylla programmets mål?
- 3.4 Kan vi utnyttja vårt omvärldsstöd på ett effektivare sätt?

#### **4. Uppföljning**

- 4.1 Vilka indikatorer skall vi använda för att få stöd i det dagliga programarbetet?
- 4.2 Hur användbara är dessa indikatorer vid bedömningar av ansökningar och egna initiativ?
- 4.3 Hur mycket och till vad användes slutrapporter och lägesrapporter ?
- 4.4 Vad bör göras för förändringar i rapporteringsrutiner för att göra dem mera ändamålsenliga från resultatstyrningssynpunkt?

---

## **5. Utvärdering**

- 5.1 Vilka utvärderingar är gjorda och har vi verkligen utnyttjat utvärderingarnas resultat?
- 5.2 Vilka är de mest brännande frågorna att få svar på via utvärderingar?
- 5.3 Vilken är den mest angelägna utvärderingen att genomföra under det närmaste året?

## Bilaga 3

### Översiktlig jämförelse mellan verksamhetsmål och resultat

VERKSAMHETSMÅL I REGLERINGSBREV	RESULTAT I KONCENTRAT
Program skall utarbetas för FUD-verksamheten	De program som uppfyller inriktningsplanens intentioner är sedan våren 1996 antagna av KFBs styrelse, se avsnitt 2.1
Den upplevda nyttan av den av KFB finansierade forskningen skall enligt mätningar uppvisa positiv utveckling	Två på varandra följande undersökningar av bl a den kundupplevda nyttan visar en mindre förbättring bland dem som har direktkontakt med KFB se avsnitt 2.3.1
Samma grad som hittills av medfinansiering från industri, transportföretag m fl skall eftersträvas	Andelen medfinansiering visar i stort en oförändrad utveckling under 1995/96 jämfört med tidigare år se avsnitt 2.2.4.
FUD-resultaten skall dokumenteras och förmedlas effektivt med anpassning till användarnas behov	Ökad publicering av korta sammanfattningar av pågående FUD-projekt, mer av kunskapssammanställningar, se avsnitt 2.3.1
Forskningsmiljöer av internationell klass skall byggas upp med minst 40 % långsiktig kunskapsuppbyggnad och ökning av antalet licentiat- och doktorsarbeten	Medelstorleken i kronor per år av KFBs insatser har stigit. I stort oförändrad andel av FUD-medlen inom huvudanslaget används till teman och projekt med minst treårig finansiering. Antalet licentiat- och doktorsavhandlingar har ökat, se avsnitt 2.2.3.
Samverkan mellan KFB-stödda forskare skall öka med stärkande särskilt av samverkan mellan samhällsvetenskaplig och teknisk forskning	Flera nya nätverk har utvecklats som innefattar forskare från flera olika discipliner. För exempel se avsnitt 3.1 om forskarnätverket "Vägar till uthållig utveckling" och avsnitt 3.2 om KFBs initiering av forskning inom olika gränsfält på Telematikområdet
Ett närmare samarbete mellan VTI och högskoleinstitutioner bör underlättas	KFB-stöd till sådant samarbete under 1995/96 särskilt på infrastrukturteknikområdet, se avsn. 3.3.4
Förbättrad internationell anknytning genom bl a ökat internationellt forskarutbyte, ökad spridning utomlands samt bättre information om utländsk forskning i Sverige	Särskilt styrelsebeslut om stöd till forskarutbyte vilket bland annat resulterat i en gästprofessur inrättad vid Centrum för transport och trafik i Göteborg, se avsnitt 3.3.2
Ökat antal artiklar i internationella vetenskapliga tidskrifter	Antalet vetenskapliga artiklar visar en ökande trend under perioden 1991/92 – 1995/96. Antalet artiklar under 18-månadersperioden 1995/96 var dock detsamma som för 12-månadersperioden 1994/95, se avsnitt 2.2.3
KFB skall svara för att Sverige deltar i europeiskt forskningsarbete till största möjliga nytta för landet	Svenska deltagare finns i 33 % av projekten inom EUs transportprogram. Av EUs programbudget går 3,8 % till svenska forskare vilket motsvarar 54 miljoner kronor, se avsnitt 2.4
Som ett genomsnitt skall 20 % av forskningsstödet utvärderas	Ca 20 % av forskningsvolymen utvärderad 1995 och 1996, se avsnitt 2.5

---

## Genomförande

Det här skisserade utvecklingsstödet kan ske med mycket varierande ambitionsnivå. Den lägsta ambitionsnivån men kanske också den mest realistiska är att utse en samtalspartner till den som närmast har ansvaret till programmet och att ange en tidpunkt då svar på de ovan ställda frågorna skall vara dokumenterat. Samtalspartnern bör helst väljas inom kansliet. Promemorian med svaren bör först diskuteras på enheten och sedan med Gd för att slutligen ge ett underlag till en programdiskussion i vederbörande ledningsgrupp eller motsvarande forum för omvärldsstöd på det sakområde programmet gäller. Inom telematikområdet kan det vara klokt att avvakta resultatet av FBA:s formativa utvärdering och bygga vidare på den.



---

## 8 Erfaring med bestilling, gjennomføring og mottak/bruk av evalueringar.

*Civilekonom Sven Faugert  
SIPU Utvärdering AB*

### Kort resumé av föredraget

Föredraget utgår från 25 års egna erfarenheter som beställare och utförare av utvärdering i Sverige. De senaste 8 åren har jag arbetat med detta som konsult, dessförinnan som statstjänsteman i myndigheter och departement.

Inledningsvis tar jag ett grodperspektiv och frågar: Vad kännetecknar en dålig och en bra beställare av utvärdering? Vad kännetecknar en dålig och en bra utförare?

Med den bakgrunden går jag sedan steg för steg igenom mina erfarenheter av vad som kan gå gale i en utvärdering och vad man kan och bör göra för att undvika det.

Utvärdering kan och bör göras i olika skeden av ett program. För varje skede finns olika utvärderingsfrågor.

Vanliga misstag och lämpliga åtgärder exemplifieras när det gäller

- val av tidpunkt för utvärdering
- klargörande av motiv för och användning av utvärdering
- precisering av utvärderingsfrågor.

Olika typer av utvärderingsfrågor kräver olika metod-ansatser.

Det finns för- och nackdelar med interna respektive externa utvärderare. Utvärderingens intressenter måste hanteras på lämpligt sätt. I särklass står uppdragsgivaren. Det är ofta lämpligt att använda flera former av interaktion med uppdragsgivaren. Utvärderaren kan - frivilligt eller ofrivilligt - få olika roller, som det också gäller att hantera.

Nu talas det en del om effektutvärderingar. Exempel ges på en ansats som vi med framgång använt i Sverige, effektkedjeanalys. Grundmodellen, effektkedjan är enkel. Med en genomtänkt undersökningsstruktur och effektiva analysmetoder kan vissa kvantitativa effektbestämningar göras. I exemplet gällde det en insats för att främja spridning av energieffektiv teknik bland hushåll i Sverige.

Utvärderarens ständiga önskan är genomtänkta skriftliga mandat från uppdragsgivaren.

---

För tilltron till utvärderingar är det viktigt att kvalitetssäkra hela utvärderingsprocessen. Särskilda kvalitetsindikatorer kan där vara till god hjälp. Till sist är det så att nyttan av utvärderingar bestäms av lärklimatet. I en organisation kan det råda olika förhållningssätt. Åtgärder kan behöva vidtas inom en organisation för att röja undan hinder för lärande.

Gemensamma åtgärder för departementen kan också vara lämpliga.

---

## 9 Erfaringer med evaluering i tre departementer

### 9.1 Utenriksdepartementet.

*Revisjonsdirektør Torbjørn Aalbu*

#### 9.1.1

Erfaringene fra Utenriksdepartementet når det gjelder evaluering er konsentrert om bistandssektoren, altså forvaltningen på departementsnivå av norsk utviklingshjelp. Dette saksfeltet har ligget i UD siden sammenslåingen med det tidligere Departementet for utviklingshjelp i 1990.

Evalueringsevirsomheten innen bistandssektoren er imidlertid langt eldre. Den startet for alvor i 1978 med opprettelsen av Kontoret for evaluering og forskning i Direktoratet for utviklingshjelp. Før dette hadde de enkelte avdelinger i Direktoratet gjennomført egne prosjektgjennomganger. På dette tidspunkt hadde ingen andre norske statsinstitusjoner etablert en tilsvarende virksomhet.

Dualismen mellom evalueringer og prosjektgjennomganger har fortsatt fram til i dag. I realiteten har det eksistert to evalueringssystemer i norsk bistand. Et desentralisert system for selvevaluering, med særlig vekt på regelmessige prosjektgjennomganger og fokus på resultater i forhold til framdriftsplaner og budsjetter, og et sentralt system med fokus på kriterier som relevans, virkning, måloppnåelse, produktivitet og bærekraft. Som det framgår har systemene hatt forskjellige mål/normer og delvis vært konkurrerende, også som følge av at arbeidsdelingen er blitt del av et institusjonelt skille mellom departement og underliggende etat.

Evalueringsevirsomheten innen bistand har også vært preget av at utviklingsbistand er et internasjonalt fagfelt hvor store aktører som f.eks. Verdensbanken har ledet an. I OECDs Komité for utviklingshjelp (DAC) ble det tidlig på 1980-tallet etablert en ekspertgruppe for evaluering hvor Norge hele tiden har vært en aktiv deltaker. Vår egen evalueringsevirsomhet i de senere år har vært basert på en egen håndbok for evaluering og resultatvurdering i bistanden, som ble utgitt i 1993.<sup>3</sup>

Det har hos oss alltid vært en stor grad av politisk enighet om betydningen av evaluering av utviklingsbistand. I Stortingsmelding nr. 19 heter det: "Det vil bli vurdert hvordan evalueringsevirsomheten best kan bidra til et helhetlig system for resultatvurdering, analyse og kvalitetssikring". Stortinget kommenterer dette bl.a. på følgende måte: "*Flertallet mener at evalueringsevirsomheten av bistanden må styrkes og at det må legges vekt på å styrke en kritisk og uavhengig evaluering.*"

---

<sup>3</sup> *Evaluering og resultatvurdering i bistanden. Håndbok for utøvere og beslutningstakere.* Utenriksdepartementet, august 1993. Norsk og engelsk versjon.

---

*Evalueringsarbeidet må sikres integritet og bredde. Dette kan oppnås blant annet ved at det i tillegg til intern prosjektgjennomgang satses mer på eksterne fritt-stående evalueringer".*

Dette innspillet fra Stortinget sammen med bestemmelsene om evaluering i det nye økonomireglementet har vært grunnlaget for et utredningsarbeid som UD satte i gang i november i fjor og som vi nå holder på å avslutte. Målet har vært å få til et helhetlig og sammenhengende evalueringssystem hvor prosjektgjennomganger, årlig rapportering og eksterne evalueringer ses i sammenheng. Dette skal utvides til også å omfatte UD's øvrige virksomhet, altså ikke bare bistandsområdet. Jeg skal komme litt tilbake til dette mot slutten av dette innlegget.

### **9.1.2**

Hva er så forutsetningene for en effektiv evalueringsvirksomhet? Erfaringene fra bistandsevaluering synes særlig å peke på følgende elementer:

- politisk og administrativ prioritering av evaluering
- tilstrekkelige faglige og økonomiske ressurser for evalueringsformål
- et evalueringsprogram som reflekterer de reelle prioriteringer
- oppfølgingsprosedyrer etter evalueringene.

Effekten av evalueringsvirksomheten i bistandssystemet har helt klart vært sterkst i de perioder hvor funksjonen er blitt prioritert både av politisk og administrativ ledelse, når ledelsen har sendt signaler om at evaluering er viktig og også påsett at de resultater og konklusjoner som framkommer gjennom evaluering anvendes i praktisk arbeid. Samtidig er det viktig at evalueringsfunksjonen er plassert slik i organisasjonen at det eksisterer nødvendig avstand til de operasjonelle avdelinger for å unngå forhold som kan sammenlignes med bukkens kontroll over havresekken. Disse forhold har i UD variert noe over tid.

En annen forutsetning er at økonomiske ressurser og faglig kompetanse er disponible for evaluering. Det er nødvendig både hos dem som skal fatte beslutninger om at evalueringer skal foretas, hos dem som administrerer og gjennomfører evalueringer og hos dem som skal bruke resultatene. Bredt faglig sammensatte evalueringsenheter, omfattende erfaring og evne til samarbeid er i denne sammenheng viktige stikkord. Løpende evalueringsfaglig skoling av ansatte på saksbehandler- og ledernivå er nødvendig. Utviklingen de senere år har vist en relativ nedgang av ressursbruk på evaluering. Dette er nå i ferd med å snu.

Ifølge Stortinget (Innst S nr 229 1995–96) skal evaluering av bistand være kritisk og uavhengig. Begrepet uavhengighet kan forstås på mange måter, og i en politisk styrt organisasjon som et departement vil det være grenser for uavhengighet. Det vil også innen organisasjonen være en balanse å ivareta mellom uavhengighet, deltakelse og læring. Uavhengighetskravet synes derfor i vesentlig grad å være forbundet med at gjennomførende konsulent har nødvendig faglig integritet. Rapporters troverdighet er avhengig av ekspertisen og uavhengigheten til oppdragstaker og hvor åpen evalueringsprosessen er.

---

Troverdighet forutsetter at både suksess og fiasko rapporteres. Det er også avgjørende for evalueringsvirksomhetens renommé at konsulenter holder tidsfrister og at evalueringer kommer tidsnok i forhold til planleggings- og beslutningsprosesser.

