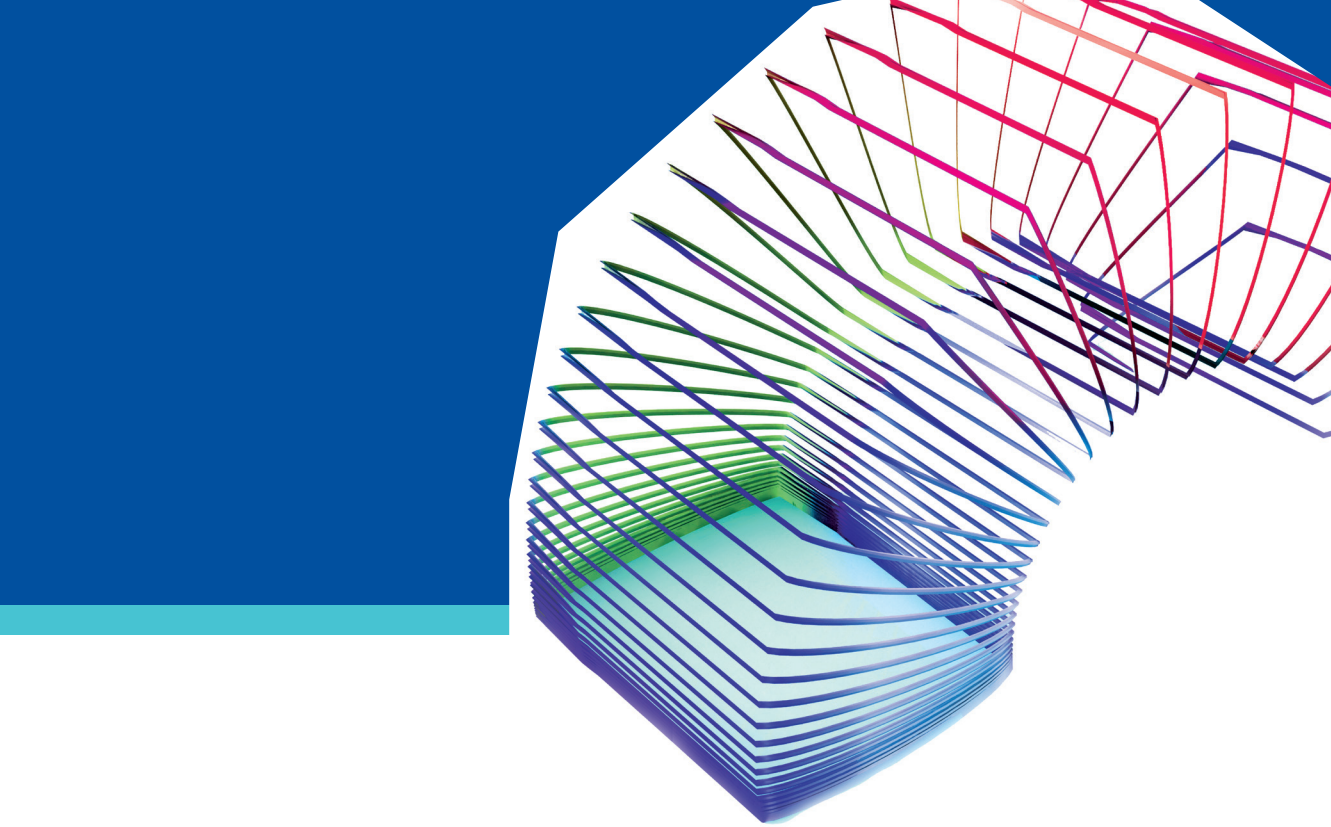




Knut Samset

Prosjektevaluering



CONCEPT TEMAHEFTE **NR. 13**

Knut Samset

Prosjektevaluering

Om forfatteren:

Knut Samset gikk nylig av med pensjon etter å ha vært professor ved Institutt for bygg- og miljøteknikk, NTNU, og leder av Concept-programmet i rundt 20 år. Han er teknolog og samfunnsviter, og arbeidet tidlig i karrieren med fremtidsstudier ved UiO og i regi av World Futures Studies Federation. Han har en omfattende produksjon av vitenskapelige artikler, faglitteratur, lærebøker og mediebidrag, og mange år bak seg som rådgiver internasjonalt, innen planlegging og evaluering av investeringstiltak.

Prosjektevaluering

Rettighetene til innholdet i dette heftet er forskningsprogrammet Concepts.

Dato: mars 2021

Utgiver: Ex ante akademisk forlag

Adresse:

Concept-Programmet : www.ntnu.no/concept

Høgskoleringen 7A

7491 TRONDHEIM

ISBN: 978-82-8433-002-0 (papirversjon)

ISBN: 978-82-8433-003-7 (nettversjon)



Innhold

1 Generelt om evaluering	1
2 Bakgrunn og dokumentasjon	3
3 Sentrale begreper	5
Prosjekter	5
Evaluering	6
Formativ og summativ evaluering	7
Evaluering på ulike tidspunkt	7
Programteori	9
Concepts evalueringsmodell	10
4 Evalueringskriterier	11
Produktivitet	11
Måloppnåelse	11
Andre virkninger	11
Relevans	11
Levedyktighet	11
Samfunnsøkonomisk effektivitet	12
Et eksempel	12
5 Informasjonens kvalitet	14
Gyldighet	14
Pålitelighet	15
Triangulering av informasjon	15
Kvalitativ og kvantitativ analyse	16
Evalueringens fokus	17
6 Metoder for datainnsamling	19
7 Evalueringens etikk	22
8 Rapportering	23
Rapportens struktur	24
Rapportens innhold	24
Detaljeringsgraden	24
9 Erfaringer med evaluering i regi av Concept-programmet	26
10 Bruk og nytte av evalueringer	29
Bruken av evalueringer	29
Tiltak for å lære av erfaringer	31
Hva statlige virksomheter kan gjøre	33
Referanser	35

1 Generelt om evaluering

Ordet evaluering betyr å vurdere eller fastslå verdien eller resultatet av noe. Det har vært brukt i undervisningssektoren til alle tider, i nyere tid i økende grad i andre sektorer, etter hvert som samfunnets kompleksitet har økt. Ikke minst gjelder det tiltak og programmer i offentlig sektor.

En evaluering innebærer systematiske og analytiske studier for å få svar på bestemte spørsmål, til forskjell fra en generell vurdering som er videre og mer upresist definert. (Tornes 2012). Evaluering er ikke eksakt forskning, men heller ikke alminnelig synsing. Den gjennomføres på en systematisk måte og bygger både på vitenskapelige metoder, ekspertkunnskap og sunn fornuft. Det viktigste formålet er å trekke ut kunnskap og erfaring som kan føre til forbedring av fremtidig virksomhet.

Evaluering er ikke noe mål i seg selv, nytten ligger i at resultater og konklusjoner blir tilbakeført til de ansvarlige parter og blir brukt nettopp til å bedre fremtidig virksomhet.

I Finansdepartementets veileder til gjennomføring av evalueringer, (Finansdepartementet, 2006), heter det at:

«Evaluering er systematisk datainnsamling, analyse og vurdering av en planlagt, pågående eller avsluttet aktivitet, virksomhet, et virkemiddel eller en sektor. Evalueringer kan gjennomføres før et tiltak iverksettes (ex ante), underveis i gjennomføringen, eller etter at tiltaket er avsluttet (ex post). Evalueringer kan utføres av interne eller eksterne fagmiljøer.

Evalueringer gjennomføres for å få oversikt og dybdekunnskap. Evalueringer skal bidra til å heve nivået på departementenes styringsinformasjon og den informasjonen Regjeringen gir til Stortinget. Evalueringer er et supplement til løpende kontroll- og styringsinformasjon internt i departementer og virksomheter, og mellom nivåene i statsforvaltningen»

Ansvar for informasjon og opplæring om Økonomiregelverkets krav til evalueringer er lagt til Direktoratet for forvaltning og økonomistyring, som har utarbeidet enkelte veiledere på fagfeltet. Materialet er tilgjengelig via nettsiden [Evalueringsportalen.no](http://www.evalueringsportalen.no).

Denne samler og gir tilgang til Statens evalueringer på ett og samme sted, både for offentligheten, fagmiljøer og allmennheten. Den gir også tilgang til offentlige veiledere på dette fagområdet.¹

¹ <https://dfo.no/fagomrader/evaluering/veiledere-til-evaluering> og www.evalueringsportalen.no

Forskningsprogrammet Concept ved NTNU er en av mange aktører på dette feltet. Vår oppgave er å drive følgeforskning knyttet til statlige prosjekter som kvalitetssikres i henhold til Statens prosjektmodell. Formålet er å oppnå økt samfunnsnyttan av investeringene i etterkant ved å forbedre konseptvalget og prosjektutformingen i forkant. Det er derfor viktig å etterevaluere prosjekter som har vært gjennom ordningen, for å trekke ut erfaringer som kan bidra til å forbedre fremtidige prosjekter. Informasjon om dette og rapportene fra prosjekter som er evaluert er tilgjengelig på programmets nettsider.²

Concept startet med etterevaluering i 2012. En pilotevaluering av fire prosjekter ble gjennomført hvor man benyttet OECDs målorienterte metodikk, bestående av fem evalueringskriterier. I tillegg vurderte vi prosjektenes *samfunnsøkonomiske effektivitet*.³

Denne metodikken ble benyttet videre, og fem år senere var det gjennomført 20 evalueringer. I 2017 ble det foretatt en metastudie basert på evalueringsrapportene fra disse, og intervjuer med ansvarlige, medvirkende og brukere.⁴

Det gjennomføres nå to til fire evalueringer årlig. Det har vært vanlig at Concept-programmet foretar enkelte evalueringer i egenregi, men lyser ut de fleste.