Når det gjelder sammensetning av evalueringsprogrammet har det i stor grad basert seg på forslag fra underliggende etat og avdelinger. Politisk og administrativ ledelse har i liten grad gitt føringer mht hva som har vært viktig. Dette er imidlertid i ferd med å endre seg. Det har også vært en generell erfaring at relevansen til evalueringene har vært begrenset fordi de viktigste spørsmålene ikke har blitt evaluert. Mindre viktige evalueringer er blitt foreslått på bekostning av de viktige. Stortingets økende interesse for bistandsevaluering fører nå til at de sentrale spørsmålene blir satt høyere på dagsorden.

En vesentlig svakhet ved bistandsevalueringene har vært at det ikke har blitt kontrollert om anbefalinger vedtatt på basis av evalueringer er blitt fulgt opp i praksis og integrert i politikk, programmer og prosjekter. Mange mener derfor at evalueringer av denne grunn har hatt liten virkning på norsk bistand. Slike kontrollsystemer vil nå bli vurdert. Det vil også kunne bli et krav at sentrale planleggingsdokumenter kan vise til at en har brukt relevant evalueringslitteratur.

### 9.1.3

Evalueringsarbeidet i bistanden har hatt to hovedformål:

- å systematisere erfaringer for å forbedre pågående og framtidig virksomhet
- å gi uavhengig informasjon om bruken av midlene til bevilgende myndigheter.

I motsetning til Verdensbankens femti mann store "Operations Evaluation Department" har hovedvekten for norske bistandsevalueringer vært lagt på å systematisere erfaringer for å forbedre pågående og framtidig virksomhet. Rapporterings- og kontrollfunksjonen har vært mindre viktig. I 1992–93 gjennomførte UD en evaluering av intern læring som følge av evalueringer.<sup>4</sup> Studien presenterte fram en rekke interessante funn:

Når det gjelder å ta evalueringsresultater i bruk er det ikke kanaliseringen/distribusjonen av informasjonen som er avgjørende, men innhold og "timing" i forhold til styrings- og planleggingssykluser. Det ble påpekt at det tok for lang tid å produsere evalueringsrapporter; når de forelå hadde ofte "toget gått" mht å nyttiggjøre seg dem.

Læringen er større fra prosjektgjennomganger enn fra evalueringer pga nærheten til prosessen. Læringen kan bli svært høy når utforskningen og løsningen på et problem foregår nær opp til den enkeltes arbeidsoppgaver.

---

<sup>4</sup> Utenriksdepartementet: *Evaluation Report 1-93: Internal Learning from Evaluations and Reviews*, by Scanteam international as, January 1993.

---

Absorpsjonskapasiteten, og følgelig graden av læring, er svakere når det gjelder abstrakt kunnskap i form av informasjon på makronivå enn ved tilbake-rapportering av mer konkret informasjon.

Rapporten understreker betydningen dominerende kunnskapsstrukturer har for læring, dvs en felles referanseramme, felles tilslutning til organisasjonenes uttrykte verdier og idéer.

Rapporten skiller mellom læring ved deltakelse og læring ved kommunikasjon. Når det gjelder det første understreker studien betydningen av at personalpolitikken tillater ansatte å arbeide nært sammen med evalueringsteam. Ledelsen må oppfordre ansatte til dette og legge forholdene til rette. Når det gjelder det andre påpeker studien betydningen av "policy"-rettede studier som gir bidrag til læring når det gjelder spørsmål av politisk art og som går på tvers av organisatoriske og faglige grenser. Betydningen av regelmessige seminarer, arbeidsgrupper, offentlige debatter, systematisk lagring av rapporter osv. framheves som mekanismer for å bedre læring ved kommunikasjon.

#### **9.1.4**

Jeg har nevnt at vi nå er i ferd med å avslutte et utredningsarbeid i UD hvor hovedformålet er å utvikle en samordnet evalueringsfunksjon for hele departementets virksomhetsområde, med utgangspunkt i Økonomireglementets krav særlig mht koblingen mellom evaluering og økonomistyring. Ledd i dette vil være et årlig evalueringsprogram for hele departementet, samordning mellom det som gjøres i avdelingene og i NORAD på den ene side og den overordnede evalueringsfunksjon for hele departementet på den annen. Det blir lagt opp til årlig rapportering om resultater av evalueringsvirksomheten i form av status-rapporter og sammendrag for bruk i budsjettsammenheng, og det legges opp til systematisk bruk av evalueringsresultater for intern læring i samarbeid med Utenriktjenestens kompetansesenter og intern og ekstern informasjon i samarbeid med Presse-, kultur- og informasjonsavdelingen. Kontroll av oppfølging forutsettes ivaretatt hovedsakelig i sammenheng med budsjettprosessen. Det foreslås også opprettet et internt rådgivende og koordinerende utvalg for å sikre den nødvendige samordning mellom fagavdelingene. Utvalget som foreslår disse tiltakene vil ha sitt avsluttende møte førstkommande mandag, og vi regner med at det ikke vil ta lang tid før departementets ledelse tar stilling til utvalgets forslag. Så snart rapporten foreligger vil den selvsagt også være tilgjengelig for interesserte utenfor vårt eget departement.

---

## 9.2 Kulturdepartementet

### *Underdirektør Rigmor Grande, Forvaltningsseksjonen*

Kulturdepartementet har ikke gjennomført så mange evalueringer, og de som har vært gjennomført har ofte vært nokså avgrenset. Vi har dermed heller ikke brukt mye ressurser på dette til nå.

Situasjonen har vel båret mest preg av at vi har foretatt evalueringer der det har vært problemer – enten i form av kritiske spørsmål fra politisk ledelse, mediaoppmærksomhet eller egen ubekvemhet/uro for at ikke alt har vært som det skal være – f.eks. ved at vi ikke har fått de forutsatte rapporter, eller at vi selv har sett at en ordning har fungert dårlig eller hatt lav/synkende legitimitet.

Det gir neppe uttrykk for en evalueringsstrategi, og heller ikke for et bevisst forhold til at visse typer tiltak bør evalueres planmessig ut fra sin karakter, betydning e.l. Ikke nødvendigvis fordi vi ikke har *forstått* at vi burde etterprøve resultater på denne måten – men vi har ikke sett det mulig å prioritere det.

Bestillingskompetanse: Vi er neppe bedre enn andre, heller tvert imot – vi strever veldig med kvalitetsbegrepet, og med at vi oftest ønsker å måle effekt, antakelig ut fra at vi opplever at offentlig kulturpolitikk har legitimitetsproblemer i opinionen. Det anses ikke vulgært i Norge å si at man ikke burde ha noen kulturpolitikk.

Effekt har for oss – som andre – sammenheng med kvalitetsbegrep – vårt mål er å skape et *kvalitativt* bedre samfunn for folk å leve i, ved hjelp av *kvalitativt* gode kulturtilbud. Hva skal man da måle mot eller sammenligne med – det kulturløse samfunn? Tilstanden *før* den offentlige kulturpolitikken, på 1940-tallet? Da vil vi fort få problemer med å definere hvilke utviklingstrekk i etterkrigstida som skyldes kulturpolitikken, hvilke skyldes velferds- og samfunnsutviklingen, utviklingen innen media, teknologi.

Undertegnede har jobbet i offentlig kulturforvaltning siden 1978, på kommunalt, fylkeskommunalt og statlig nivå. Har grublet mye, deltatt i mange diskusjoner om HVORDAN MÅLE KVALITET – noen ganger minner disse om typen diskusjoner som HVOR GÅR VI, HVOR STÅR VI, HVEM SKAL BETALE? Vi kommer ingen vei. Hvorfor går vi slik i ring?

Jeg har erkjent at det er et blindspor å lete etter tall som kan beskrive kreativitet, intellektuell utvikling, følelsesmessig modenhet, erkjennelse og forståelse av det dypt menneskelige – og at vi må nøye oss med å finne en troverdig måte å beskrive dette på med *ord*. Vi må stole på fagfolk og profesjonelle "evaluatorer" innen vårt felt – sier Hans Rossiné at det er godt teater, så får vi tro at det er det. Videre får vi trøste oss med at van Gogh ikke solgte ett eneste bilde så lenge han levde, mens hans bilder har hatt STOR betydning i kunsthistorien og i historien om menneskets utvikling. I dag kan vi gjenkjenne smerten og galskapen, fordi han våget å male den.

---

Uten sammenligning for øvrig – over til økonomireglementet og dets krav til evalueringer.

Noe som preger økonomireglementet, og som MÅ prege et slikt verktøy, er bildet av at den byråkratiske (og dermed politiske) verden kan fremstilles skjematisk. (Men det kan den jo ikke.)

Jeg vil trekke fram noen tankekors i den sammenheng – men først har jeg lyst til å understreke:

Selv om jeg nok tror at en annen forsamling enn oss, kanskje inkludert NOEN av oss, vil sitte her, eller et annet sted, om kanskje ti år, og le like hjertelig av vår tro på at NÅ hadde vi – EUREKA! – funnet løsningen på hvordan vi skulle få oversikt og kontroll – som vi lo i går av "DEN NYE STATEN" – så ser jeg like fullt det nye økonomireglementet som et kjærkomment dytt i baken for å få ryddet opp i mye gammelt grums, mange lite gjennomtenkte ordninger, tilfeldigheter, historisk betingede merksnodigheter, utdaterte virkemidler, manglende eller dårlige rutiner osv. osv. Jeg er nemlig et ordensmenneske og synes det er DEILIG med orden og system – og så har jeg en naiv og usvikelig tro på at demokrati og rettferdighet er de beste goder noe folk kan bli begunstiget med. Jeg synes det er viktig at "det offentlige" bruker skattekrone til de beste formål og på mest mulig effektiv måte – slik at vi i størst mulig grad kan oppnå DET GODE LIV, som vi snakket om i går.

Så jeg liker økonomireglementet. Jeg tror vi kan få LITT mer strømlinjeforma virksomheter, litt bedre tak på interne rutiner i forbindelse med tilskuddsforvaltning, litt bedre grunnlag for politisk styring osv. osv. Kanskje nokså mye bedre også.

Hvis man lurer på om man har beveget seg framover, må man se seg tilbake og måle avstanden til utgangspunktet. Og selv om vi ved hjelp av DEN NYE STATEN og andre lignende dokumenter ikke kom så langt som vi hadde ambisjoner om, har vi faktisk kommet videre. Det gjør vi nok denne gangen også.

Men ved siden av å være opptatt av orden og system, og ikke minst gjennom-siktighet (det skal være *synlig* hva vi driver med) – så er jeg altså også veldig opptatt av kulturpolitikken og dens kår. Og da kommer tankekorsene.

Det blir nå krevd at vi legger langt større vekt på evalueringer enn før. Særlig skal vi evaluere tilskuddsordninger. I hvert fall er jeg særlig opptatt av dette temaet, siden mitt lille departement har en SVÆRT stor tilskuddsportefølje.

Vi opplever – og det er vi neppe alene om – at Stortinget gjør vedtak om tilskudd, uten at premissene for tilskuddet går fram. Over år kan slike vedtak om tilskudd utvikle seg til å bli en "tilskuddsordning" fordi flere aktører greier å "komme inn på statsbudsjettet" når døren først er satt på klem. Vi har et slikt eksempel i tilskudd til festivaler.



---

Når vi nå, på grunn av det nye økonomireglementet, arbeider med et regelverk for denne "ordningen" – møter vi på problemer. Det fins f eks ingen indre logikk i sammensetningen av tilskuddsmottakere. Noen er store, kjente festivaler med nasjonalt og internasjonalt tilsnitt og ry, andre er små og av en atskillig mer lokal karakter. Noen er bare for profesjonelle musikere, mens andre er for amatører. Sjangrene variere selvfølgelig sterkt. Vi støtter i tillegg også andre typer festivaler enn musikkfestivaler. Noen får noen millioner kroner, andre får et par hundre tusen. Beløpsstørrelsene er utviklet over år, ved konkrete vedtak i Stortinget.

Man skal ikke vite så mye om norsk politisk hverdag før man finner bakgrunnen for dette bildet som tegner seg, dvs for en del av de elementene vi ikke finner noen logisk forklaring på, i hvert fall ut fra et kulturpolitisk ståsted:

- distriktpolitikken som overstyrende for mange/de fleste andre hensyn i norsk politikk
- politiske kompromisser (noe mindre pent sagt hestehandel) som nødvendig forutsetning for et saldert budsjett.

Dette er helt legitime hensyn man må ta i politikk generelt og i norsk politikk spesielt. De er bare ikke alltid så lett å sette inn i en sektor-faglig ramme.

Så både Finansdepartementet og Riksrevisjonen må antakelig ha oss unnskyldt når vi har visse problemer med den skjematiske fremstillingen av virkeligheten.

Vår utfordring framover blir:

- i størst mulig grad å innpasse vår del av virkeligheten i "skjemaveldet", slik at vår virksomhet kan beskrives og vurderes på linje med andres
- samtidig å synliggjøre de delene som ikke så lett lar seg innpasse –

tilskudd som kanskje ikke *kan* målstyres og følges opp så mye – tenk på typen pressestøtte –

tilskudd som kanskje ikke *bør* underlegges statlig målstyring, som typen frivillige organisasjoner eller private kulturinstitusjoner der statens tilskudd bare utgjør en liten del av deres totale omsetning.

For alle disse (de er nemlig mange) må vi konsentrere oss om å kontrollere at midlene blir brukt etter forutsetningene, at våre egne rutiner er gode nok og at det er synlig for Stortinget og andre at det *blir* brukt penger på disse formålene – og på hvilke premisser det blir gjort.

I årets budsjettproposisjon har vi foreslått en tredeling av tilskuddene når det gjelder *styringsambisjonsnivå* – og dette vil selvsagt også være styrende for våre valg når det gjelder evaluering. Vi har eksplisitt bedt Stortinget melde tilbake dersom de ikke finner våre styringsambisjoner høye nok, slik Økonomi-

---

reglementet legger opp til. Komitéen har kvittert ved å si at den «er enig med departementet i at det verken er ønskelig eller praktisk mulig å fastsette presise resultatkrav for alle ansvarsområder til departementet».