Evalueringene publiseres fortløpende på nettsidene i en app som gir en kortfattet og lett tilgjengelig fremstilling av prosjektets leveranser og prestasjoner. Der kan også rapportene lastes ned i sin helhet.

Dette temaheftet sammenfatter en del av erfaringene forskerne har hatt til nå (2021) med denne evalueringsvirksomheten og forskning som gjelder evaluering.

² <https://www.ntnu.no/concept/etterevaluering-av-prosjekter>

³ Volden og Samset (2013)

⁴ Volden og Samset (2017)

2 Bakgrunn og dokumentasjon

Kvalitetssikring av store statlige prosjekter ble innført av Finansdepartementet i år 2000 med vekt på å sikre prosjekters operasjonelle ytelse (kostnad, fremdrift og kvalitet). I 2005 ble ordningen utvidet til også å omfatte selve konseptvalget og taktisk og strategisk ytelse (effekten, og nytten for samfunnet). Forskningsprogrammet Concept ved NTNU ble fra starten av etablert for å følge ordningen, og gi Finansdepartementet og departementer/etater forskningsbaserte råd underveis.

De prosjektene det her gjelder har en tidligfase og gjennomføringsfase som tar flere år, i noen tilfeller tiår. Skal man etterevaluere for å lære av erfaringer, må prosjektene i tillegg ha vært i drift minst 3-4 år. Bare da kan man vurdere om en har oppnådd forventet samfunnsnytte.

Concept har i skrivende stund (2021) evaluert om lag 30 prosjekter. Men alle disse er fra ordningens første fase 2000-2005, som ikke omfattet kvalitetssikring av selve konseptvalget. Årsaken er at ingen av prosjektene fra den utvidete fasen fra 2005 er kommet så langt at de kan evalueres.

Concept-programmet har utarbeidet et sett med retningslinjer for å sikre at evalueringsarbeidet vurderes og tolkes likt.⁵ Disse tar for seg sentrale begreper, og aspekter ved evalueringskriteriene som det forventes blir besvart i evalueringene. retningslinjene gir også anvisning om hvilken type dokumentasjon som skal hentes inn og presenteres i rapportene.

Formålet med etterevalueringene er som nevnt å trekke lærdom for fremtidig virksomhet. Derfor benyttes samme evalueringsmodell for alle prosjektene. Det gjør at en skal kunne trekke erfaringer på tvers av prosjekter og innenfor sektorer.

Av samme grunn har vi også innført et standardisert scoringssystem. Evaluator er pålagt å aggregere opp sine vurderinger for hvert evalueringskriterium på en skala fra 1 til 6. Det skal brukes for å kommunisere resultatet av evalueringsteamets avsluttende overordnede vurdering av i hvilken grad prosjektene har lyktes.

Evalueringene har også gitt mye erfaring med selve evalueringsarbeidet. Om hva som skal til for å fremskaffe tilstrekkelig omfattende og etterrettelig informasjon, og å sikre at vurderingene er konsistente, uavhengig av hvem som foretar analyser og vurderinger.

Metodikken er nærmere beskrevet i en lærebok utarbeidet ved NTNU, (Samset, 2014). Kunnskap om gjennomføring og anvendelse finner man i rapportene:

⁵ Concept-programmet (2020)

- Concept-rapport nr. 30 etterevaluering av statlige investeringsprosjekter, Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter, Gro Holst Volden og Knut Samset (2015)
- Concept-rapport nr. 52 Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene, av Gro Holst Volden og Knut Samset (2017)
- Arbeidsrapport 2020-3, Bruk og nytte av etterevalueringer av prosjekter, Ingri Bukkestein, Morten Welde og Gro Holst Volden

Dette temaheftet bygger på og sammenfatter vesentlige deler av dette materialet.

3 Sentrale begreper

Evaluering handler om hvorvidt en lykkes. Men suksess er et sammensatt fenomen som handler om flere perspektiver og nivåer, innenfor ulike fagområder som undervisning, forskning, offentlige tjenester, industriproduksjon eller prosjekter. Vi foretar alle evalueringer i det daglige. Men det er blitt et eget fag og en omfattende virksomhet i et stadig mer komplekst samfunn, i økende grad utført av spesialister. I dette temaheftet handler det om evaluering av offentlige investeringsprosjekter.

Prosjekter

Et prosjekt er et tidsbegrenset tiltak som er igangsatt for å oppnå en bestemt effekt. Man bygger en vei for at man skal komme seg raskere fra A til B. Ettersom samfunnets ressurser er begrenset må det stilles spørsmål ved nytten av tiltaket for å kunne prioritere i bruken av ressurser. I etterkant vil det være naturlig å evaluere om man oppnådde det som var forventet. Evalueringen er testen på om man har lykkes.

Det handler altså om å fastslå graden av suksess. Trekker man inn tidsdimensjonen i det som skjer og hvor ansvaret skal plasseres (eventuelt æren for å lykkes) er det nyttig å skille mellom operasjonell og taktisk/strategisk suksess, som vist i figur 1. Det operasjonelle gjelder tiltaket som sådan (produktiviteten), det taktiske gjelder måloppnåelsen, og det strategiske de samfunnsmessige konsekvensene av tiltaket (nytten i sin helhet). Evalueringsmetodikken brukes for å komme frem til kvalifiserte svar på mer konkrete spørsmål, som antydnet i figuren. Var det behov for det? Vil nytten vedvare? Og så videre.



Figur 1 Concepts definisjon av prosjekters suksess

Evaluerings

Det er mange måter å evaluere på, avhengig av sektor, formål, tradisjon etc. En interessant variant av evalueringsmetodikk er den som er etablert av den internasjonale organisasjonen for økonomisk samarbeid og utvikling, OECD, og som tatt i bruk i bistandssektoren internasjonalt, og anbefalt blant annet av FN, EU-kommisjonen og de internasjonale utviklingsbankene. Den benyttes også av Concept-programmet i evaluering av store statlige prosjekter som faller innenfor Statens prosjektmodell, og har en svært konkret og operasjonell definisjon av fenomenet⁶: OECD's definisjon er som følger:

En evaluering er en undersøkelse, så systematisk og objektiv som mulig, av et igangværende eller avsluttet prosjekt, program eller tiltak, og utformingen, gjennomføringen og resultatene av dette. Hensikten er å fastslå prosjektets *relevans, effektivitet, måloppnåelse, virkninger og levedyktighet*. En evaluering skal resultere i informasjon som er troverdig og nyttig, og som bringer erfaringer inn i beslutningsprosessen.

Prosjektet skal altså vurderes i et samfunnsmessig perspektiv. Den angir fem evalueringskriterier, i definisjonen ovenfor markert i kursiv, som til sammen skal gi et omfattende, men allikevel enkelt bilde av status for et prosjekt, som vist i figur 2.



Figur 2 OECD's evalueringskriterier

⁶ Vår oversettelse

Det finnes som nevnt et mangfold av evalueringsformater som brukes i forskjellige sektorer, og til ulike formål innen det offentlige, næringslivet, utdanning, helse og så videre. Vi går ikke nærmere inn på noe av dette i denne sammenhengen.

Formativ og summativ evaluering

I mange tilfeller benyttes evalueringer for å vurdere gjennomføring og resultatoppnåelse, for eksempel ved avslutningen av et prosjekt. Det kalles summativ evaluering og skal gi beslutningstakere og allmennheten dokumentasjon for bruken av ressurser og resultatene som er oppnådd. Det skal bidra til bedre forståelse av prosjektet, dets muligheter og begrensninger.

I andre tilfeller er hensikten å endre prosesser etter hvert som de skjer, ved å gjøre forløpende undersøkelser. Det kalles formativ evaluering, og har læring og påvirkning som formål. Det skal bidra til å gjøre fremtidige prosjekter mer relevante og effektive, ved å trekke ut, analysere og formidle erfaringer fra igangværende og avsluttede prosjekter.

Når det gjelder evaluering i regi av Concept, som arbeider med følgeforskning av landets største offentlige investeringsprosjekter, dreier det seg om en rekke summative evalueringer som samlet sett har et formativt formål. Ambisjonen er å trekke kunnskap ut av aggregater av evalueringer som kan gi nyttig kunnskap for eksempel på sektornivå, og når det gjelder bestemte typer prosjekter, for eksempel vei, bygg, forsvar og IKT.

Evaluering på ulike tidspunkt

Evalueringer gjennomføres på ulike tidspunkt i forhold til prosjektets gjennomføringsfase, som vist i figur 3:

Ex ante evaluering

skjer i forkant for å avgjøre om et prosjektforslag skal gjennomføres eller ikke. Perspektivet bør være bredt, mye på samme måte som med senere evalueringer, for å vurdere om det vil være økonomisk levedyktig, relevant i forhold til brukernes behov, og bærekraftig. I de største norske statlige prosjektene gjøres dette i form av en formell ekstern kvalitetssikring (KS1 og KS2)

Underveisevaluering.