Så blir det jo spennende å se da – hva som er morsomt om ti år?

---

## 9.3 Kirke-, utdannings- og forskningsdepartementet

*Rådgiver Morten Nordlie, Avdeling for videregående opplæring*

### 9.3.1 Innledning

Kirke-, utdannings- og forskningsdepartementet (KUF) har – som NIFUs kartlegging viser – gjennomført et betydelig antall evalueringer i perioden 1994–96. Denne fremstillingen er konsentrert om den løpende evalueringen av Reform 94, som ble initiert i 1993, iverksatt parallelt med reformen i august 1994 og som vil pågå frem til innføringsfasen for Reform 94 er avsluttet i 1998.

Et av motivene for departementet bruk av evalueringer knytter seg til et løpende behov for kunnskap for å sikre at alle våre virksomhetsområder skal preges av god styring, kvalitet i alle ledd og effektiv bruk av ressursene. Evalueringene er ett av flere tiltak for å fremskaffe data som er nødvendige for å danne et tilfredsstillende grunnlag for styring, tilsyn og kontroll i utdanningssystemet. En forutsetning for å utøve effektiv styring og for å kunne vurdere måloppnåelse både i etat- og sektorstyringen, er at det er tilgang på relevant og pålitelig informasjon som er lett tilgjengelig og som kommer til rett tid. Det nye økonomireglementet for staten forsterker dette behovet.

På denne bakgrunn har en arbeidsgruppe i KUF (*Statistikkgruppen* under ledelse av departementsråden) nylig utredet departementets behov, gjennomført en rekke analyser knyttet til behov, systemer, ansvar m m og presentert strategier og tids- og aktivitetsplaner for de forskjellige delene av dokumentasjonsarbeidet. Målet er å utvikle et samlet og sammenhengende system for rapportering om tilstanden og løpende vurderinger av situasjonen på KUFs virksomhetsområder.

### 9.3.2 Evalueringen av Reform 94

#### *Om reformen*

Reform 94 er et svar på en rekke utfordringer videregående opplæring stod overfor ved inngangen til 1990-tallet. Utfordringene gjaldt i særlig grad fag- og yrkesopplæringen.

Denne delen av opplæringen hadde: *en mangfoldig, uoversiktlig og overlappende tilbudsstruktur, mange opplæringsløp førte ikke frem til kompetanse, gjennomstrømningen var dårlig fordi mange elever beveget seg på tvers i systemet, spesialiseringen startet i stor grad på grunnkursnivå og det var svak kobling mellom skole og lærebedrift.*

For videregående opplæring totalt sett var utfordringene: *å utvikle et overordnet måldokument for all videregående opplæring, utforme målstyrende læreplaner, øke mulighetene for flere til å skaffe seg høyere utdanning, lage et system som kunne følge opp elever som falt fra og skape læringsmiljø som i større grad*

---

*involverer og ansvarliggjør elever og lærlinger når det gjelder opplæringen opplæringen.*

Omlegningen ble innledet med Blegenuutvalget, som leverte sin innstilling i 1991 (NOU 1991:4 «Veien videre til studie- og yrkeskompetanse»).

Hovedprinsippene i innstillingen ble fulgt opp av Regjeringen i St meld nr 30 (1992– 93), som Stortinget senere sluttet seg til.

### ***De sentrale målene***

*Reform 94 sikter blant annet mot: å gi all ungdom en rett til treårs videregående opplæring, å gi all ungdom muligheter til å nå frem til en formell kompetanse – enten studiekompetanse, yrkeskompetanse, delkompetanse eller annen form for videregående opplæring, å gi god gjennomstrømning av elever og lærlinger nærmest mulig hjemsted, å få til en fremtidsrettet yrkesopplæring som kombinerer yrkesopplæring i skole med avsluttende spesialisering i bedrift, å sikre voksne mennesker gode og tilpassede tilbud om videregående opplæring, å tilrettelegge innholdet i den videregående opplæringen slik at den gir kompetanse for yrke og samfunnsnivå i fremtiden.*

Iverksettingen av Reform 94 bygger på et grundig utredningsarbeid. Gjennom utviklingsprosjekter i regi av de tidligere faglige rådene og ved en rekke lokale utviklingsprosjekter var det også bygget opp en beredskap på skolenivå til å kunne møte det nye i reformen. På denne bakgrunn og fordi en ønsket å gi nye årskull en fullverdig kompetanse så raskt som mulig, ble det ikke gjennomført utprøvinger i mindre skala i forkant.

For å finne i hvilken grad sentrale mål i reformen blir innfridd og for å kunne foreta eventuelle justeringer skulle det tilrettelegges en løpende forskningsbasert evaluering med en tidsramme som dekket innføringsfasen av reformen (1994– 98).

### ***Mål og tema***

På bakgrunn av St meld 33 Kunnskap og kyndighet Om visse sider ved videregående opplæring (1991–92) og andre grunnlagsdokumenter ble det i forkant av selve iverksettingen av Reform 94 av departementet utviklet *et mål- og rammedokument* for evalueringen. Forskere og brukere med kunnskapsinteresser knyttet til reformen deltok i prosessen.

Dokumentet skisserer mål, fokus, organisering, økonomi og krav til anbudsokumentene. På basis av dette dokumentet ble relevante forskningsinstitusjoner invitert (høsten 1993) til å komme med tilbud. Forsommeren 1994 inngikk syv forskningsmiljøer kontrakt med KUF.

Siden kunnskapsbehovene vil kunne skifte i løpet av evalueringsperioden, forhandles kontraktene på nytt hvert år. Dette for å sikre fleksibilitet i forhold til nøkkelspørsmål og kompetanse. Det kunne jo vise seg å være behov for å endre i forkant lagte prioriteringer forholdsvis raskt i forhold til uforutsette kunnskapsbehov under iverksettingen.

Evalueringen tar for seg fire hovedområder: 1) *dimensjonering og kapasitet,*

---

2) gjennomstrømning og kompetanse, 3) organisering og samarbeid og 4) innholdet i og tilretteleggingen av opplæringen.

I tillegg er det egne prosjekter knyttet til: a) oppfølgingstjenesten, b) fremmedspråklig elever og lærlinger, c) elever uten rett og d) elever og lærlinger med behov for særskilt tilrettelagt opplæring.

På grunnlag av innspill fra brukerne og i samarbeid med forskerne har oppdragsgiver (departementet) utarbeidet egne nøkkelspørsmål innenfor hvert av områdene og delprosjektene. Spørsmålene er gjenstand for eventuelle justeringer i tilknytning til de årlige kontraktsforhandlingene. Partene i arbeidslivet, elev- og lærerorganisasjonene, Oppfølgingsgruppen (se nedenfor) og forskerne er med i denne årlig prosessen. Den endelig versjonen fastsettes gjennom kontraktsforhandlingene mellom oppdragsgiver og forskningsinstitusjonene. Den årlige resursrammen vil være avgjørende for hva det er mulig å få belyst.

### **Organisering og legitimitet**

Siden evalueringen blant annet skal *gi kunnskap og innsikt underveis i en form som gjør det mulig raskt å kunne justerer i forhold til målene*, vil den i særlig grad være beslutningsorientert. Fra oppdragsgivers side vil det blant annet kreve tette oppfølging og nærhet under hele evalueringen.

Forskerne på sin side har behov for handlingsrom – en viss distanse, slik at arbeidet sikres høy forskningsmessige kvalitet og at det er forskningsmessig interessant og mulig.

Siden departementet er oppdragsgiver på evalueringen av egen reform, var det en utfordring å finne en organisering som kan balansere disse to hensyn på en slik måte at den sikrer evalueringens legitimitet. Løsningen ble å organisere en såkalt *Oppfølgingsgruppe*.

*Oppfølgingsgruppen* ledes av en forsker oppnevnt av Norges forskningsråd. De øvrige 10 medlemmene representerer fylkeskommunale myndigheter, elev- og lærerorganisasjonene, partene i arbeidslivet og departementet. Et av medlemmene er hentet fra Danmark for å få inn et komparativt perspektiv og grenseflater mot EU.

Gruppen har som mandat å føre en overordnet kontroll med fremdriften av evalueringen, *kommentere og diskutere utkast til rapporter med forskerne* og gi råd til oppdragsgiver om hva som er problematisk og hva det bør gjøres noe med. Anbefalingene fra Oppfølgingsgruppen inngår som en viktig del av departementets beslutningsgrunnlag når det gjelder styringen og gjennomføringen reformen.

Oppfølgingsgruppen fungerer som en viktig arena for kommunikasjon mellom forskerne, brukerinteressen og oppdragsgiver. Bredden i gruppens sammensetning har bidratt til og bidrar til gjensidig læring slik at de ulike parter – forskerne, brukerne og departementet – har fordel av hverandres kunnskap og kompetanse. De spenninger og interessenemotsetninger som alltid vil være knyttet til *følgeforskningens dilemma mellom oppdragsgivers behov for relevant*

---

*kunnskap til rett tid og forskernes behov for tid til refleksjon og analyser er fortsatt til stede, men det er skapt en arena for meningsfylt kommunikasjon slik at partene kan ha fordel av hverandres kunnskaper og kompetanse.*

### **Løpende rapportering**

Gjennom systematisk underveisrapportering er det tatt sikte på å fremskaffe relevant kunnskap av høy kvalitet til rett tid. Når oppdragsgiver i utgangspunktet har valgt en forskningsbasert evaluering, er begrunnelsene flere. Grepet er først og fremst valgt fordi den kunnskapen som fremkommer vil være knyttet til metoder som borger for at kunnskapen kan inngå som en del av et planleggings- og beslutningsgrunnlag – i departementet så vel som i fylkeskommunene og på skolenivå. Det ligger også til grunn en forholdsvis sterk tro på at forskning er nyttig i møte med usikkerhet.

Forskernes dokumentasjon og analyser skal komme direkte til nytte her og nå. På startsamlingen for evalueringen i juni 1994 ble det derfor fra oppdragsgivers side spesielt understreket at forskerne i sine rapporter også skulle komme med råd og anbefalinger om hva som burde endres og av hvem dersom det viser seg at sider ved reformen ikke virker som planlagt. I sine rapporter har forskerne fulgt opp dette i den grad datagrunnlaget har gitt muligheter for det. Dette grepet har økt rapportenes nytteverdi sett i forhold til målene for evalueringen.

Det er lagt opp til og kontraktsfestet to rapporter hvert år. En hovedrapport leveres **1. februar** i en slik form at den inngår i institusjonens publikasjonsserie. Den andre underveisrapporteringen skjer i form av en artikkelsamling, som publiseres i **juni** i forbindelse med skoleavslutningen. Et profesjonelt forlag legger til rette utgivelsen av artikkelsamlingen.

I første omgang ble det lagt opp til to større halvårsrapporter. Etter råd fra forskerne, ble dette opplegget justert etter halvannet år til det løpet som er nevnt i avsnittet over. Hovedårsaken var kort og godt at forskerne opplevde at det ble for meget bruk av resurser til rapportering og liten tid til analyser og anbefalinger.

Alt forskerne produserer innenfor evalueringen er offentlig. Det er viktig for departementet at kunnskapen blir spredt og brukt.

I forkant av hver hovedrapportering kommenterer *Oppfølgingsgruppe* utkast til rapporter. Dialogen mellom Oppfølgingsgruppen og forskerne har vist seg svært nyttig. Som lederen av oppfølgingsgruppen – professor Kristin Tornes – uttrykker det i artikkelsamlingen «Utdanning for alle» Tano 1996 om denne prosessen: *Den kunnskapen vi da får springer ikke bare ut av ett sett problemstillinger, men mange. I tillegg forandrer de seg underveis ettersom reformen går sin gang.....Vi er altså i ferd med å prøve ut en ny måte å produsere kunnskap på. Dette er kunnskap som framfor alt skal være handlingsrettet og inkludere løsninger på problemer (s.19).*

---

To uker etter møtet med Oppfølgingsgruppen overleveres rapportene til oppdragsgiver, som etter avtale med oppdragstakerne bestemmer tidspunktet for offentliggjøringen.

### ***Helhet og sammenheng***

Evalueringen gjennomføres i dag av syv selvstendige forskningsinstitusjoner:

- 1) Norsk instituttet for studier av forskning og utdanning
- 2) Høgskolen i Lillehammer
- 3) Gruppe for flerfaglig arbeidslivsforskning, UiBergen,
- 4) Norsk voksenpedagogikk forskningsinstitutt
- 5) Forskningsstiftelsen FAFO
- 6) Arbeidsforskningsinstituttet
- 7) Møreforskning, Volda.

En av årsakene til spredningen på flere forskningsinstitusjoner var bredden på oppdraget. For å få alle sidene belyst var det nødvendig å hente inn kompetanse fra flere miljø. For helheten og sammenhengen i evalueringen er det en særlig utfordring. Det har vist seg å være vanskelig å finne gode mekanismer som kan bidra til planmessig kommunikasjon mellom prosjektene. Dette var særlig åpenbart i den første fasen. I dag – 1997 – er bildet noe annerledes. Dialogen og samarbeidet mellom forskerne har det siste året blitt tettere – noe som blant annet viser seg i rapporteringen. Ideelt sett er ønsket at de ulike prosjektene utfyller hverandre og bidrar til en dypere forståelse av reformen og gjennomføringen av den.

### ***Bruk av kunnskapen***

Når det gjelder bruken av resultatene fra evalueringen, må den bygge på grundige analyser, der en blant annet i utgangspunktet så godt som mulig får avdekket hva som er innføringsproblemer og hva som er reelle problemer av mer permanent karakter.

En løsning på et problem vil kunne ha negative konsekvenser på andre områder. Reformen er en sammenhengende vev der alle delene virker inn i hverandre.

Gjennom et så omfattende og langvarig evalueringsprogram fremkommer en betydelig dokumentasjon. Ved den siste hovedrapporteringen var f eks det samlede sidetallet 1200 sider. Det sier seg selv at skal dette kunne utnyttes på en god og hensiktsmessig måte kreves det systemer for formidling og syntetisering av kunnskapen slik at den når frem til brukerne og i en form som er hensiktsmessig. For øyeblikket er dette den oppgaven som har særlig prioritet. Hvordan få brukt kunnskapen her og nå til god nytte for reformarbeidet på alle nivå i opplæringssystemet.

### ***Kompetanseutvikling***

En side ved evalueringen er at den også skal bidra til læring og kompetanseutvikling – i departementet – i fylkene – i skolene og i lærebedriftene –  
**i forskningsmiljøene**

---

For departementet er det en viktig intensjon at evalueringen kan styrke og bygge opp norsk utdanningsforskning og gi forskningsmiljøene slagkraft til å møte nye utfordringer og som kan motivere til relevant utdanningsforskning blant annet på våre ansvarsområder.

Departementet har gjennom EVA 94 hittil investert 22,3 millioner kroner og vil totalt sett ha investert nærmere 30 millioner kroner når evalueringen avsluttes. Det er mao gjort et solid og viktig løft for norsk utdanningsforskning gjennom denne evalueringen.



---

## 10 Erfaringer med evaluering i tre forskningsmiljøer

### 10.1 Norsk institutt for by- og regionsforskning

*Instituttssjef Jon Naustdalslid*

#### **Noen refleksjonar sett fra eit forskarsynspunkt**

(Dette manuskriptet er noe lenger enn innlegget som blei presentert på konferansen)

Dette innlegget bygger i utgangspunktet på mine erfaringar med evaluering sett fra instituttsektoren, nærmare bestemt Norsk institutt for by- og regionforskning (NIBR), som er eit av dei frittstående forskingsinstitutta som er sentrale som utførande fagmiljø for evalueringar. Eg skal ikkje gå i detalj når det gjeld enkeltevalueringar, men heller trekke fram noen meir allmenne og prinsipielle synspunkt som innkast til den debatten som eg veit at det er lagt opp til den etterfølgande gruppediskusjonen og plenumsdebatten.

Sett fra ein forskarsynsstad kunne det vere freistande å bruke ei anledning som denne til å kritisere oppdrags- og brukarsida. Dette er likevel ikkje mi primære vinkling, sjøl om noen av dei synspunkta eg skal gjere meg til talsmann for også kan – og bør – oppfattast som (rett nok konstruktiv) kritikk til brukarsida. Mitt utgangspunkt vil heller vere å fokusere på noen av dei spørsmål vi bør vere opp-tatt av for å få til betre evalueringar og bruke evalueringar på ein meir fornuftig måte.

Det er også på sin plass å seie heilt innleiingsvis at etter mitt syn har det ein gjerne kan kalla «brukarkompetansen» blitt betre dei siste åra, sjøl om dette heilt opplagt varierer ein god del. Særleg i departementa har kompetansen når det gjeld evne til å formulere problemstillingar, innsikt i metode og i dei krava som generelt kan og bør stillast til fagleg arbeid klart blitt betra. Delvis har dette å gjere med rekruttering av forskingskompetente sakshandsamarar, betre kontakt mellom fagmiljøa og departementa og utveksling av personell mellom forskingsmiljø og departement.

La meg først gi uttrykk for eit litt kjettersk synspunkt sett i lys av at eg kjem fra ein institusjon har evalueringar som ein viktig del av levebrødet:

Blir det kanskje gjort for mange evalueringar? Eller sagt på ein anna måte: Ligg det ein fare i at kravet om meir evaluering fører til at fleire evalueringar blir bestilte og gjennomførte meir som rituale enn som uttrykk for eit genuint ønske om å få fram kunnskap? I alle fall er eg ganske overtydd om at ein ikkje ubetydeleg del av dei evalueringane som blir gjort ikkje blir gjennomført fordi oppdragsgivar har ei genuin interesse i evalueringa, men fordi oppdragsgivar har eit ytre krav på seg om å gjennomføre evaluering, som rutine, eller som legitimering. Eg skal ikkje ta på meg å kvantifisere dette, men som utførande

---

evalueringsmiljø møter vi i varierende grad mangel på interesse fra oppdragsgivers side.

I andre tilfelle kan ein møte det motsette problemet: Ei overfokusert interesse i evalueringa fra oppdragsgivers side i den forstand at oppdragsgivar er svært interessert i å få eit bestemt evalueringsresultat – eller meir generelt legge rammer for evalueringa som gjer resultatet meir predikerbart. Eit strengare obligatorisk krav om evaluering, slik som det som ligg i Økonomireglementet, kan føre til at oppdragsgivarar blir meir opptatt av å beskytte seg mot evaluering gjennom å styre og påverke innhaldet.

Her kan det ligg ein fare i at forvaltninga i sterkare grad vil bli interesserte i å kontrollere fagmiljøa. Eg trur ikkje ein løyser dette problemet ved å bryte ned dei nettverka og felles arenaene som eksisterer mellom utførande evalueringsmiljø og forvaltninga – og som Brofoss har peika på i si kartlegging av evalueringspraksis i departementa.<sup>5</sup> Men det er viktig at ein innafor desse nettverka – og på desse arenaene – aukar medvitet om behovet for uavhengighet, fagleg integritet, rollefordeling og kvalitetskontroll.

Det kan også ofte bli stilt for store (eller feilaktige) forventningar til evalueringar. Både oppdragsgivarar og forskarar har etter mitt syn ofte ein tendens til å overdrive betydninga av evalueringar. Oppdragsgivarar spør ofte etter noe som det ikkje går an å skaffe fram – og forskarane eller konsulentane blir ofte freista (eller lokka) til å selge noe dei i grunnen ikkje kan levere.

Mellom anna er det mi oppfatning at det altfor ofte blir stilt opp forventningar om at evalueringar av tiltak og program skal kunne komme fram til eintydige og endelege svar på om eit tiltak er godt eller dårleg, om det har realisert dei mål som blei stilt for det, osv. I praksis er det ofte slik at dei måla som eit tiltak skal evaluerast opp mot ikkje er eintydige – og følgeleg heller ikkje er så enkle å måle eintydige resultat opp mot. Like eins er det ofte slik at dei tiltak som skal evaluerast er konfliktfylte og framstår som positive og vellykka for noen interessentar og som uheldige og problemskapande for andre.

Kravet om at evalueringar skal komme fram til eintydige konklusjonar kan difor av og til vere problematisk. I praksis vil ofte eit slikt krav innebere at evalueringa berre ser på enkelte sider ved eit tiltak – eller belyser det einssidig.

Her kan det ofte ligge kjelder til konflikt mellom oppdragsgivar og evaluator. Dette er konfliktstoff som ein kan møte på allereie i utforminga av oppdraget – formulering av problemstillingar og metodisk opplegg. Det er riktig at evaluering skil seg fra uavhengig forskning ved at eit evalueringsoppdrag først og fremst skal gi svar på oppdragsgivers spørsmål. Dette er likevel ikkje uproblematisk alltid. Særleg når det er tale om å evaluere offentleg politikk, der evalueringar lett blir ein del av den politiske prosessen, har etter mitt syn evalueringsinstansen eit sjølstendig ansvar for ikkje å la seg bruke til å gjennomføre ei evaluering som er for einssidig eller fagleg mangelfull. Det å kunne bestille og gjennomføre

---

<sup>5</sup> Sjå Karl Erik Brofoss' presentasjon av resultatata av undersøkinga første dag av konferansen.

---

evalueringar – og bestemme kva for problemstillingar som skal tematiserast i ei evaluering – kan vere eit maktmiddel.

Eg skal ikkje påstå at vi i dag har eit utstrakt misbruk – det meiner eg ikkje. Men eg meiner at vi kunne ha større oppmerksomhet retta mot evalueringars rolle som styringsreiskap. Eg skreiv for 2-3 år sidan ei bok om «Kunnskap og styring» saman med Marit Reitan der vi forsøkte å sjå kritisk på korleis anvendt forskning – og særleg evalueringar – blir brukte i styringssamanheng.<sup>6</sup> Her brukar vi ein god del plass på å forsøke å grunngi kvifor samfunnsforskning så sjeldan kan brukast som eit eintydig grunnlag for styring. Det har m.a. å gjere med at samfunnsforskningas objekt – samfunnet – består av konkurrerande og ofte motstridande interesser. Sjøl om ei evaluering er aldri så objektiv, vil den svært ofte gripe berre deler av virkeligheta. Følgeleg vil det som regel også vere rom for fortolkingar av korleis eit evalueringresultat skal forståast og brukast. I praksis finn vi da også ei lang rekke eksempel på at evalueringar blir brukte selektivt og fortolka ulikt av ulike interessentar.

Det poenget eg vil fram til her er for så vidt enkelt: Det ligg eit ansvar både på evaluator og på oppdragsgivar i til å akseptere at evalueringar sjeldan leier fram til eintydige løysingar på avgjerdsproblem. Evalueringar i offentleg politikk fjernar ikkje behovet for politiske avgjerder.

I dette ligg det også at etter mitt syn er ein kanskje for mye oppteken av resultat-evaluering eller såkalla summative evalueringar. Både oppdragsgivarane, og samfunnet meir generelt, ville truleg bli klokare dersom vi la meir vekt på evalueringar som vurderer tiltak (f. eks. organisasjonsreformer, forsøk, politikk-endingar) ut fra *ulike* interessentars synsstad og ikkje berre ut fra dei styrande sitt perspektiv. Ei evaluering kan også vere ei kjelde for læring for brukarar og deltakarar, ikkje berre for dei som er styringsmessig ansvarlege.

Kravet om eintydighet og klare svar fører også ofte til at oppdragsgivar forlanger *kvantifiserbare konklusjonar*. Eg er på ingen måte imot bruk av kvantitative metodar. Tvert imot trur eg at det kan vere eit problem at skuleringa i kvantitativ metode er altfor dårleg i ein del evalueringssmiljø. Det kan bli trekt vidtrekkande konklusjonar om årsakssamanhengar på grunnlag av enkle bivariate samanhengar og magert datagrunnlag. Mi bekymring har meir å gjere med at ein ofte forsøker å tvinge ei vanskeleg kvantifiserbar virkelighet inn i kvantitative uttrykk. Oppfordringa mi er å legge større vekt på allsidig metodebruk – kombinasjon av kvalitative og kvantitative teknikkar og større varsemd i tolking og bruk av resultat. Hindringa er som regel for knappe ressursar og ønsket om enkle og eintydige svar på kompliserte og samansette spørsmål.

Oppdragsgivars forventningar kan også vere overdrivne i forhold til ressursar som tid og pengar – eller fordi spørsmålet oppdragsgivar vil ha svar på ikkje er forskbart. Dette gjeld kanskje særleg der det er spurt etter effektar og årsaker. Det er ofte metodisk vanskelegare enn oppdragsgivar er villig til å akseptere å

---

<sup>6</sup> Jon Naustdalslid og Marit Reitan, *Kunnskap og styring. Om forskningens rolle i politikk og forvaltning*. TANO, Oslo 1994.

---

skaffe fram pålitelege svar på kva effektar bestemte tiltak har hatt, m.a. fordi dette som regel krev kontroll over variable som (det alle fall i ettertid) er vanskeleg å kontrollere. Korleis skal ein t.d. kunne evaluere dei sosiale og økonomiske effektane av eit distriktsutbyggingsprogram i eit u-land i løpet av eit tre vekers evalueringsoppdrag når ein veit at dette programmet berre er ein av ei lang rekke faktorar som har virka saman for å påverke forholda i det aktuelle området?

Særleg der oppdragsgivar ønsker effektstudiar burde planlegginga av evalueringa langt oftare enn det som er tilfellet vere kopla saman med planlegginga av sjølve tiltaket. Da kunne ein m.a. fått gjennomført baseline-studiar og kartlegging av før-situasjonen. Dersom planlegginga og gjennomføringa av evalueringa kjem i gang først langt ut i prosessen, eller –dersom det er tale om eit forsøk – når forsøket nærmast er avslutta, manglar ein ofte data om før-situasjonen som vil vere nødvendige for i det heile tatt å kunne seie noe om effektar og verknader.

Ei anna side ved dette er tidsperspektivet. Tiltak og reformer har ofte verknader og effektar på lang sikt som ein ikkje utan vidare kan fange opp i ei evaluering som blir gjennomført etter eit år eller to. Slik sett er oppdragsgivarar ofte altfor kortsynte. Straks ei sak er ute av den aktuelle politiske debatten, mister også gjerne myndighetene interessa for å systematisere læring. Eg etterlyser vilje og evne til å tenke evaluering over tid .

Her er det relevant å peike på den samanhengen det kan vere mellom evaluering og eit anna styringsreiskap som vinn popularitet i offentleg politikk, nemleg konsekvensvurderingar. Ei konsekvensvurdering er ikkje anna enn ei ex ante evaluering. Ein viktig måte å forbetre grunnlaget for slike ex ante evalueringar er nettopp å systematisere erfaringar med pågåande og tidlegare tiltak.