Dette er formativt og gjøres vanligvis for å gi styringssignaler eller som svar på forespørsel eller press fra aktører eller publikum. Årsaken kan være at evalueringen er programmert i utgangspunktet, at prosjektet går inn i en ny fase, at det har oppstått problemer, at det er et behov for å analysere virkningene, etc. Dette faller utenfor den

norske ordningen for kvalitetssikring. Det gjøres i noen grad systematisk i andre land, for eksempel Storbritannia, men da hovedsakelig av virksomhetene selv.

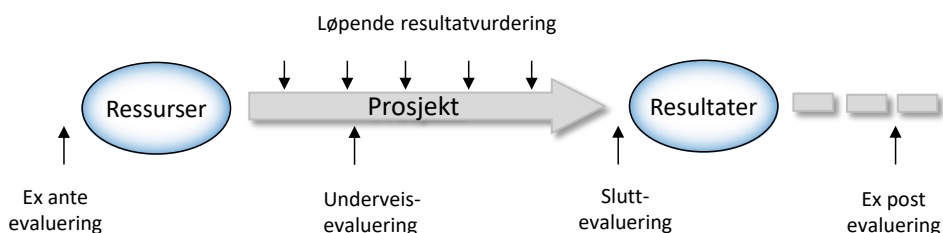
Sluttevaluering.

Dette gjøres som en formell øvelse for å dokumentere resultatoppnåelse ved prosjektets avslutning. Den fokuserer i hovedsak på prosjektleveransen, målt i form av omfang, kvalitet, fremdrift og kostnader, men også i hvilken grad formelt avtalte mål er eller vil bli realisert. Denne typen evaluering skjer internt der det er krav om sluttrapportering fra ansvarlig part, normalt ikke ved hjelp av uavhengige eksterne evaluatorene.

Ex post evaluering.

Dette gjøres for å fastslå og vurdere de mer langsiktige effektene av prosjektet, og i hvilken grad det har bidratt til realisering av formelt avtalte mål, med mere. Motivet kan være å trekke lærdom som kan være nyttig for liknende prosjekter i fremtiden. Slik etterevaluering kan kreve analyser med et bredt samfunnsøkonomisk perspektiv.

Det er ikke krav til etterevaluering av statlige prosjekter generelt. Men evalueringer i regi av Concept-programmet gjennomføres, fordi de inngår i en større følgeforsknings-sammenheng, hvor målet er å trekke erfaringer som kan bidra til bedre prosjekter i tiden som kommer, men også for å utvikle forskningsbasert kunnskap generelt.



Programteori

En evaluering er ikke bare en enkel sammenstilling av hva som går inn i, og hva som kommer ut av et prosjekt. Av hensyn til læring om fremtidig virksomhet er det hensiktsmessig å forholde seg til hele prosessen som fører frem til resultatet.

Programteori, eller teoribasert evaluering, innebærer at evalueringen tar utgangspunkt i, og bygger på en teori eller konseptuell modell som beskriver det fenomenet som skal vurderes. Det kan i sin enkleste form være en årsak-virkningskjede. Den andre ytterligheten kan være en systemdynamisk modell med mange noder og koplinger/tilbakekoplinger. Det kan også være en vitenskapsbasert teori.

Teorien eller modellen danner grunnlaget for utvikling av evalueringsdesign og valg av analysemetoder. Den teoribaserte tilnærmingen er i utgangspunktet metodeneøytral, og baserer metodevalget på hva som er formålet med evalueringen. Det vil si at den benytter den metodikken som i hvert tilfelle er best egnet til å besvare forskningsspørsmålene som ligger til grunn for evalueringsstudien (Lervåg, 2017). En trend de senere årene har vært såkalt *teoribasert evaluering* som supplement eller alternativ til empiriske effektevalueringer. Fokus rettes gjerne mot en begrenset del av målstrukturen som beskrives i detalj. Det kan være aktuelt der tiltaket er komplekst eller innovativt, i tillegg til at datatilgangen er begrenset. En begrunnelse er at dersom evalueringsresultater skal kunne ha overføringsverdi til andre, lignende tiltak, er det ikke nok å beskrive effektene, men det må også redegjøres for hvilken kontekst og mekanisme som utløste dem.

Det kan være nyttig å skille mellom effekter som har solid støtte i litteraturen og slike som ikke har det. For eksempel er bilisters respons på endringer i reisetid og -kostnader godt dokumentert i reisevaneundersøkelser. Derfor kan man med stor grad av sikkerhet anta at en observert trafikkøkning i etterkant av en kapasitetsforbedring er en effekt av prosjektet.



Figur 4 En enkel programteori for lineære prosjekter

Teoribasert evaluering er først og fremst nyttig i tiltak som griper inn i folks atferd og holdninger. Men også i innovative og komplekse tiltak, hvor det er stor usikkerhet om hvorvidt og hvordan tiltakene vil virke.

I evaluering av prosjekter er det vanlig å betrakte det som skjer forenklet som en lineær prosess. Programteorien blir da det som til vanlig kalles prosjektstrategien, med andre

ord prosjekteiers eller oppdragsgivers forestilling om hvordan og hvorfor de planlagte virkemidlene skal føre til det ønskete resultatet, se figur 4.

Årsakssammenhengen og hvert enkelt ledd skal være nærmere definert og kvantifisert, som referanseverdier for evaluator.

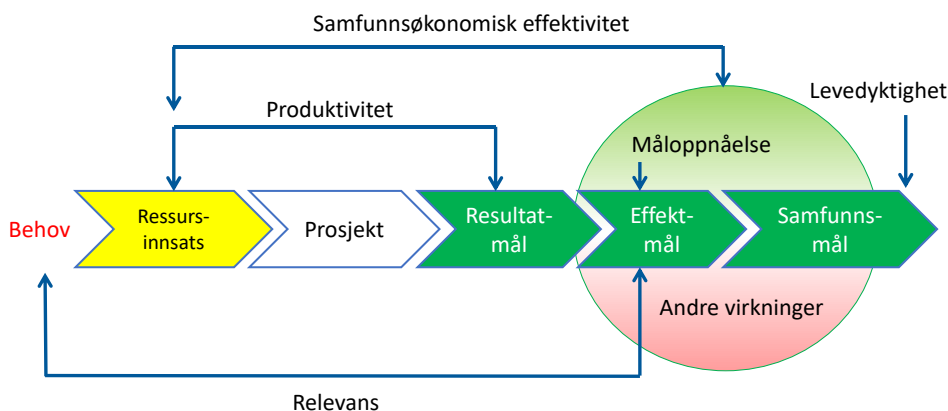
Erfaring viser at det ofte er store svakheter ved programteorien. For eksempel at logikken er feil, ressursene utilstrekkelige, timingen for knapp, eller at målene er trivielle eller urealistiske. Dette gir i så fall et dårlig grunnlag både for planlegging i starten av prosjektet, og evaluering i etterkant. Saken er at dersom de formelt vedtatte målene er urealistisk ambisiøse, vil de ikke kunne realiseres, og evalueringen vil måtte konkludere negativt, selv om resultatet reelt sett er godt. Ingen er tjent med dette.

Det første steget i evaluators arbeid er derfor å foreta en kritisk gjennomgang av programteorien, og melde tilbake til prosjekteier og oppdragsgiver, for å avklare hva som skal legges til grunn ved evalueringen.

Concepts evalueringsmodell

Evalueringsmodellen som benyttes i alle evalueringer i regi av Concept, tar utgangspunkt i programteorien i figur 4, og benytter OECDs evalueringskriterier som vist i figur 2, men har i tillegg tatt med et sjette evalueringskriterium, samfunnsøkonomisk *effektivitet*.

Figur 5 viser hvilke ledd i programteorien som danner grunnlaget for analysen.



Figur 5 Concept-programmets evalueringsmodell

Evalueringskriteriene er nærmere beskrevet i neste kapittel.

4 Evalueringskriterier

Produktivitet

Dette gjelder selve gjennomføringen av investeringsprosjektet, og spørsmålet om hvor effektivt prosjektorganisasjonen har omsatt ressurser til leveranser. En spør om resultatmålene for henholdsvis kostnad, tid og kvalitet ble nådd som avtalt.

Utgangspunktet for vurderingen er Stortingets vedtak og KS2-rapporten med de anbefalingene som ble gitt der. En vurderer også om arbeidet kunne vært gjort billigere, raskere eller med bedre kvalitet.

Måloppnåelse

Dette gjelder realiseringen av effektmålet, og handler derfor om behovstilfredsstillelse for de aktuelle brukergruppene. Som nevnt kan det effektmålet som er formelt vedtatt i mange tilfeller ikke legges til grunn ved evaluering, for eksempel fordi det er trivielt eller urealistisk.

I slike tilfeller må evaluator få en avklaring i forhold til oppdragsgiver og prosjekteier om hvilket ambisjonsnivå som skal gjelde, og legges til grunn i evalueringen, for eksempel ved å vurdere hva som er realistisk oppnåelig i andre tilsvarende eller liknende tiltak.

Andre virkninger

Vurderingen her gjelder alle andre virkninger utover måloppnåelsen som kan tilbakeføres som resultat av prosjektet. En må da tenke bredere og vurdere de samlede konsekvensene av tiltaket, både positive og negative, tilsiktede og utilsiktede, for målgruppen og for andre berørte parter.

Relevans

Spørsmålet her er i hvilken grad samfunnet og de sentrale brukergruppene har et reelt behov for det prosjektet leverer. Dette er altså en vurdering av målene, men denne gangen om disse er i samsvar med brukernes og samfunnets behov eller prioriteringer. At en har lyktes i å nå formelt avtalte mål er altså ikke tilstrekkelig.

Levedyktighet

Dette gjelder det langsiktige og strategiske, altså om hvorvidt de positive effektene av tiltaket kan vedvare over hele levetiden, og om nytten for samfunnet er positiv på sikt.

Dette er en vurdering av om man kan anta at behov og prioriteringer vedvarer, fremtidig ressurstilgang, fremtidig evne og -vilje til å betale for tjenestene eller vedlikeholde infrastrukturen, etc.

Samfunnsøkonomisk effektivitet

Samfunnet har ikke ubegrenset med ressurser, og intet mål er så viktig at det må nås for enhver pris. Midler som brukes på ett offentlig tiltak går på bekostning av andre goder innbyggerne kunne fått.

Samfunnsøkonomisk analyse ser på sammenhengen mellom ressursbruken og samfunnsnyttens som blir resultatet. Men nytten er ofte vanskelig å omregne til kroner. I mange tilfeller må vi derfor nøye oss med å vurdere kostnadseffektiviteten og forhold til andre tilsvarende tiltak. Men også den nytten ikke kan omregnes til kroner skal kartlegges, og en analyse av slike ikke-prissatte virkninger må alltid være med.



Eiksundsambandet. Kilde Statens vegvesen

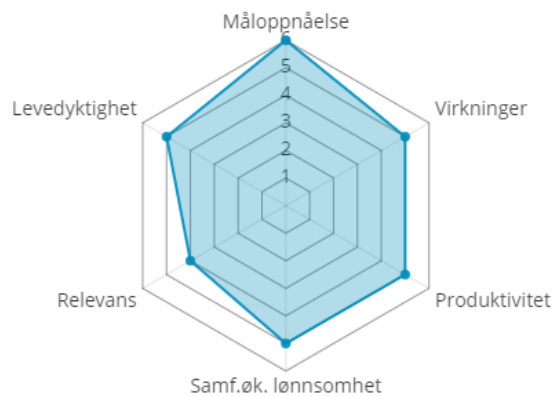
Et eksempel

For å illustrere bruken av disse har vi valgt et veiprojekt som eksempel, kjent som Eiksundsambandet (Ulstein m.fl., 2014): Det gir fast veisamband mellom fire øykommuner og fastlandskommunene Ørsta og Volda, med til sammen 40 000 innbyggere og betydelig industrivirksomhet. Prosjektet omfattet om lag syv kilometer med undersjøisk tunnel, to broer, og bygging og utbedring av tilslutningsveier. Grunnlaget for vurderingen av de seks evalueringskriteriene var først og fremst:

1. Produktivitet: omfang og kvalitet av veianlegget, og sluttkostnad i forhold til det vedtatte budsjettet.
2. Måloppnåelse: reduksjon i reisetid og tidskostnader
3. Andre virkninger: påvirkningen på arbeidsmarkedet og virkninger på det offentlige og private servicetilbudet
4. Relevans: tidsinnsparing og trafikkvekst, som kan slå ulikt ut for forskjellig brukere og berørte parter.

5. Levedyktighet: varighet av positive virkninger
6. Samfunnsøkonomisk effektivitet: nytte i forhold til investering

Evalueringen konkluderte med at prosjektet var meget vellykket. Når det gjelder måloppnåelse ble det full score, det vil si 6 på en skala fra 1 til 6, fordi en med det nye veianlegget oppnådde betydelig reduksjon i reisetid og tidskostnader. Prosjektets relevans ble vurdert lavere, score 4, fordi en mente at nytten var størst for arbeidsmarkedet og servicetilbudet, og at en annen trase kunne gitt større nytte også for andre deler av samfunnet.



Figur 6 Evalueringsteamets samlede konklusjon presentert i et såkalt spider-diagram.

5 Informasjonens kvalitet

Styrken ved og nytten av evalueringer er den informasjonen den bygger på. Det viktigste skillet når det gjelder informasjonens kvalitet går mellom fakta på den ene siden, og erfaringsinformasjon/antakelser på den annen. Troverdigheten må dokumenteres. I tillegg kommer spørsmålet om hva som er aktuelle og ikke minst tilgjengelige informasjonskilder, og hvordan informasjonen fremskaffes og blir brukt. Det handler også om på hvilket aggregeringsnivå informasjonen er nyttig. Primærdata som fremskaffes av evaluator selv antas å ha høyere troverdighet enn informasjon fra sekundære kilder. Ikke minst kommer spørsmålet om kostnaden av informasjonen inn, høy kostnad innebærer ikke nødvendigvis høy kvalitet, det kan medføre at en ikke har ressurser til å fremskaffe annen, viktig informasjon, slik at kvaliteten samlet sett blir lavere.

Erfaringsmessig er det meste av informasjonen man konfronteres med både relevant og av god kvalitet. Men noe kan være motstridende, feilaktig, subjektivt skjev-vurdert, preget av individuelle preferanser og forutinntatthet, og i noen tilfeller direkte forsøk på å desinformere. Det kan være vanskelig å avgjøre hva som er sant og ikke sant. Men ofte er det lettere å vurdere troverdigheten, for eksempel om informasjonen bygger på systematisk analyse eller bare individuelt skjønn. Problemet er at det bare gir en indirekte indikasjon på informasjonens godhet.

Kvalitetskravene som stilles til evaluering er de samme som i forskning, det vil si at informasjonen skal være *gyldig og pålitelig*.⁷

Gyldighet

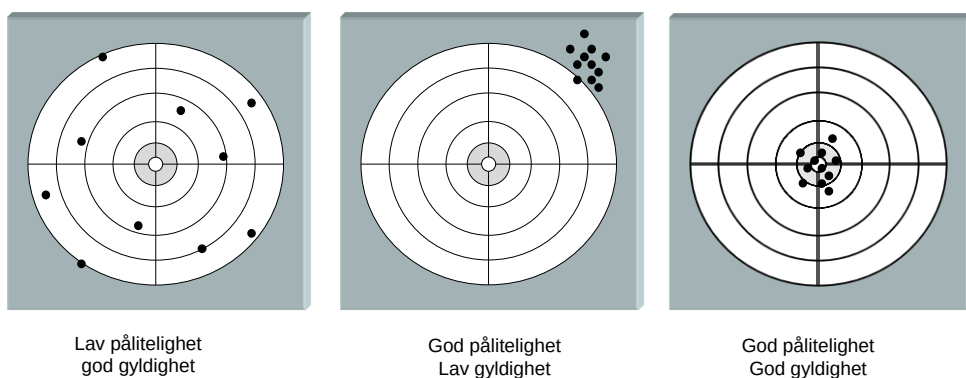
Testen på informasjonens gyldighet er i hvilken grad den gir svar på det vi ønsker å måle. Det handler altså om tolkningen av dataene, og ikke om målingen. Gyldighet betyr at informasjonen er relevant for nettopp den problemstillingen som skal belyses.

Informasjonens gyldighet lar seg ikke fastslå direkte empirisk. Her må det brukes skjønn. Dette vanskeliggjøres ved at man ofte har med kompliserte eller sammensatte problemstillinger å gjøre. For å sikre gyldigheten av informasjonen må man legge vekt på å finne frem til indikatorer som gir et mest mulig *direkte* mål på det man skal undersøke. I tillegg kan man triangulere, det vil si å bruke forskjellige mål samtidig for å beskrive det samme fenomenet.

⁷ Faguttrykkene som brukes i forskning er henholdsvis *validitet* og *reliabilitet*

Pålitelighet

Testen på om informasjonen er pålitelig er om en får samme resultat ved å foreta omtest, med andre ord om resultatet er etterprøvbart. Kravet er at informasjonen skal være så pålitelig at en uavhengig undersøkelse av samme fenomen ville kunne gi tilnærmet identisk resultat. Eller at en får samme resultat selv om en bruker forskjellige mål på å måle det samme fenomenet. Evaluatør må derfor i tillegg til å fremskaffe god nok informasjon, også gjøre rede for de metodene som er brukt, slik at det kan være mulig for utenforstående å etterprøve resultatet. Å sikre at informasjonen er pålitelig er spesielt viktig i spørreundersøkelser der forskjellige personer deltar i datainnsamlingen, eller i løpende registreringer der samme type informasjon registreres over tid av forskjellige personer.



Figur 7 Sammenhengen mellom informasjonens gyldighet og pålitelighet

Både vurderingen av gyldighet og pålitelighet avhenger i en viss grad av om de antakelsene som gjøres og konklusjonene som trekkes i evalueringen har støtte i faktaopplysninger eller konkret erfaring fra virkeligheten. Evaluatør plikter å oppgi kilder i de tilfellene der det er andre som har dokumentert de sammenhengene som beskrives.

Triangulering av informasjon

Evaluatør går inn i situasjoner hvor det er aktører med forskjellige interesser og ulik tilgang til informasjon. Det kan gjøre at respondenter har ulik oppfatning og gir forskjellig informasjon om samme sak. Triangulering er det viktigste hjelpemiddelet man har for å sikre informasjonens kvalitet og trekke riktige konklusjoner. Triangulering vil si å benytte flere respondenter og informasjonskilder, eventuelt angripe problemstillingen fra forskjellige vinkler, eller benytte forskjellige metoder til å skaffe informasjon.

Dette kan være en kostnadseffektiv måte å løse problemet på, for eksempel dersom alternativet er å gjennomføre tidkrevende spørreundersøkelser. Det forventes at

evaluator har vurdert behovet for og muligheten til å benytte triangulering allerede i tilbudsbeskrivelsen for evalueringsoppdraget.