Eg kan også tenke meg å kommentere kort spørsmålet om forholdet mellom forskning og evaluering. For meg er god evaluering meir eller mindre samanfalle med god anvendt forskning. Det er riktig og viktig at oppdragsgivar skal og må ha eit avgjerande ord når det gjeld å bestemme kva for spørsmål evalueringane skal gi svar på. Men det gjeld likevel – som eg alt har peika på – at evaluator ikkje berre skal vere eit passivt metodereiskap for ein oppdragsgivar. Det gjeld kriterier for fagleg redelighet, for metode, for å stille kritiske spørsmål, osv. som langt på veg er dei same for evaluering som anna anvendt forskning. Alt dette gjer at det er viktig at evalueringar blir gjennomført av fagleg kompetente miljø

Oppdragsgivarar burde kanskje av og til også stille strengare krav til prosjekt-opplegg og design enn det som er vanleg i dag. Eg ville ikkje synest at det var urimeleg om departementa – i alle fall når ein bestiller store og ressurskrevande evalueringar – i større grad enn i dag nytta seg av uavhengig fagleg ekspertise til å vurdere søknader og anbud.

Gode evalueringar bør bidra til noe meir enn «eingongskunnskap» – dvs vere relevant utover den enkelte evaluering. Det bør vere eit mål å bygge opp kumulativ kunnskap på dei ulike felt. Gode evalueringar fungerer også slik. Ta

---

eksempelvis evalueringane av frikommuneforsøka på slutten av 80-åra. Kunnskapen som blei bygd opp om kommunal organisering og omorganisering har heilt opplagt bidradd med eit betre fagleg grunnlag for vidare evalueringar av kommunale reformer i 90-åra.

For at dette skal kunne fungere slik, krevst det at evalueringane er forankra i fagleg teori. Her er mange oppdragsgivarar i dag altfor nærsynte og kortsiktig nytteorienterte. Men ein skal heller ikkje gløyme at evaluering stiller ein del spesielle krav til design, metod og ikkje minst gjennomføring. Evaluering er også eit eige fagfelt på tvers av disiplinær. Kompetansen på dette feltet er ikkje alltid like god i evalueringsmiljøa. Med andre ord: God evaluering stiller på mange måtar større krav til teori og metode enn anna ordinær anvendt forskning fordi den krev både ordinær samfunnsvitskapleg teori og metode og evalueringskompetanse. I tillegg er det viktig at evaluator har god substansinnsikt i det feltet som skal evaluerast. Oppdragsgivarar synest ikkje alltid å vere klar over dette, verken når dei vel evaluerer eller når dei vurderer behovet for ressursar.

Når eg understrekar betydninga av at evalueringar må bidra med meir enn eingongskunnskap, har dette også å gjere med at – i alle fall etter mitt syn – skal evalueringsbasert kunnskap vere nyttig også for andre enn den eller dei som skal ta konkrete avgjerder i konkrete situasjonar. Særleg evalueringar av offentleg politikk bør sjåast på som ein del av det allmenne kunnskapsgrunnlaget som samfunnet treng for å forhalde seg til styring, forvaltning og politikk. Ein kan seie det slik at *jo viktigare evalueringar blir som grunnlag for offentleg styring og politikk, jo viktigare er det at evalueringskunnskapen blir spreidd til andre enn dei som ev. brukar kunnskapen til å ta avgjerder.*

Det er med andre ord viktig å skille mellom oppdragsgivarar og brukarar. Ein må jo vente at oppdragsgivaren også er brukar (sjøl om det ikkje alltid er tilfelle). Men i tillegg er den alminnelege ålmenta – publikum – ofte ein heilt relevant brukar. Eksempelvis har vi masse erfaring for at folk flest er interesserte i evalueringar av norsk bistand. Men eg føler meg ikkje overtydd om at bistandsmyndighetene legg til rette evalueringsverksemda i bistanden slik at folk flest skal få eit betre kunnskapsgrunnlag for å delta i denne debatten.

Pressa er ein anna viktig brukar, og eg skulle ønske at pressa i større grad en det som dessverre er tilfelle i dag kunne ta si rolle som forskingsbrukar seriøst. Altfor ofte seg vi at pressa og journalistar er meir interessert i å trekke detaljar og enkeltfunn ut av sin samanheng og forvrengje dei til det ugjenkjennelege for å få salgbare «oppslag» enn dei er i å bidra til ein opplyst samfunnsdebatt.

Forskningsbaserte evalueringar kan også vere relevante som grunnlagsmateriale og case til bruk i undervisning. Her ligg det eit potensiale for bruk og formidling som er altfor dårleg utnytta.

Endeleg er det viktig å hugse på at dei som faktisk blir evaluerte er ei viktig brukargruppe. Evalueringar bør leggest til rette slik at dei som blir evaluerte sjøl har ein sjanse til å lære av evalueringa. Ofte er dette minst like relevant som å

---

bruke evalueringa som grunnlag for f.eks. tildeling av midlar eller som styringsreiskap for leiinga.

---

## 10.2 Diakonhjemmets Internasjonale Senter

*Fagkonsulent Stein-Erik Kruse*

### Evaluering innen bistandsarbeid

#### Utgangspunkt

Mitt utgangspunkt er evaluering av bistand og arbeider p.t. en studie av WHO's landprogrammer og en evaluering av institusjonsutvikling i norsk bilateral bistand.

Siden tiden er begrenset vil innlegget bli konsentrert om to temaer:

1. Evaluering som resultatvurdering – eller hvilke problemer/dilemmaer står vi overfor når resultater skal måles og vurderes.
2. Evaluering er svaret – men hva var nå spørsmålet? Noen kommentarer til hva evalueringer faktisk kan svare på.

#### Vedr. 1. Økte krav til resultater

Ikke bare i Norge, men også internasjonalt er det økte krav og forventninger om at evalueringer skal dokumentere resultater av bistand. Valter Angell ved NUPI tegnet for et par uker siden et unyansert svart bilde av norske bistandsevaluering i et intervju i Arbeiderbladet. Han krevde også mer og bedre resultatvurderinger. Men hva ber han da om? Det er både sant og riktig at hensikt og gode intensjoner om å hjelpe fattige land ikke lenger er nok – resultater skal måles og dokumenteres – ikke bare i forhold til enkeltprosjekters mål, men i forhold til de overordnede og langsiktige mål for norsk bistand.

Årsakene til en slik dreining av oppmerksomhet mot resultatvurdering er også forholdvis klare:

- økte politiske krav til at norsk bistand kan dokumentere verdi, nytte og relevans,
- økt bruk av målstyringsverktøy i NORAD – hvor rapportering av resultater står mer sentralt enn tidligere,
- et stort antall evalueringer som beskriver, analyserer og vurderer mål, planer, prosesser og gjennomføring ved prosjekter og programmer, – men som i svært liten grad innfrir krav og forventninger om dokumenterte resultater. Rapportene kommer godt fra det, ut fra overordnede vurderinger, men fravær av dokumenterte resultater er en gjennomgående svakhet ved de fleste evalueringer og følgelig en hodepine for de fleste evalueringsteam.

Problemene er mange. Det mest vanlige er mangel på data og informasjon for å vurdere resultater, og hvis det finnes data er kvaliteten som oftest tvilsom, eller

---

det er store hull i informasjonen. Et annet problem er arbeidsmåten. Et typisk evalueringsteam består av 3-4 personer med 3-4 uker til rådighet for å finne dekkende svar på en komplisert og krevende oppdragsbeskrivelse. Slike team har få praktiske muligheter til å samle inn egne data. Deres eneste alternativ er som oftest å finlese tilgjengelige rapporter, samt å intervju informanter blant prosjektdeltakere og stab om hvordan fortid og nåtid vurderes mot hverandre. Dette gir som oftest interessant informasjon, – men preget av subjektive oppfatninger og vurderinger. Slike evalueringer er nyttige for visse formål, men illustrerer behovet for å innføre et skille mellom resultatvurdering og evaluering. Eller rettere sagt: se resultatvurdering som en underkategori av samlebetegnelsen evaluering. Resultatvurderinger forutsetter egne metoder og mekanismer, krever andre tidsrammer enn tradisjonelle enkeltstående evalueringer, og er den form for evaluering som mest nærmer seg anvendt forskning.

Kravet om å dokumentere resultater av bistand er forståelig og viktig, men jeg vil også trekke fram en del forhold som ikke bare vanskeliggjør, men i enkelte tilfelle også umuliggjør resultatvurdering av bistandsinnsats – i hvert fall i snever forstand.

**a. Bistandskjeden blir stadig lengre og får flere og flere ledd.**

Den tid skal være forbi hvor norsk bistand finansierte og drev egne prosjekter. I dag legges det vekt på en samarbeidsstrategi hvor mottakerne selv står for planlegging og gjennomføring, og hvor mottakerne befinner seg på ulike nivåer og trinn i bistandskjeden. Et eksempel fra frivillige organisasjoner illustrerer poenget med et mangfold av aktører og nivåer mellom norsk ressursinnsats på den ene siden og resultater på den andre. (Se Vedlegg 1: Frivillige organisasjoner i bistandskjeden). I mange tilfelle kan det være snakk om fra fire til fem nivåer hvor mål på et nivå, blir midler på et annet osv.

I tillegg til interne aktører, kommer det inn også andre bistandsgivere som operer innen samme område, hvis innsats virker sammen med den norske. I mange programmer kan det være mellom fem og ti bistandsgivere med lignende målsettinger og overlappende programområder. I tillegg foregår det meste av NORADs arbeid i land med en svært dynamisk og omskiftelig kontekst. Alt dette betyr at f.eks. NORADs innsats kommer inn som brikke av en større mosaikk – et faktum som ikke gjør resultatvurdering enklere. Å bestemme retning og sammenheng i en årsaks- virkningskjede – mellom bistandstiltak på den ene siden og kort- og langsiktige effekter på den andre – volder ikke bare teoretiske, men også praktiske vansker.

Et sitat fra en forfatter som heter Loasby gir en treffende oppsummering av et velkjent metodisk problem i de fleste evalueringer:

*It is dangerous to assume – either that what has been decided will be achieved, or that what happens is what was intended.*

Hele bistandskonseptet er bygd på forutsetningen om at det går tråder mellom planlagte intervensjoner i den ene enden av bistandskjeden, og måloppnåelse og



---

resultater i den andre, men man skal ikke ha arbeidet mye med bistands-evalueringer for å få anskueliggjort de ofte skjøre, tvetydige, indirekte og komplekse forbindelsene mellom det som planlagt går inn og det som faktisk kommer ut. Dette er på en måte selvsagt, men det oppstår et problem hvis oppdragsgiver krever klarere svar enn hva det er mulig å gi.

**b. Målsettingene blir stadig mer omfattende.**

Et kjennetegn ved de fleste bistandsmål er at de er ideelle, krevende og omfattende.

Norske bistandsmål er knyttet til fattigdomsreduksjon, styrking av miljø og bærekraft, organisatorisk og institusjonell utvikling osv. Målene er også rettet inn mot stadig flere nivåer – fra landsbynivå, gjennom regional utvikling til nasjonal policy planlegging – med økt vekt på politiske og institusjonelle forutsetninger for utvikling.

Mangfoldet i målsettingshierarkiet og intervensjonsnivåer blir godt illustrert i det følgende diagrammet, som er utviklet for frivillige organisasjoner, men som illustrerer generelle problemer. (Se vedlegg 2) Det viser at når vi snakker om resultater av bistand – så snakker vi for det første om resultater på lokalt, nasjonalt, regionalt og globalt nivå. I tillegg opererer vi med målsettinger i forhold til en rekke variable som: «policy environment, management/ organisation, standard of living» som muligens kan kvantifiseres («tangibles») og på den andre en rekke kulturelle og verdimesige variable av kvalitativ art («intangibles»).

I teorien bør alle disse nivåene trekkes inn en evaluering. Konsekvensene for evalueringsarbeidet er mange.

Tradisjonell evalueringsmetodikk er imidlertid primært utviklet og tilpasset vurderinger av enkelt prosjekter og programmer, og de er langt dårligere egnet til vurdering av institusjonelle eller policy – orienterte målsettinger, hvor kriteriene for vellykkethet ofte er andre, mer komplekse og vanskelig målbare.

For det andre er det et voksende gap mellom mål og det som måles. De fleste bistandsorganisasjoner opererer med aggregerte nasjonale og regionale mål, men det som finnes i rapporter er som oftest informasjon om aktiviteter og umiddelbare virkninger av bistandstiltak. For å bruke termene fra logisk prosjektanalyse (se vedlegg 3) så ligger målene på «impact» (resultat) nivå, mens rapporteringen og evalueringene skjer på «output» og i beste fall på «outcome»-nivå. Problemet er selvsagt å skaffe fram god resultatinformasjon, men det er også knyttet til vanskene med å aggregere effekten av enkeltinnsatser. Norsk bistand har f.eks. delmål for enkelt prosjekter og programmer i Tanzania, men operer også med overordnede mål for landprogrammet – som så igjen går inn i et målhierarki for regionen og i siste instans i de globale policy formuleringer. Men hvordan er det mulig – og er det mulig å legge sammen effekten av enkeltinnsatser for så å kunne si noe om måloppnåelse på program, land eller globalt nivå?

---

Svaret fra mange bistandsorganisasjoner er bruk av indikatorer. Ideen er å finne fram til et sett felles indikatorer som kan gjengi måloppnåelse for enkeltprosjekter, men som også kan legges sammen, da de søker å uttrykke et sett med felles kjennetegn for flere prosjekter og programmer. NORAD skal være i ferd med å bygge opp et informasjonssystem hvor alle bilaterale prosjekter skal legges inn i en database og knyttes opp mot en eller flere indikatorer, hvor et sentralt mål er å kunne si noe om måloppnåelse innenfor en sektor, et land, en gruppe med prosjekter e l.

Selv var jeg ansatt av DANIDA og plassert i planleggingsministeriet i Kenya for å etablere et resultatoppfølgingssystem for en gruppe på 40-50 programmer, men mislyktes ganske ettertrykkelig av flere grunner: Vi ble ikke enige om valg av indikatorer, de gode indikatorene var prosjektspesifikke og ikke allmene, det var ikke mulig å enes om felles prosedyrer for innsamling av informasjon mellom flere donorer og til syvende og sist var opplegget for avansert og krevende for et ministerium preget av sterkt indre og ytre turbulens, og store problemer med å innfri elementære krav til planer, budsjetter og regnskapsførsel.

Jeg sier ikke at veien gjennom indikatorer til resultater er stengt, men det er viktig å trå varsomt og ta ett skritt av gangen Mange bistandsorganisasjoner utvikler for tiden kompliserte monitorerings- og evalueringssystemer gjennom bruk av indikatorer, mens erfaringene så langt er blandede og burde mane til nøktern forsiktighet.

### **c. Økt sektororientering**

Et annet kjennetegn ved nyere former for bistand er den økende sektororienteringen, hvor bistandsinnsatsen ikke knyttes til enkeltprosjekter og programmer, men har som overordnet mål å få en sektor til å fungere bedre. I økende grad fokuseres politiske og institusjonelle forutsetninger. Verdensbanken står ofte sentralt i drøfting og planlegging av slike sektor programmer, og de bilaterale partnere gis deloppgaver innen en større helhet. Dette synes som en svært fornuftig vei å gå, men reiser spørsmål og nye krav til evalueringer av slike sektorprogrammer.

Riksrevisjonen i Danmark gjennomførte for en tid tilbake en evaluering av et av DANIDAS sektor programmer i Uganda, hvor det ble pekt på – ikke bare vanskeligheten, men også umuligheten av å evaluere og etterprøve effekten av DANIDAS innsats. Enkeltinnsatser rett og slett forsvant i en kontekst hvor det var å umulig og uønskelig å forfølge og måle effekten av danske skattepenger. En slik anonymisering av et enkelt lands bistand var ikke en uheldig sideeffekt, men en sentral hensikt med sektorprogrammet.

Det er et generelt problem at det ingen finnes en entydig “bottom-line” som gjør det mulig å gi klare svar på hvor vellykket eller mislykket bistanden er. Evalueringer må i stor grad ta hensyn og stilling til verdimeslige spørsmål. Det er også et dilemma at resultater lettere kan vurderes i forhold til enkeltstående prosjektinnsatser, mens bredere sektorprogrammer ikke kan etterprøves på samme målbare måte, men representerer muligens en bedre bistand for mottakerlandet – om enn ikke for Riksrevisjonen i Danmark.

---

### **Konklusjon:**

Det er ikke min konklusjon at resultatvurdering er umulig. Det er ingen vei tilbake til en tid hvor intensjoner var viktigere enn resultater. Det er bare en vei framover, hvor vi må finne løsning på en rekke faglig metodiske evalueringsspørsmål. Dette er viktige og vanskelige spørsmål som krever felles innsats, og som bistandsorganisasjonene ikke må se på som unyttige problematiseringer fra forskernes side. Problemene er reelle nok.

Et siste anliggender: evaluering må beholdes som et omfattende begrep, hvor resultatvurdering er en del av et større sett med virkemidler. Oppstår det et behov eller et spørsmål, er ikke svaret automatisk evaluering, men det er et annet virkemiddel fra den rikholdige verktøykassen evaluering. Det er oppdragsgiver som til syvende og sist må og skal definere sine behov, men evalueringer ville tjene på en grundigere diskusjon av formålet med en evalueringen mellom evalueringsteam og oppdragsgiver – før arbeidet starter:

Den følgende oversikten viser en del alternativer:

- Hvis problemet er knyttet til at beslutningstakere forsøker å unngå/omgå eller utsette en avgjørelse, er svaret ikke evaluering, men økt press for å komme fram til en rask beslutning
- Hvis problemet er mistanke om økonomisk rot eller misligheter, er svaret ikke evaluering, men en revisjon
- Hvis problemet er personalkonflikter, er svaret heller ikke evaluering, men organisasjonsutvikling
- Hvis problemet er manglende informasjon, kan svaret være bedre rapporterings- (monitorerings-)rutiner
- Hvis problemet er for lite eller for dårlig informasjon om årsakssammenhenger, er ikke svaret evaluering, men forskning
- Hvis strategiske mål og metoder for et prosjekt er uklare, kan en ekstern prosessevaluering være svaret
- Hvis problemet er svak lokal deltakelse og respons, kan en selvevaluering eller brukerundersøkelse være svaret
- Hvis behovet er å få vite mer om måloppnåelse og resultater, kan svaret være et eget opplegg for resultatvurdering

<b>SPØRSMÅL</b>	<b>LØSNING</b>
Forsøk på å unngå/omgå/utsette avgjørelse	1. Press for å få til rask beslutning
Mistanke om økonomiske misligheter	2. Revisjon
Personalkonflikter	3. Organisasjonsutvikling
Manglende informasjon om prosjektgjennomføring	4. Bedre monitorering/rapportering
Svak kunnskaper om sektor eller årsakssammenhenger	5. Anvendt forskning
Uklare faglig strategiske mål	6. Ekstern prosess evaluering
Lite deltakelse	7. Selvevaluering
Mangel på kunnskap om resultater	8. Resultatvurdering

---

Svaret kan være evaluering, men det avhenger av spørsmålet. Og i mange tilfelle vil andre virkemidler være bedre egnet og få større betydning.

---

## 10.3 Arbeidsforskningsinstituttet

*Forsker Jon Frode Blichfeldt*

### **Erfaringer fra evalueringsmiljø**

Det er to utgangspunkt for dette innlegget. For det første rammen for denne konferansen: Nytt økonomireglement for Staten. For det andre de praktiske erfaringene med organisering og samarbeid omkring et evalueringsoppdrag Arbeidsforskningsinstituttet (AFI), sammen med en rekke andre forskningsinstitusjoner har for KUF. Oppdraget er å evaluere iverksettingen av skolereformen i Videregående skole, Reform 94. AFIs del av oppdraget er å evaluere «organisering og samarbeid.»

Håpet er at noen av våre konkrete erfaringer fra reformevalueringen kan vise til forhold ved det nye økonomireglementet det kan være verd å diskutere.

Mer sammenfattende synspunkter på evalueringsarbeidet kan fås gjennom en artikkelsamling «Utdanning for alle?» Tano 1996, der også synspunkter omkring gjennomføringen av evalueringen som sådan tas opp. Det er en del av de samme synspunktene som legges fram i dette innlegget. (Blichfeldt 1996).

#### ***Det nye økonomireglementet.***

Økonomireglementet har, slik Engh viste i går – en forhistorie fra begynnelsen av 70-åra med stadig klarere vekt på styring, planlegging og kontroll/vurdering av hvorvidt planer blir fulgt og resultater oppnådd. «*Departementet skal påse at det med mellomrom gjennomføres analyser av virksomhetenes oppgaveløsning for å få informasjon om man oppnår de fastsatte målsettinger med virksomhetens aktiviteter. Ved utvelgelse av aktiviteter er det naturlig å legge vekt på bl.a. politisk viktighet, økonomiske størrelser, resultatutvikling samt antatt effektiviseringspotensial.....Periodiske evalueringer skal gjennomføres slik at resultatet senest foreligger i januar.*» (Økonomireglementet s. 44)

I sammenheng med evalueringer fokuseres det på krav om klare mål- og resultatmål som forutsetning for slik evaluering. Begrepet om klare resultatmål knyttes foruten til ønsket om og mulighetene for styring og kontroll – og også til et begrep om læring. Den klare fokuseringen kan forstås som et ønske om å lære – slik at man får større forutsigbarhet og treffsikkerhet i forhold til tiltak som settes i verk – at man ikke sløser med de ressursene man har til rådighet.

#### ***Erfaringer med å evaluere gjennomføringen av Reform 94.***

I mitt lille innlegg vil jeg snu begrepsbruken litt, og si at evaluering er systematisk synliggjøring og refleksjon omkring erfaringer, slik at man kan lære av dem.

I begrepet erfaring legger jeg noe mer enn resultater og effekt som prestat kan måles. Erfaringer kan synliggjøres i tall – på hvor mange, hvor ofte, hvor lenge – eller hvor kort, men slett ikke alle erfaringer kan det, kanskje de færreste – når det kommer til stykket. Det betyr ikke at man ikke skal være skikkelig og ryddig

---

på det som *kan* måles – og at man måler det som kan måles, men kanskje bør man besinne seg på *hva* man måler.

Når jeg kort skal redegjøre for våre erfaringer med R94, vil jeg forsøke å fortelle en liten historie, fordi den vanligste måten for oss alle er å feste erfaringer til hendelsesforløp, historier som utspiller seg over tid. Det er selvsagt en historie fortalt fra mitt spesielle ståsted, liksom de erfaringene som er lagt fram fra departementet – er fortalt med departementet som ståsted.

Historiene vil være forskjellige – uten at noen av dem derfor behøver være uriktige eller usanne – selv om det for en stor del dreier seg om et handlingsforløp vi deler erfaringer med.

Det gjelder for alle historier at det alltid er noe før begynnelsen (og etter slutten) – men ett sted må man begynne, noen avgrensninger må gjøres.

Jeg begynner samme sted som departementet – med at KUF inviterte norske forskningsmiljøer til en anbudsrunde i forbindelse med den store reformen i videregående skole, R94.

Invitasjonen, som kom før sommeren 93, viste at departementet var godt forberedt, de hadde tatt begrepet «brukerkompetanse» på alvor. Dette hadde de jobbet med en stund. De ville ha forskningsbasert evaluering som kunne gi innspill underveis som korreksjon i gjennomføringsfasen av reformen, og oppdraget skulle også innebære en sluttevaluering av iverksettingen av reformens første fulle gjennomføring – altså et tidsperspektiv på 4-5 år. De hadde presisert hovedområder og mål for hva de ønsket evaluert, hensikten med evalueringen. Allerede innledningsvis var det klart at oppdraget innebar hva man kan kalle underveisevaluering eller følgeforskning. (Statsråden brukte senere, under et seminar ved prosjektoppstart, også begrepene aksjonsforskning og policy-research med henvisning til Coleman 72).

For de fleste forskningsmiljøene ble det umiddelbart klart at oppgaven eller oppdraget var så omfattende at ikke noe forskningsmiljø kunne dekke det alene. Det krevde ulik type metodisk og faglig kompetanse. Det var også slik at de færreste miljøene hadde omfattende erfaring med forskning/evaluering i forhold til videregående skole, enn si arbeid knyttet til den forestående reformen. De hadde bare unntaksvis «jobbet med dette en stund.»

I løpet av høsten 93 ble det derfor samarbeidet intenst mellom ulike institutter for å lage «forskningkonsortier» – som kunne presentere en «samlet pakke» til søknadsfristen ved årsskiftet. Samtidig som noen prøvde å presentere slike sammenhengende «pakker» var man forsiktig med å binde seg for tett inn i konsortiene – man visste ikke hvordan departementet ville gjøre sine valg. Men det innledende samarbeidet nødvendiggjorde at man tittet litt over faggrensene til hverandre – begynte å bli litt kjent.

Departementet valgte miljøer fra litt ulike «konsortier» – og gikk inn i forhandlinger med de enkelte miljøer mot slutten av våren. Slik ble det forberedende samarbeidet som hadde pågått mellom institutter et halvt års tid brutt opp. Det var ikke derfor bortkastet, men man måtte til dels begynne på nytt, og denne

---

gangen under vanskeligere forhold, fordi det måtte gjøres samtidig med at selve arbeidet tok til.

Utkast til kontrakter ble laget – disse inneholdt tidspunkter for halvårslige tilbakerapporteringer, noe som også er nedfelt i departementets virksomhetsplaner. I forhandlinger med de enkelte miljøene ble problemstillinger presisert i forhold til såkalte nøkkelspørsmål. Det praktiske arbeidet ble startet høsten 94 – omtrent da det første reformkullet begynte i grunnkurset. Men kontraktene forelå ikke ferdig undertegnet før godt ut på nyåret.

Det området vi skulle følge, og følger fra AFIs side er «organisering og samarbeid». Det finnes ikke noen enkle, presise eller målbare oppskrifter eller fasitsvar verken på hvordan en organisasjon kan eller skal beskrives – eller på hva slags organisasjons- eller samarbeidsformer som er best. Men det finnes ulike teorier og begreper om det.

Uansett hva slags organisasjon eller organisering vi ønsker å forstå, må konkretiseringen av begrepene «organisering og samarbeid» gjøres i forhold til oppgaver og aktører og i samspillet mellom disse over tid. For å forstå organisasjonen må vi få et begrep om *oppgavene* som skal løses, om *hva*, om hvordan oppgaver kan være gjensidig avhengige, og vi må ha et grep om *hvem* som løser oppgavene *hvordan* – hvilke handlingsvalg og beslutninger som treffes.

Fordi dette var snakk om en sentralinitiert reform som gjennomføres på flere nivåer, valgte vi å følge arbeidet med gjennomføringen ved «vertikale snitt» – fra departementet – via to fylkeskommuner – til fem skoler til sammen i de to fylkene. Vi la opp til observasjoner, intervjuer og dokumentstudier på alle nivåene, og vi la opp til regelmessige tilbakemeldinger og diskusjoner av iakttakelsene våre. (Blichfeldt og Lauvdal 94).

De første månedene måtte vi bruke til å etablere referansegrupper i de to fylkene (med representanter fra alle relevante aktører fra elever – til fylkesmyndigheter og Statens Utdanningskontor), velge skoler – og etablere kontakt, bli kjent.

Ideelt sett ønsket vi å starte sløyfen med tilbakemeldinger og diskusjoner på skolenivå – for så å gå videre til fylkeskommunale referansegrupper som vi fikk satt ned – for til slutt å ende opp med tilbakemelding til det sentrale nivået – først til Oppfølgingsgruppen for diskusjon – og så ferdigstilling til departementet en måned senere. Opplegget var relativt ambisiøst, det tok sikte på mulige læringseffekter underveis på alle de ulike nivåene der vi hentet opplysninger og gjennomførte tilbakemeldinger og drøftinger av erfaringene.