Kvalitativ og kvantitativ analyse

Informasjonens troverdighet avhenger av hva den bygger på. Mye er kvantitativ informasjon basert på underliggende systematisk analyse av fakta, mye er kvalitativ informasjon med utgangspunkt i antakelse og skjønn. Å kommunisere med begreper gjør det lettere å visualisere og drøfte komplekse forhold. Men presisjonsnivået blir høyere dersom argumentene underbygges med tall.

Samtidig vet vi at kvalitativ analyse er viktig for å kunne forstå og beskrive helheten. Det er derfor vi hovedsakelig bruker tekstlig informasjon når vi kommuniserer.

Det er flere motforestillinger mot å basere seg på kvalitativ eller tekstlig analyse i for stor grad: at presisjonen blir lavere, muligheten for etterprøvbarehet ofte er begrenset, og at kvalitativ informasjon ofte preges og/eller avhenger av individers fortolkning, og at det derfor kan oppstå troverdighetsproblemer og misforståelser. Da kan det reises spørsmål om fortolkning og kategorisering.

Tallbasert informasjon er i større grad enn tekstlig informasjon egnet til systematisk analyse som kan gi grunnlag for generalisering eller påvisning av samsvar med stor grad av sannsynlighet. Gyldigheten eller godheten av resultatet avhenger av de antakelsene som er knyttet til dataene som blir brukt. Åpenbart må det være samsvar mellom metodene som brukes og kvaliteten på dataene.

Ved evaluering av enkeltprosjekter er det ofte slik at verken kvaliteten på dataene eller utvalgsstørrelsene gjør det meningsfullt å bruke avanserte statistiske metoder. Det er også slik at mye av informasjonen ganske enkelt ikke kan kvantifiseres. Og det er et kjent fenomen at statistiske analyser av et stort tallbasert datamateriale ofte kan være vanskelig å kommunisere, og derfor kan bidra til å tilslore heller enn å klargjøre et forhold.

Ettersom en i evalueringsarbeid har knapt med ressurser og tid, kan en ikke stille samme krav til presisjon som i forskning. Derfor vil det være aktuelt å benytte enkle analysemetoder.

Kvantitativ metode

- Tallbasert informasjon
- Få opplysninger om mange undersøkelsesenheter
- Stor grad av etterprøvbarehet
- Stor vekt på presisjon
- Generalisering og samsvar som mål

Kvalitativ metode

- Tekstlig informasjon
- Mange opplysninger om få undersøkelsesenheter
- Etterprøvbarehet er ofte problematisk
- Stor vekt på relevans
- Helhetsforståelse som mål

Men man skal ikke overse de viktigste fordelene ved å bruke tekstlig informasjon, nemlig at slik informasjon er grunnleggende nødvendig for å gi en helhetsbeskrivelse av komplekse forhold der det inngår mange påvirkningsfaktorer.

Evaluators må derfor finne en rimelig balanse for å gi en tilstrekkelig helhetsforståelse og presisjon, som dokumenterer det faktiske og underbygger teamets konklusjoner og evalueringsscore.

Evalueringsfokus

Målgruppen for evalueringsrapporter er både spesialistene og allmennheten. Rapporten skal skrives med de minst avanserte leserne i tankene. Dersom analysen bak funnene er basert på avansert statistisk analyse, må de presenteres på en enkel måte, slik at alle kan forstå dem. Enkle tabeller, prosenter og gjennomsnittstall er i mange tilfeller den beste måten å presentere kvantitative data på for de som ikke har spesiell kunnskap om kvantitativ metode. For mer avanserte leserne kan sannsynlighetsfordelinger, percentiler og standardavvik gi nyttig informasjon om hvordan en variabel er fordelt over et populasjonsutvalg. I noen evalueringer er det nødvendig med mer avanserte metoder, både de som bygger på store mengder kvantitative data, og der hvor det er behov for å teste sammenheng mellom variabler. Utfordringen ligger da kanskje først og fremst i å finne måter å kommunisere resultatene på slik at flest mulig kan forstå dem.

I kvalitativ analyse er det viktig å få frem det essensielle. Tekstlige fremstillinger har alltid en hierarkisk dimensjon, det vil si at noe er overordnet og andre deler er underordnet. Det er også årsakssammenhenger, noe forårsaker noe annet. Videre at noe står sterkere enn noe annet fordi det forekommer hyppigere, må tillegges større vekt, osv. Disse tre dimensjonene er sentrale å synliggjøre og drøfte i kvalitativ analyse.



Figur 8 Den vertikale dimensjonen i et informasjonshierarki

Fremstillingen av et saksforhold består av informasjon på forskjellige aggregeringsnivå, som vist i figur 8. Data om enheter på nederste nivå aggregeres opp til informasjon som er nyttig på høyere nivåer for å kunne ta beslutninger der. Det øverste nivået er i prinsippet en beslutning om ja/nei. I en evaluering om hvorvidt prosjektet er vellykket/ikke vellykket.

Evalueringen skal bygge et informasjonshierarki som er sterkt nok til å gi robuste vurderinger av prosjektets vellykkethet. Fremstillingen på overordnet nivå vil i stor grad være kvalitativ, underbygget med noe tallbasert informasjon. På underordnet nivå vil det tallbaserte være mer sentralt.

Evalueringsfokus er viktig. I prosjektets tidlige konseptfase er fokus på det overordnede, det vil si nytten av investeringen for samfunnet. Dels handler det om å avklare overordnede spørsmål om formål, kapasitet, finansiering, etc. Detaljinformasjon er mindre viktig og dessuten usikker fordi mye vil bygge på antakelser.



Figur 9 Prosjektet er et middel til å realisere det som var den overordnede visjonen i starten. Fokus ved ettrevalueringen er derfor i stor grad også på det overordnede og i begrenset grad på det operasjonelle

Men under planlegging og gjennomføring dreier det seg i stor grad om å håndtere mengder av detaljinformasjon.

Ved etterevalueringen er perspektivet igjen løftet til det overordnete, dvs. prosjektets nytte for samfunnet, selv om også gjennomføringen av selve prosjektet inngår i vurderingen som vist i figur 9. Dette illustrere samtidig at informasjonen i en evalueringsrapport må ha hovedvekt på det overordnete, mye vil bygge på kvalitative analyse og vurderinger, tall og faktabasert informasjon er viktig for å sikre dokumentasjon og troverdighet.

6 Metoder for datainnsamling

Evalueringer gjennomføres ofte under betydelig tidspress. Det gjør at mye må baseres på sekundær informasjon, uformelle intervjuer og observasjon. For å sikre kvalitet må det stilles krav til bruken av datainnsamlingsmetoder.

Metodene kan rangeres for eksempel i forhold til kostnad og tid ved bruk. Uformelle metoder på den ene siden, formelle på den annen side, med klart definerte prosedyrer. Nedenfor er det en liste over vanlig brukte metoder, rangert fra uformelle, til presise men ressurskrevende metoder, med en kort karakteristikk av disse.

1. **Bruk av eksisterende data** Basert på gjennomgang av dokumenter både av generell karakter og knyttet til prosjektet som evalueres. En økonomisk og effektiv måte å skaffe informasjon på. Vanskelig å vurdere gyldighet og pålitelighet når det gjelder sekundære data.
2. **Intervju med nøkkelinformanter** Fleksibel tilnærming med mulighet for å gå i dybden. Enkel å gjennomføre. Nyttig både tidlig for å skape innsikt og sent for å verifisere annen informasjon. Fare for subjektiv skjev-vurdering og tolking både fra informanter/intervjuer.
3. **Direkte måling** Registrering av kvantifiserbare eller klassifiserbare data ved hjelp av analytiske instrument. Nøyaktig, pålitelig, og ofte lite ressurskrevende. Registrer bare fakta, ikke forklaringer.
4. **Direkte observasjon** Innebærer inspeksjon, feltbesøk og observasjon for å forstå prosesser, infrastruktur og hvordan denne utnyttes. Egnert for å gi innsikt og kontekst. Avhengig av observatørens forståelse og tolkning.
5. **Fokusgruppe** For analyse av konkrete, komplekse problemer, for å identifisere holdninger og prioriteringer i mindre grupper. Rimelig i forhold til kostnader og effektivt. Nyttig for å generere nye ideer. Fare for ensidighet både hos deltakere og moderator.
6. **Uformell spørreundersøkelse** Innebærer kvantitative undersøkelser med små utvalgsstørrelser. Utstrakt bruk av ikke-sannsynlighetsbaserte utvalg. Rimelig og raskt. Fare for utvalgsfeil/skjevheter og derfor mindre egnert for generalisering.
7. **Casestudier** Grundig gjennomgang av en eller et fåtall tilfeller. Godt egnert til å forstå prosesser og for å formulere hypoteser som skal testes senere. Kan gi omfattende, svært detaljert informasjon, men med begrenset potensial for generalisering.

8. **Deltakende observasjon** Omfattende observasjon på stedet over en lengre periode. Godt egnet til å forstå prosesser, men med begrenset potensiale for generalisering
9. **Formell spørreundersøkelse** Muntlige intervjuer eller skriftlige spørreskjemaer i et statistisk representativt utvalg av respondenter. Datainnsamlingen er krevende, men gir ofte meget pålitelig informasjon

Hvilken metode som er best egnet i en gitt situasjon avhenger på den ene siden av forventet nivå når det gjelder presisjon og troverdighet av informasjon, og på den andre siden de praktiske begrensningene en har i forhold til kostnad, tid og kompetanse. Generelt vil evaluering av virkningene av et tiltak ofte kreve formelle metoder, mens en kombinasjon av mindre formelle metoder kan være mer hensiktsmessig ved evaluering av prosjektets gjennomføring og måloppnåelse.

Den mest brukte datainnsamlingsmetoden er intervju med informanter. Det vil uunngåelig resultere i informasjon som kan være subjektiv og skjev-vurdert og som må verifiseres. Evaluator har derfor behov for supplerende informasjon for å opprettholde en rimelig grad av presisjon og etterrettelighet uten at dette blir for tidkrevende og kostbart.

7 Evalueringens etikk

Evalueringer skal resultere i uavhengig og upartisk vurdering med utgangspunkt i relevant informasjon, fagkunnskap og systematisk analyse. En evaluering kan få betydelige konsekvenser for de ansvarlige og medvirkende i prosjektet som evalueres. Det stilles derfor både faglige og etiske krav til evaluator når det gjelder:

1. Anonymitet/konfidensialitet

Evaluator skal respektere folks rett til å opptre anonymt som informasjonsgivere, og sørge for at data ikke kan spores tilbake til den enkelte.

2. Hensynet til respondenter

Å avgi informasjon er frivillig. Respondenter skal informeres i forkant om evalueringen, formålet og deres eventuelle bidrag. Intervjuet bør avgrenses til de sentrale temaene og gjennomføres effektivt.

3. Hensyn til helheten

Evaluator har et ansvar for også å belyse forhold som ikke er direkte nevnt i oppdragsbeskrivelsen, dersom dette er nødvendig for å forstå helheten.

4. Kritiske vurderinger og dissens

Å gi uttalelser på vegne av evalueringsteamet forutsetter at alle medvirkende har hatt anledning til å ta stilling til spørsmålet. Eventuelle dissenser med hensyn til konklusjoner eller anbefalinger skal nevnes i rapporten.

5. Evaluering av individer

Å evaluere enkeltindivider er ikke en del av en evaluering. Det er funksjonen som skal vurderes, ikke individet.

6. Avdekking av feil eller regelbrudd

Evalueringer vil iblant avdekke feil eller regelbrudd. Evaluator skal i slike tilfeller sikre at individer ikke eksponeres i rapporten.

7. Informasjonens gyldighet

Medvirkende i prosjektet og berørte parter har rett til å bli hørt i spørsmål om hvorvidt informasjonen er korrekt. Det er imidlertid evaluator som har det siste ordet. Enkeltpersoner skal ikke gis anledning til å påvirke utfallet.

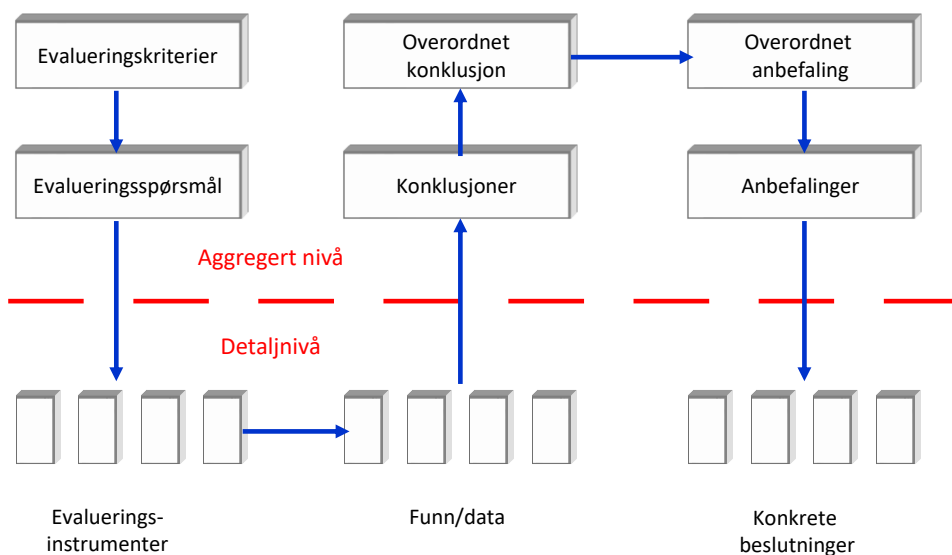
8. Offentliggjøring

Prosjektets medarbeidere og ansvarlige i berørte organisasjoner er gjerne interessert i å få evalueringsteamets konklusjoner så tidlig som mulig. Generelt bør evalueringsteamet vise skjønn med hensyn til hva som kan sies før overlevering av evalueringsrapporten, som skal skje direkte til oppdragsgiver.

8 Rapportering

Evalueringsrapporten er sluttresultatet uansett hva som er evaluerings formål, og om det er en summativ eller formativ evaluering. Den skal bli lest og forstått, men fremfor alt være tilpasset det som er formålet. Dette handler i stor grad om organisering og aggregering av informasjonen, som skissert i figur 10.

Det er viktig at formatet er enkelt, rapporten godt strukturert, og med et språk som er lett tilgjengelig. Rapporten skal skrives med tanke på at også lesere med lite kjennskap til sakskomplekset skal kunne forstå innholdet. Figur 10 illustrerer informasjonsprosessen i evalueringer. Evalueringsrapporten skal hovedsakelig presentere informasjon på aggregert nivå, underbygget med funn og data på detaljnivå der det er nødvendig. Den skal ende opp med anbefalinger som kan gi grunnlag for konkrete beslutninger i etterkant. I noen tilfeller er svaret opplagt, men som en regel, skal anbefalingene være på et tilstrekkelig aggregert nivå til at det gir rom for å vurdere alternative beslutninger.



Figur 10 Aggregering og disaggregering av informasjon i løpet av evalueringsprosessen. Hensikten er å komme frem til godt funderte konklusjoner og anbefalinger på et aggregert nivå.

Det er et mål at evalueringsrapportene ikke blir veldig omfangsrike og uoversiktlige. De skal være godt strukturert i den forstand at de ikke presenterer for mye detaljinformasjon sammen med de mer overordnede vurderingene og konklusjonene.

Rapporten skal ikke presentere informasjonen sekvensielt, det vil si med mange detaljer først som bygger opp mot konklusjonene som presenteres til slutt. Prinsippet skal være det omvendte: å starte med konklusjonene, og deretter underbygge disse med det som er nødvendig av informasjon og analyse - men ikke mer. Her er noen generelle råd om strukturering og innhold i evalueringsrapporter:

Rapportens struktur

- Rapporten bør ha et sammendrag som er utfyllende i den forstand at det kan leses uavhengig av rapporten, men allikevel presenterer det som er essensielt for beslutningstakere og som gjelder bakgrunnen for evalueringen, de viktigste konklusjonene, anbefalingene og erfaringene man kan trekke av rapporten.
- Rapporten bør være bygget opp rundt konklusjonene og anbefalingene som underbygges med mer detaljert informasjon, i den grad det er nødvendig.
- Mer detaljert informasjon bør tas med i vedlegg. Rapporten skal i tilstrekkelig grad referere til dette materialet med sideangivelser. Evalueringsmandat og evalueringsprogrammet med intervjuliste og dokumentliste inngår som vedlegg.

Rapportens innhold

- Hovedpunktene bør om mulig tydeliggjøres ved hjelp av illustrasjoner, diagrammer, tabeller, etc., utstyrt med figurtekster som får frem poenget uten at det er nødvendig å lese teksten.
- De viktigste faguttrykkene bør defineres i teksten og tas med i egen ordliste som vedlegg. Lange og sammensatte setninger bør brytes opp i kortere setninger. Språkformen bør være aktiv i størst mulig grad, mens passiv språkbruk hovedsakelig brukes til å rette oppmerksomheten mot ting som skal vektlegges.
- Bruk av forkortelser bør generelt unngås i teksten. Akronymer, fagtermer og lokalt brukte betegnelser, må forklares.
- Tall bør presenteres i tabeller og diagrammer. I teksten bør man avgrense til å bare referere de viktigste tallene og forklare betydningen av disse. Prosenttall bør i teksten avrundes til nærmeste heltall eller titall.

Detaljeringsgraden

Evalueringsrapportens detaljeringsnivå er derfor sentralt. Tendensen er at dette ofte blir for høyt og at det kan tilsløre det overordnede perspektivet. Behovet for detaljering øker som nevnt nedover i et beslutningshierarki og er størst på det nederste utøvende nivå.

Evalueringsprosjekter er i større grad rettet mot det overordnede nivået, det vil si departement og etat.

Det er en kjent sak at i digitaliseringens tidsalder er tekstmengden blitt et problem i seg selv. Det er blitt for enkelt å skrive omfangsrike rapporter, og vi ser at sannsynligheten for at rapportene blir lest og benyttet avtar med tykkelsen. Det er blitt en utfordring å bidra til å snu denne trenden. I en studie om hvorvidt regjeringen har fått et bedre beslutningsunderlag som følge av Statens ordning for kvalitetssikring, mente flere at detaljeringsnivået ofte blir for høyt allerede i starten ved utredningen og kvalitetssikringen av konseptvalget, (Kvalheim, m.fl., 2015).