Vi tok det for gitt at den ene av de to halvårslige tilbakemeldingene ville være et relativt kortfattet notat, og at en noe mer omfattende rapport skulle skrives en gang i året – mens hovedrapport skulle komme etter endt løp.

Kontrakten innebar at første tilbakemelding allerede skulle gjøres 1.november til det sentrale nivået. I løpet av høsten hadde vi såvidt fått etablert kontakt på de ulike nivåene slik planen var. Vi hadde fått erfare at det ble jobbet under

---

høytrykk – og ikke enkelt å innpasse intervjuer. Vi hadde verken data av noe omfang eller analyser å melde tilbake lokalt i løpet av de første månedene. Vi regnet derfor også med at den første tilbakemeldingen til departementet måtte bli en enkel framdriftsrapport.

Men Oppfølgingsgruppe og departement ventet åpenbart langt mer, og vi fikk klar beskjed om å utvide og utdype til en langt mer omfattende ferdig rapport 1. desember. Dermed forsvant for en del november som feltarbeids- og datainnsamlingsperiode – samtidig som vi ikke egentlig hadde så mye data å fylle rapporten med.

Erfaringene fra denne første høsten var altså at departementet ga uttrykk for ganske store forventninger til halvårsrapporteringens omfang og dybde. (Disse ble også forsterket av statsråden ved seminaret omkring andre rapportlevering). Vi kunne altså raskt fastslå at det ville bli svært problematisk å få samsvar mellom krav til tidspunkt og omfang for halvårsrapportering – og muligheter til å gjennomføre det feltarbeidet med tilbakemeldinger som rapporteringen skulle bygge på. Vi oppdaget at vi rent praktisk – for å kunne levere til rett tid på høsten, måtte basere oss på å feltarbeid i skolene gjennomført fra midt i juni til slutten av august. Og da er det ikke mange på skolene.

Disse erfarte problemene meldte vi straks tilbake – allerede som del av den aller første rapporten. (Blichfeldt og Lauvdal 94). Dette var det jo den grundigste erfaringen vi hadde med «organisering og samarbeid» så langt. Disse erfaringene liknet jo unektelig også på inntrykkene vi hadde fått gjennom våre første kontakter – om et system som jobbet under høytrykk – og på konstant etterskudd. Vi var heller ikke det eneste forskningsmiljøet som hadde dette problemet. Etter atskillige forhandlingsrunder, og etter ett år – før det som skulle vært den fjerde rapporten, ble også rapporteringsrutinene noe endret. Vi syntes kanskje ikke departementet «lærte» spesielt raskt i denne sammenhengen.

I mellomtiden mener vi å ha klart å gjøre en forsvarlig jobb – dels ved at vi har hatt anledning til å sette i sving en del hovedfagsstudenter som jobber uavhengig av departementets virksomhetsplaner, dels ved å legge inn ekstra tid i den faste staben. Dels har vi måttet gi slipp på deler av vårt opprinnelige opplegg om tilbakemeldinger – spesielt til skolene er tilbakemeldingene blitt dårligere. Vi har lagt inn tidligere bruk av kvantitative metoder, spørreskjema enn vi hadde tenkt. Vi mener likevel vi har klart å få inn en rimelig «rytme» på arbeidet, har klart å beholde noen av de opprinnelige intensjonene. Men denne rytmen og til dels innholdet ble altså noe annerledes enn vi hadde planlagt.

At vi kun hadde kontrakt for ett år av gangen hang også over oss som en slags trussel – og kunne oppfattes som latent mistillit. Miljøene ble også minnet på at det ikke var noen selvfølge at kontrakt ble fornyet. To av forskningsmiljøene ble da også tatt av oppdraget etter et snautt år – noe som var høyst diskutabelt – men det ble klart at de andre forskningsmiljøene ikke var ønsket inn i denne diskusjonen.



---

Nå kan dette høres ut som om vi har hatt et veldig konfliktfylt forhold til departementets medarbeidere. Det har vi faktisk ikke. Tvert om synes vi at vi i diskusjoner og tilbakemeldinger har hatt mange fruktbare og gode møter – og selvsagt også kontroversielle, noe det har vært rimelig åpenhet omkring.

Våre oppfatninger har nok gått i retning av at departementet som oppdragsgiver har vært svært bundet opp i mål- og resultatstyringsbegreper, som i praksis nedfeller seg i sterkt fokus på tempo, tidsplaner og kontroll – og dels også på krav om metodisk representativitet på en måte som kanskje ikke alltid samsvarer så godt med en praktisk virkelighet eller faktiske oppgaver. Det vil si – det samsvarer nok svært godt med *departementets* praktiske virkelighet – der en omfattende nasjonal reform skulle planlegges og gjennomføres på rekordtid. Vi oppfatter det slik at departementets medarbeidere fort blir sittende i en litt «klemt» situasjon mellom politikernes krav og forventninger på den ene siden, og faglige, profesjonelle krav stilt av medarbeiderne i ytre etat (i dette tilfelle skoler, bransjer/lærebedrifter og fylker) på den andre.

Man kan få – slik vi skrev i vår aller første rapport – et misforhold mellom teoretisk tidsbruk og en praktisk fungerende tidsbruk – et misforhold som synes å tendere til å øke jo lenger ut i «gjennomføringskjeden» man kommer.

Jeg bruker gjerne historien om evalueringserfaringene – fordi den jo i logikk og struktur nettopp likner på de historiene vi har fått gjennom feltarbeid lenger ut i «gjennomføringskjeden». Fylker og skoler er også underlagt nokså spesifikke krav om mål-, virksomhets- og resultatrapporteringer. Det blir et praktisk problem når fylkeskommunen sliter med å få nye roller og rutiner som følger nytt inntektssystem til å fungere, når krav om plan- og resultatrapportering skal tilpasses en budsjettprosess som følger kalenderåret, mens skoleåret går fra august til juni – og med helt andre plan- og rapporteringssykluser som ønskelige – eller om man vil «naturlige». Når skoler skal lage planer for etterutdanningsbehov nærmest før de kjenner hvilke læreplaner de skal etterutdannes for – og i alle fall før de har hatt mulighet til å bli kjent med dem.

### ***Ulik vinkling av oppgaver på ulike nivåer.***

Jeg vil understreke at vi synes dette er en uhyre spennende og lærerik prosess. Vi mener den evalueringsmodellen departementet har lagt opp til er spennende og god. Vi har ikke oppfattet noen stor uenighet mellom departementet som oppdragsgiver og forskerne om selve oppgaven, om hva som skulle gjøres – om oppdragets art og retning. Problemene synes å ha oppstått omkring forståelse av oppgaven slik den utspiller seg i praksis.

Dels med referanse til den foregående historien skal jeg prøve å gjøre noen mer generelle refleksjoner basert på disse erfaringene – og mer med blick for det generelle evalueringskravet som ligger i det nye økonomireglementet.

---

Min første refleksjon er knyttet et begrep om *oppgave*.

Iverksettingen av Reform 94 er en uhyre omfattende og komplisert oppgave. Oppgaven som helhet betraktet er politisk bestemt og sentralt gitt. Som vedtak forventes det at den underliggende organisasjonen – fra departementet ned til den enkelte lærer tar den til etterretning. Dette gjelder selvsagt også for forskerne, selv om vi i denne sammenhengen er eksterne, og *ikke* en del av organisasjonens linje.

Enkelte av målene og resultatforventningene i reformen er klart forhåndsdefinert – som dimensjoneringen: Alle 16-19-åringer har rett og skal få et tilbud innenfor tre valg.

Andre mål og mulige resultater – og ikke minst løsningsmåter er imidlertid høyst diskutabile, det gjelder kanskje særlig innhold, undervisningsopplegg. Undervisning og opplæring av unge mennesker kan (eller bør) i begrenset grad ha preg av standardisert forutsigbar produksjon. Tjenlige måter å samarbeide – organisere arbeidet på, bør utvikles i samspill mellom elever, lærere og ledelse – noe innholdsdelen av reformen faktisk også har som siktemål.

Når man i definisjonene for evaluering legger vekt på begrepet *systematisk* er det viktig å være oppmerksom på at det kan, vil – og kanskje bør være helt ulike systematikker knyttet til ulike oppgavetyper – og trolig også ulike bruksområder.

Oppgaver som er klart spesifiserbare, som ofte dreier seg om økonomi, om gjennomstrømming, eller i det hele tatt forhold som kan telles vil nok klarest kunne knyttes til resultatvurderinger slik det legges vekt på i økonomireglementet – og også til instrumentell bruk. Oppgaver som er av et åpnere – og dermed diskutabelt slag, vil snarere kunne knyttes til det Naustdalslid og Reitan (1992) kaller begrepsdanning, som bidrar til å skape de problemoppfatninger vi omgir oss med. Begrepene blir «ting å tenke med» – bidrar både til å feste, forstå og skape erfaringer.

Gjennom vårt evalueringsarbeid har vi foreslått en «dynamisk trekant» som synes å kunne hjelpe oss både til å ramme inn sentrale trekk ved den reformen vi evaluerer gjennomføringen av, og dessuten hvordan gjennomføringsoppgaven kan forstås og håndteres ulikt – og til dels motsetningsfylt mellom ulike aktører og ulike nivåer. (Blichfeldt, Deichman-Sørensen og Lauvdal 96). Modellen kan være et eksempel på en «ting å tenke med.»

Reformen skal ideelt imøtekomme tre typer målsettinger. For det første politiske målsettinger som uttrykkes gjennom rettigheter, idéer om likeverd og like muligheter. For det andre et mål om en rasjonell og økonomisk bruk av tilgjengelige ressurser, «mer kunnskap for hver krone» – om man vil. En av begrunnelsene bak reformen var den sløsing med ressurser som man i det gamle systemet oppfattet lå i at svært mange elever aldri fikk, eller hadde mulighet til å få fullført noen fagutdanning, men gikk sidelengs i systemet fra grunnkurs til grunnkurs. For det tredje faglige mål om en kvalitativt god

---

utdanning som svarer til krav og forventninger fra samfunns- og arbeidsliv, fra bransjer og fagprofesjoner.

Sett fra departementets nivå skal – og bør man selvsagt ha med seg alle målene, ideelt se sammenhengene mellom dem. Våre erfaringer tilsier likevel at problemstillinger knyttet til hva vi kan kalle aksene langs det politiske og økonomiske hjørnet har fått forrang så langt. I vår erfaringshistorie fra evalueringen er det tydelig fokusering på *tempo – tidsplaner og kontroll* i departementets oppgaveforståelse.

Som forskningsinstitutt vil vår forankring være i det «relevante» hjørnet, i at vi metodisk og analytisk kan legge opp et faglig forsvarlig løp.

I det praktiske arbeidet får vi i samspillet med departementet en problematisk relasjon i forholdet mellom det «effektive» og «relevante» hjørnet i trekanten – en relasjon som måtte være (og ble) gjenstand for dialog og forhandlinger om man skulle komme videre.

De politiske spørsmålene har først og fremst vært knyttet til hvorvidt alle får plass, hvorvidt det er opprettet lærlingeplasser nok, hvorvidt tidsplanene for ferdigstilling av læreplaner, handlingsplaner, etterutdanning, evalueringsrapportering osv. holder – om resultatene i forhold til dette. Når læreplan-grupper vil ha nye runder for å sikre faglig kvalitet, eller man går inn i omfattende forhandlinger med lærerorganisasjonene om forståelsen av faglige og/eller avtalemessige forhold, blir det fort plunder og heft i en stramt oppsatt virksomhetsplan.

Og departementet pustes i nakken av den politiske ledelsen og det politiske systemet.

Departementets medarbeidere kan få en problematisk relasjon a) i forhold til den politiske ledelse i spennet mellom det politiske hjørnet og det praktisk effektivt mulige på denne ene side, og b) i forhold til fagforeningene i spennet mellom relevans/profesjonelle faglige krav og effektivitetshjørnet på den andre siden. At fagforeningene også har sin egen politiske agenda gjør ikke situasjonen enklere.

Oppgaver som er diskutabile – og som ikke minst skal håndteres i vårt korporativt forhandlingsbaserte system *må* ta tid.

Fylkenes oppfatning av oppgaven synes faktisk noe ulik. I ett av våre fylker har rettighetskravet, og orienteringen mot skolene – langs aksene likhet/relevans vært sentral, mens det i det andre har vært reformens hovedmodell, kanskje med særlig vekt på relevans i forhold til arbeidsliv og lærlingeplegg som har vært sentralt.

For medarbeiderne i skolene står nok relevansbetraktningene på topp i trekanten – og da krav om relevans i forhold til faglig innhold, i forhold til å kunne ivareta elevenes behov og situasjon, krav som fordrer fleksibilitet – som ofte er av ganske åpen og diskutabel karakter. Tids- og effektivitetskravene fra sentralt

---

hold oppleves fort å komme på kollisjonskurs med det som oppfattes relevant – og dermed faktisk også som ineffektivt i forhold til hvilke oppgaver man faglig synes det er viktig å prioritere.

Oppgaven sett fra vår side som evaluatorene – vil altså være fokusert på relevans. Faglig forsvarlighet, slik vi forstår det, nødvendiggjør et visst spillerom – til å etablere og jobbe i felt, til analyse før tilbakemelding. Som antydnet i historien, må vi forsvare dette spillerommet både for å kunne imøtekomme profesjonelle krav – og for å imøtekomme krav til integritet. Den departementale styringen må ikke gå på bekostning av evalueringens troverdighet, og det er heller ikke vår oppgave å «kjøpe» andre aktørers oppgaveforståelse – enten det er fagforeninger eller politikere.