Finansdepartementet har utarbeidet en egen veileder om detaljeringsgraden i konseptvalgutredninger i prosjektets tidligste fase (Finansdepartementet, 2010). Der heter det at kravet til presisjon og detaljering nødvendigvis ikke skal være meget høyt. Det gjelder dels definisjonsgraden av kostnadsestimatet på forskjellige tidspunkt i prosessen, men også at hvert konsept kun beskrives til den detaljeringsgraden hvor det er mulig å bedømme det i forhold til andre konkurrerende prinsipielle løsningsforslag. Videre at utredningen av konseptvalg ikke bør være på mer enn 30-50 sider, med mindre det er mange alternativer som vurderes. Det en bør sikte mot er først og fremst en systematisk grovsiling av alternative konsepter med tanke på å fjerne de dårligste. Det er en kjent sak at de fleste KVV-utredningene er langt mer detaljert og omfattende.

Det samme bør gjelde også for evalueringsrapporter. Disse er også overordnede dokumenter som ikke skal gå for mye ned i detaljer. Det skal gi en tilstrekkelig god beskrivelse og vurdering av hva prosjektet har oppnådd strategisk, taktisk og operasjonelt, samt hva man kan lære av dette for fremtidige prosjekter i tilsvarende sektor.

9 Erfaringer med evaluering i regi av Concept-programmet

I 2021, 20 år etter at Finansdepartementets ordning for ekstern kvalitetssikring av landets største statlige investeringsprosjekter⁸ ble etablert, har i overkant av 300 prosjekter vært gjennom ekstern kvalitetssikring i tidligfasen. En tredjedel av prosjektene gjennomført, og om lag 30 av disse er etterevaluert i regi av forskningsprogrammet Concept.

I 2013 gjennomførte programmet en pilotevaluering av fire prosjekter for å teste ut og velge metodikken for etterevaluering. (Volden og Samset, 2013). Studien konkluderte med at metodikken som var benyttet var hensiktsmessig, og ga et godt grunnlag for læring på tvers av prosjekter. Noen sentrale anbefalinger var:

1. Evalueringsopplegget må være generisk fleksibelt slik at konkrete indikatorer og datakilder kan tilpasses det enkelte investeringsprosjekt.
2. Det er viktig at evalueringsteamet er uavhengig, upartisk og har god evalueringskompetanse og erfaring fra evalueringsarbeid.
3. Teamleder være den som har mest evalueringskompetanse.
4. En mente at gjennomføringen burde kunne gjøres med en arbeidsinnsats på omtrent tre månedeverk,
5. Sluttrapporten bør være begrenset mht. sidetall.
6. Fokus bør være på hva som er «godt nok», men det bør stilles høye krav til å redegjøre for forutsetninger og begrensninger.
7. Informasjonens gyldighet må stå sentralt ved valg av indikatorer.
8. Triangulering er derfor viktig.
9. Statlige virksomheter og departementer som mottakere av informasjonen må finne et opplegg som passer inn i den øvrige virksomheten slik at evalueringene kommer til nytte og bidrar til forbedringer av fremtidige prosjekter.

I 2017 ble det gjennomført en meta-evaluering basert på de første 20 evalueringene som til da hadde vært gjennomført, (Volden og Samset, 2017). Figur 11 viser resultatet på øverste aggregeringsnivå der teamene har gitt en poengscore for hvert av de seks evalueringskriteriene som angir teamets samlede vurdering av investeringstiltaket

⁸ Omtales i dag som Statens prosjektmodell

vellykkethet. Det benyttes en skala fra 1 (helt mislykket) til 6 (svært vellykket). Score 4 brukes der resultatet er akseptabelt, men ikke noe mer.

Sektor	Antall prosjekt	Produktivi- tet	Mål- oppnåelse	Andre virkninger	Relevans	Leve- dyktighet	Samf.øk. effektivitet
Bygg	5	5,4	4,2	4,6	4,6	4,8	3,8
Forsvar	2	4,5	4,5	4,5	4,5	3,5	3,5
IKT	2	5,0	5,5	4,5	4,0	5,5	4,0
Jernbane	3	4,3	3,3	4,0	4,7	4,7	2,7
Vei	8	4,4	5,3	4,3	4,6	4,5	5,3
Snitt		4,7	4,7	4,4	4,6	4,6	4,2

Figur 11 Sammenfatning av evalueringsscore for de første 20 prosjekter som er evaluert

Studien viste at i gjennomsnitt scoret de 20 prosjektene gjennomgående godt, med et gjennomsnitt mellom 4 og 5. Samfunnsøkonomisk effektivitet fikk lavest score, mens produktivitet og måloppnåelse scoret litt høyere. Videre ser vi at veisektoren scorer høyest og jernbanesektoren lavest.

De viktigste konklusjonene i denne grundige og godt dokumenterte studien var at:

- Evalueringene bør få frem mer informasjon om bakgrunnen for og tidligfasen i prosjektene. Nyttene av evalueringene avhenger av hva man kan lære av dem. For å lære må man forstå årsakene
- Formelt avtalte mål (programteorien) må være realistisk oppnåelige, ikke trivielle eller uoppnåelige. Evaluatoren må vurdere målstrukturen kritisk, og ikke legge til grunn en prosjektlogikk med store feil
- Det må jobbes mer med å involvere departementer og etater slik at de ser nytten og læringspotensialet i evalueringene. Informasjon om hensikten med arbeidet og enkel tilgang til nøkkelresultater er viktigst
- Kvaliteten på evalueringene hviler sterk på at evaluatoren anvender en bred tilnærming og triangulerer mellom ulike kilder og metoder. De fleste gjør dette, men noen har et forbedringspotensial
- Mange av teamene har hatt et flertall av samfunnsøkonomer. Faglig bredde er viktig for å unngå en skjev tilnærming og mangelfull vurdering av prosjektet. Hvert team skal ha tilstrekkelig kompetanse på fire områder: evalueringsfaglig,

samfunnsøkonomisk, prosjektfaglig (herunder estimeringskompetanse) og sektorfaglig.

- Uavhengighet til prosjektet som skal evalueres er et selvsagt krav, med mindre en ønsker å involvere medarbeidere i den aktuelle etat i evalueringsarbeidet for å styrke læringseffekten.
- Det er nyttig med en viss sektorkunnskap – noe som vil variere med typen prosjekt.
- Avgrensning til prosjektet (ikke tiltaket) og evaluering +/- 5 år ut i driftsfasen, er et godt utgangspunkt som gjør evalueringen håndterbar. Prosjektet må ses i sin kontekst.
- I neste runde blir det særdeles viktig å evaluere prosjekter som har vært gjennom KS1⁹. Da får en vurdert effekten av den utvidete ordningen hvor en også er opptatt av selve konseptvalget, for å sikre at Staten også velger det strategisk beste, ikke bare gjennomfører prosjektene på best mulig måte operasjonelt og taktisk. Evalueringen av de 20 første prosjektene vil utgjøre en kontrollgruppe.
- Bruk av poengscore for å oppsummere vurderingene er nyttig både for evaluator og brukerne. Men det må være tilstrekkelig sikkerhet for at vurderingene er konsistente på tvers av evaluatorene. Det er derfor behov for felles retningslinjer.

Dette siste blir ivaretatt ved at Concept har laget et hefte med retningslinjer for scoresetting. Hftet ble første gang publisert i 2015 og oppdatert i lys av metaevalueringen i 2017. Siden er det gjort noen mindre oppdateringer. Gjeldende versjon kan til enhver tid lastes ned fra nettsiden her

<https://www.ntnu.no/concept/etterevaluering-av-prosjekter>.

Hftet gir også nærmere anvisninger som gjelder kravet til bakgrunnsdokumentasjon, hvilke kostnadsestimater og tall som skal registreres, og gjennomføringen av selve evalueringsprosessen.

⁹ Per år 2021 er svært få av disse ferdigstilt og kommet så langt (3-5 år inn i driftsfasen) at de kan evalueres, men om relativt kort tid vil det være mulig

10 Bruk og nytte av evalueringer

Bruken av evalueringer

Evalueringer går inn i problemstillinger som i mange tilfeller er svært komplekse, hvor det er mange parter involvert med ulike interesser i forhold til prosjektet. Verdien av evalueringer er først og fremst at de presenterer informasjon som er blitt til ved systematisk innsamling og sikring av kvaliteten og kan ansees som fakta.

Formålet er å fremskaffe relevant informasjon for beslutningstakere på riktig tidspunkt, og nyttig erfaringskunnskap i læringsprosesser for fremtiden. Men erfaringen er at evalueringer ikke blir brukt i særlig grad. Skal det skje må informasjonen inn i de riktige kommunikasjonskanalene og frem til brukerne.

Den internasjonale organisasjonen for økonomisk samarbeid og utvikling, OECD, skriver om dette: *«Tilbakemelding er en viktig del av evalueringsprosessen som gir en kobling mellom tidligere og fremtidige aktiviteter. For å sikre at resultatene benyttes i fremtidig policy og prosjektutvikling, er det nødvendig å etablere mekanismer som involverer alle parter. Dette inkluderer tiltak som bedømmelseskomitéer, seminarer, workshops, systematiske systemer og rapporterings- og oppfølgingsprosedyrer.»*

Det gjøres en del på dette området. I offentlig sektor gjennomføres stadig flere evalueringer. Disse er åpent tilgjengelig for allmennheten, for eksempel gjennom www.evalueringsportalen.no. Den inneholder alle evalueringer utført på oppdrag fra statlige virksomheter fra 2005 og til i dag, i tillegg til et utvalg sentrale evalueringer fra perioden 1994 til 2004.