Vår oppgave er å forsøke å forstå, beskrive og analysere f.eks. hvordan «organisering og samarbeid» kan forstås som et prekært samspill omkring gjennomføringen av en felles oppgave – der ulike nivåer har ulike deloppgaver, og der den enkeltes ståsted i forhold til sin oppgave kan være svært forskjellig slik jeg har forsøkt å vise, og de vil kunne være bestemmende for hvilke handlingsvalg som treffes.

Vi har imidlertid som forskningsinstitutt også interne krav om effektivitet og strategisk tenkning ut over dette spesielle oppdraget. Og dette er nok også relevant for å forstå samspillet mellom forskere og oppdragsgiver. Hvordan lar oppdraget seg kombinere med andre oppdrag? I min virksomhetsplan for 94 sto det svært lite om evaluering av Reform 94. Og som medarbeider i en statsinstitusjon kreves det at virksomhetsplanen er klar. Siden jeg ikke kan basere årets inntjening på ett usikkert anbudskort, må jeg som en annen snekker ha flere baller i luften for ikke å risikere å stå uten oppdrag og inntjening – og må forhandle og justere underveis. Som offentlig forskningsinstitusjon med oppdrag for offentlig etat står vi på en måte mellom to sett mål- og resultatrapporteringskrav – begge med utgangspunkt i historien bak økonomireglementet, og med den type krav som der stilles til presisjon og resultatvurdering.

Av og til føles spillerommet litt trangt – i forhold til forskningsoppgavens egenart. Samtidig er det utvilsomt slik – at det kan gi konstruktive dytt i ryggen å bli stilt overfor den type krav, om da ikke ryggen ryker.

### ***Faren ved å sette kontroll og kvantitet i sentrum.***

Vi antar altså at ulike målhjørner av trekanten får forrang avhengig av hvilket nivå og hvilke oppgaver vi går inn på – det dynamiske samspillet mellom ulike mål blir forskjellig alt etter hva som får forrang, ulike typer problemer, paradokser og dilemmaer vil framstå.

Her vil jeg gå tilbake til innledningen – og problematisere mål- og resultatstyringsbegrepene.

Jeg tror det er viktig at man – i anvendelse av styringsbegrepene balanserer mellom de tre «målhjørnene» – at det nok kan være riktig med ulikt fokus i ulike

---

faser eller perioder f eks av et reformprogram, og i forhold til deloppgaver av ulik karakter – f eks evaluering.

Den mest nærliggende faren jeg kan peke på ut fra egen erfaring – er at man i for stor grad lar «effektivitetshjørnet» bli institusjonalisert, premissgivende – og førende i stor detalj.

Dersom man fra sentralt hold i for stor grad forhåndsspesifiserer, detaljerer eller kvantifiserer de resultater som skal nås og vurderes i forhold til – risikerer man å låse eller svekke dynamikken på de underliggende nivåene. Det gjelder kanskje særlig hvis også tidsmarginene blir knappe..

Mellomledere med stor sans for styring og kontroll av sine medarbeidere kan i økonomireglementet ha fått retningslinjer og verktøy som en knagg også for *utidig* kontroll og styring. Med det mener jeg kontroll som får utilsiktede dysfunksjonelle konsekvenser – i form av opplevd mistillit og manglende spillerom for bruk og utvikling av faglig autonomi, ansvar og initiativ.

I forbindelse med gjennomføringen av Reform 94 ser vi tendenser til dette – både på saksbehandlernivå i fylkene, blant ledere og lærere i skolen. Vi ser det også blant medarbeidere i ytre etater i helsevesenet, på forskningsinstitutter og universitet, i samferdselssektoren.

Man blir pålagt rammer, regler og rapporteringsrutiner som oppleves eller erfares som irrelevant i forhold til det å gjøre en god jobb, til dels ved at diskuterte og kvalitative oppgaver skal kvantifiseres og tallfestes i kategorier som ikke passer. Kanskje oppleves de også å gå direkte på bekostning av den tiden man skulle brukt til primæroppgaven, og på kvaliteten av resultatet. «Fordi vi pålegges så mange og omfattende resultatrapporteringer rekker vi ikke riktig å utføre de oppgavene vi skal rapportere om.»

Det synes å være ulike måter å møte en slik situasjon på: Mange vil «sabotere» – i den forstand at de i hovedsak fortsetter å jobbe slik de har gjort før. Noen vil utøve oppfinnsomhet og fantasi i måten å tolke retningslinjene på – «kreativ budsjettering» er ikke noe ukjent begrep, eller også vise stor aktivitet for fremme alternativer de mener er bedre. Noen vil «melde seg ut» – enten ved å bli sjuke, eller ved faktisk å gå over i annen virksomhet. Man risikerer å bidra til utvikling av en organisasjon med dårlig evne til å lære selv om det motsatte man ville. En risikerer at hovedvekt på kontroll og kvantitet går på bekostning av medvirkning, ansvar og kvalitet.

Oppgavene på sentrale nivåer må være å definere klare rammer og retning – men også å sørge for at det er spillerom nedover i organisasjonen for ulike måter å forstå og løse oppgaver på.

Innenfor definerte enheter, virksomhetsområder (som f eks en skole eller en departementsavdeling) bør man kanskje snarere enn detaljerte tallfestinger etterspørre system og opporganisering for diskusjoner av diskuterte oppgaver – og for hvordan man trekker konsekvensene av det man kommer fram til.

---

Oppgaver som ofte nettopp vil være knyttet til kvalitetssikring og evaluering. Kall det gjerne dialogbasert evaluering – eller å bygge inn evne til læring, intern evalueringskompetanse på alle nivåer i en organisasjon.

Jeg tror for all del at man skal ha oversikt over det som kan måles, teller det som kan telles, men økonomi og tempo kan vanskelig sees som noe hovedmål i vår «dynamiske trekant». God husholdning med ressurser skal være hjelpemidler for å sikre god undervisning, pleie eller transport, ikke mål i seg selv.

Også forutsigbarhet kan være problematisk – kanskje mer jo bedre man lykkes. (Og i denne sammenheng stilles det ofte urealistiske og gale forventninger til forskning). I forhold til det entydig definerte mål og sikre resultat, kan man ha tilsynelatende full kontroll – på det nivået man har avgrenset. Men samtidig blir læring som begrep fort uinteressant – som diskusjonen blir det, og også begreper om medvirkning. Det oppgave- og organisasjonsnivået en har avgrenset vil alltid inngå i større sammenhenger, sammenhenger som kan bli *mer* ustabile og uforutsigbare om man absolutterer kontroll og forutsigbarhet på et underordnet nivå uten forståelse for det større samspillet.

Læringsevne, oppfinnsomhet og kreativitet henter næring fra det som ennå ikke er definert.

-----

#### Litteratur

Blichfeldt, J.F. og Lauvdal, T. (1994) Evaluering av R94. Underveisnotat 1 AFI-notat 9/94

Blichfeldt, J.F., Deichman-Sørensen, T og Lauvdal T. (1996) Evaluering av R94 Underveistrapport IV. AFI-notat 9/96.

Blichfeldt, J.F. m fl: (1996) Utdanning for alle? Evaluering av R94. Tano-Aschehoug.

Naustdalslid, J. og Reitan M. (1992): Kunnskap og styring. NIBR. Rapport 1992:5

---

## 11 Tema frå gruppediskusjonane

Den andre delen av gruppediskusjonane tok for seg sjølve evalueringsarbeidet, frå planlegging og mandatutforming til gjennomføring og bruk av evalueringar.

For å få eit best mogeleg utgangspunkt, var det mange som la vekt på å bruka tid innleiingsvis på å klargjera rammene for evalueringsprosessen. Eit viktig moment her er at både oppdragsgivar og evaluator har ei felles forståing av rollene sine. I tillegg bør ein danna seg ei oversikt over andre interesser innan det området som skal evaluerast. Noko som og er heilt sentralt for korleis evalueringsarbeidet skal gjennomførast, er kva evalueringa til sjuande og sist skal brukast til. Det er viktig at ein i heile prosessen er så realistisk som mogeleg i forhold til tidsbruk og i forhold til kva ein kan få svar på.

Møtet mellom fag og politikk kan ofte vera vanskeleg. Evaluator må få rom til å kunna bevare sin eigen integritet samtidig som ein ikkje kan sjå bort frå den politiske kvardagen som departementa lever med. Det er og evaluator sitt ansvar å ta vare på eigen integritet gjennom ei kritisk haldning til mandat og ved å leggja forskingsetiske retningsliner til grunn.

Det kom fram sterke innvendingar mot bruk av anbudsprinsippet. For det første representerer det ei enorm sløsing med ressursar, for forskningsmiljøa enkeltvis og samfunnsmessig sett. Slik prinsippet blir brukt i dag, er det heller ikkje ein nedre terskel med omsyn til storleiken på oppdrag som blir lyste ut. For det andre fremmar det ikkje gode evalueringar. Oppdraga er ikkje av ein slik karakter at dei kan spesifiserast som i til dømes byggebransjen, og dei bør ikkje spesifiserast av oppdragsgivar åleine. Det skjerpar konkurransen mellom institutta på ein måte som går ut over kvaliteten, og det gjer det umogeleg å dra nytte av den samla kompetansen til miljøa. Som ein mogeleg måte å kombinera ønsket om konkurranse med omsynet til kompetanseutnytting og kvalitet, vart det foreslått å leggja opp ein to-trinnsmodell, ein oppdrags-utformingsfase med dialog mellom departement og relevante fagmiljø og ein gjennomføringsfase der oppdraget blir sett ut på anbod.

Evalueringstiljøa må ikkje bli for avhengige av departementa når det gjeld oppdrag og finansiering. Dei må ha høve til å planleggja si eiga verksemd og utvikla sin eigen kompetanse. Det er dessutan viktig at dei er uavhengige for å kunna stilla kritiske og politisk ukorrekte spørsmål.

Eit forhold som må vera avklart ved innleiinga av eit evalueringsarbeid, er eigeomsretten til sluttproduktet og i kva grad resultata kan offentleggjerast. Det er viktig for evalueringstiljøa å få publisera arbeidet dei gjer. Sjølv om det kan vera vanskeleg for departementa å få offentleggjort informasjon som kan vera negativ for statsråden, bør ein størst mogleg grad gå inn for offentleggjering. Det bør då alt i utgangspunktet for evalueringa avtalast ein offentleggjeringsstrategi og ein fordelingsplan. Oppdragsgivaren bør få rapporten eit par veker før offentleggjering slik at ein kan førebu seg.

---

Når det gjeld bruk av evalueringar, klaga fleire over lange rapportar og lite tilgjengeleg materiale som gjorde det ressurskrevjande å gå skikkeleg inn i arbeidet. Etablering av møteplassar mellom oppdragsgivar og evaluator gjennom bruk av seminar og foredrag kan vera med på å sikra ein meir omfattande bruk av det evalueringsarbeidet som blir gjort.



---

## Innhold

1 Innleiing .....	1
2 Åpning av konferansen .....	3
3 Evaluering i nytt økonomireglement for staten .....	6
4 System og praksis i Danmark.....	9
5 Departementenes evalueringspraksis – en kartlegging .....	11
5.1 Innledning .....	12
5.2 Departementenes evalueringspraksis – et overblikk .....	13
5.2.1 Omfang.....	13
5.2.2 Evalueringstyper .....	15
5.2.3 Tilknytning til departementale prosesser.....	17
5.3 Valg av evaluator .....	19
5.4 Evalueringenes organisatoriske forankring .....	21
5.4.1 Initiativ .....	21
5.4.2 Formålet med evalueringene.....	21
5.4.3 Identifiserte brukere.....	22
5.4.4 Konklusjon .....	22
5.5 Departementenes oppfølging i gjennomføringsfasen .....	23
5.6 Vurdering av rapportene.....	24
5.7 Bruken av evalueringsresultatene .....	25
5.7.1 Omfang.....	25
5.7.2 Brukstyper.....	25
5.8 Evalueringer som et verktøy i økonomistyringen.....	27
5.8.1 Nytt av evalueringene.....	27
5.8.2 Nødvendige forutsetninger målstyring .....	28
5.8.3 En systematisk evalueringsstrategi .....	29
6 Tema frå gruppediskusjonane .....	31
7 Vad bör evalueras? Evalueringar i ett resultatstyringsperspektiv.....	33
8 Erfaring med bestilling, gjennomføring og mottak/bruk av evalueringar. ....	41
9 Erfaringer med evaluering i tre departementer.....	43
9.1 Utenriksdepartementet.....	43
9.1.1 .....	43
9.1.2.....	44
9.1.3.....	45
9.1.4.....	46
9.2 Kulturdepartementet.....	47
9.3 Kirke-, utdannings- og forskningsdepartementet .....	51
9.3.1 Innledning.....	51
9.3.2 Evalueringen av Reform 94.....	51
10 Erfaringer med evaluering i tre forskningsmiljøer .....	57
10.1 Norsk institutt for by- og regionsforskning.....	57
10.2 Diakonhjemmets Internasjonale Senter.....	63
10.3 Arbeidsforskningsinstituttet .....	69
11 Tema frå gruppediskusjonane .....	79

Vedlegg:	Litteraturliste
	Deltakerliste
	Program

---

## Litteraturliste

Blichfeldt, J.F. m.fl. (1996) *Utdanning for alle? Evaluering av Reform -94*. Tano- Aschehoug

Naustdalslid, J. Og Reitan, M. (1994) *Kunnskap og styring. Om bruk av forskning i politikk og forvaltning*. TANO, 1994

SOU 1997:15 *Det svåra samspelet*

Utenriksdepartementet: *Evaluation Report 1-93: Internal learning from Evaluations and Reviews*, by Scanteam international as, January 1993

*Evaluering og resultatvurdering i bistanden. Håndbok for utøvere og beslutningstakere*.

Utenriksdepartementet, august 93. Norsk og engelsk.