Når det gjelder evalueringer i regi av Concept er det lagt spesielt til rette for dette, ved en søkbar app på programmets hjemmeside der man både kan få bred oversikt over evalueringsresultatene fra flere prosjekter innenfor ulike sektorer, og også kan klikke seg ned i informasjonshierarkiet og få mer detaljert informasjon som gjelder hvert enkelt prosjekt og individuelle evalueringskriterier.

En Concept-studie om bruk og nytte av etterevalueringer (Bukkestein m.fl., 2020) fant at det er stor interesse for evalueringene både i departementer og etater. Mange av informantene mente at det er flere i egen organisasjon som vil ha nytte av dem. Men de fant samtidig at målgruppen har varierende kjennskap til Concepts evalueringer. Videre at KVVU-er og KS1-rapporter sjelden eller aldri viser til tidligere evalueringer. De fant også at etatene kun i liten grad foretar systematisk etterprøving eller evaluering av om prosjektene når sine mål. Evaluering skjer ad hoc, primært når noe har gått galt, og det kan være uklart hva som er departementenes og hva som er etatenes ansvar å følge opp.

De ansvarlige etatene har ordninger for erfaringsinnhenting og legger vekt på at resultater fra gjennomførte prosjekter skal benyttes i planleggingen av nye. Det er imidlertid utfordringer knyttet til systematisering og tilgjengeliggjøring av informasjon, og det meste handler kun om realisering av resultatmålene, ikke om den overordnede samfunnsnyten.

Studien gir en rekke anbefalinger om hvilke forutsetninger som må innfris, eller barrierer som må overkommes, for å sikre at etterevalueringer kommer til nytte:

1. Departement/etat bør konsulteres tidlig for å få innspill til hva som er spesielt interessant å få belyst i evalueringen.
2. Det bør jobbes aktivt med å finne riktige kontaktpersoner, og involvere dem i utformingen av evalueringsopplegget.
3. Dette gjelder spesielt i utforming av læringspunktene.
4. I tillegg til å besvare *hvorvidt* prosjektet ble vellykket, må evalueringen også gi svar på *hvorfor* man lyktes eller ikke, og hva man kan lære av det.
5. Man må skille mellom informasjonsbehovet hos de to målgruppene (departement og etat). Fem av evalueringskriteriene retter seg i hovedsak mot departementsnivået, det siste (produktivitetskriteriet) mot etatene. En bør få frem dette skillet bedre enn i dag. Ikke minst bør man sortere læringspunktene etter målgruppe.
6. Mer fokus på formidling er nødvendig. En må innse at det er få i målgruppen som kan ta seg tid til å lese hele evalueringsrapporter. Det tilsier at det er behov for gode og dekkende sammendrag og tilsvarende, som på en lettfattelig måte kan oppsummere resultatene.
7. Det er tilrettelagt for akkurat dette på Concepts nettside, men det må gjøre mer for at målgruppen er kjent med denne kilden til relevant informasjon om prosjekter og læringspunkter som kan være nyttig i planlegging av nye prosjekter.

Tiltak for å lære av erfaringer

Å lære på grunnlag av erfaring betyr at man trekker ut kunnskap som kan være generelt aktuell i flere situasjoner, ikke bare én bestemt situasjon. Evaluering er bare én kilde til læring i en kompleks verden, som illustrert i figur 12. I tillegg vil institusjonens eget minne veie tungt, og ytre stimuli eller kilder som forskning, media og politiske prioriteringer og debatt.

- Lærdom handler ikke om detaljene, men bør knyttes til overordnede vurderinger basert på evalueringens funn konklusjoner og anbefalinger. Prosjektets relevans, ytelse og suksess er eksempler på slike.
- Målgruppen bør ta inn over seg både positive og negative erfaringer, for eksempel beste og verste praksis, og forhold som har betydning utover et enkelt prosjekt eller program.
- Å trekke erfaringer fra flere prosjekter gir åpenbart et bedre grunnlag for generisk læring.

Evalueringer representerer et stort potensial for institusjonell læring. Desto mer dersom det innebærer fortløpende testing og konvertering av dette til generell kunnskap med vekt på det som er relevant for organisasjonen. Videre at det gjøres tilgjengelig for hele organisasjonen. Det er også viktig at det legges vekt på i hvilken grad kunnskapen er anvendbar.

Studier av institusjonell læring viser at de viktigste læringsarenaene er å:

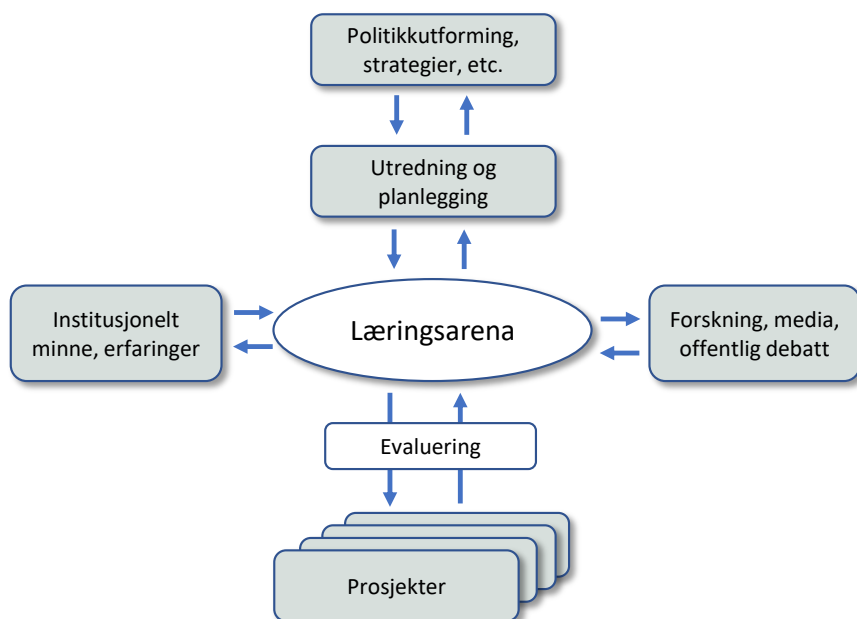
- selv være til stedet der hendelsene skjer,
- delta i interne granskinger,
- trekke på erfaring fra andre organisasjoner,
- diskutere med kolleger.

Evalueringsrapporter representerer først og fremst en omfattende og pålitelig kilde til informasjon som er et viktig supplement til dette.

Dette betyr at evalueringsrapporter kan brukes til mer enn som underlag for beslutning i det konkrete tilfellet. Det er mye å tjene på å bruke evalueringer mer proaktivt, for eksempel ved sørge for at organisasjonens medarbeidere deltar i evalueringsteam for å få innsikt og læring. Dette gjøres i svært liten grad, til tross for at fordelene kan være store for institusjonen.¹⁰ Det kan også være nyttig å stimulere til mer aktiv bruk av

¹⁰ Problemet med at teamet blir oppfattet som mindre uavhengig, er erfaringsmessig ubetydelig.

evalueringsrapporter for lignende prosjekter produserte av andre. Databaser med evalueringsresultater fra mange tiltak, for eksempel evalueringsportalen.no, kan være en nyttig ressurs.



Figur 12 Læringsarenaer som grunnlag for politikkutforming, strategiarbeid, konseptvalgutredninger m.m. i statsforvaltningen

I tillegg kommer muligheten for institusjonen til å gjennomføre eller bestille bredere eller inngående tematiske studier basert på evalueringsresultater. Det kan være studier som ser på tvers av ulike prosjekter, sektorer, institusjoner, geografisk beliggenhet, etc.

Funn fra evalueringer brukes i økende grad for å gi offentligheten korrekt informasjon i kontroversielle saker, eller på et tidlig tidspunkt for å avklare forhold som ellers kan resultere i vanskelige situasjoner.

Uansett middel eller mål, er det å sikre god kvalitet og innhold i evalueringsrapportene den beste måten å sikre at informasjonen blir brukt. Det primære formålet med evaluering er å forbedre kvaliteten av prosjekter og prosesser. I den grad de også fører til større åpenhet og offentlig innsyn kan det være positivt, og anses som en bieffekt

Hva statlige virksomheter kan gjøre

Kunnskapsutvikling i organisasjoner er krevende. Det er individene som gjør erfaringer og gjennom det utvikler sin kunnskap. Organisasjonen er avhengig av disse for at kunnskapsutvikling skal skje på et organisatorisk plan. Det later til at det er mangel på gode systemer for bruk av evaluering internt i virksomhetene, og Concept-studien har en rekke konkrete anbefalinger for å bedre på situasjonen (Bukkestein m.fl., 2020):

1. Ansvar for evaluering bør plasseres hos den som er ansvarlig for måloppnåelsen

Situasjonen er at ansvaret for evaluering fremstår uklart og fragmentert, både mellom berørte parter og internt i de ulike virksomhetene. Realiseringen av resultatmålene er fokus for den gjennomførende part, mens interessen og eierskapet for mer overordnede varierer. Anbefalingen er at det bør være en prosjekteier på departementsnivå som har et tydelig ansvar for prosjektets suksess basert på klare retningslinjer.

2. Tidlig tilrettelegging for evaluering, vil øke kvaliteten på evalueringene

Evaluering skjer i etterkant, og fremstår i stor grad som en isolert aktivitet. Man bør allerede i tidligfasen, for eksempel som del av KS1 og KS2, tilrettelegge for evaluering. Da vil man måtte klargjøre programteorien, og tilrettelegge for å fremskaffe og lagre nødvendig informasjon underveis.

3. Evaluering bør inngå i departementenes virksomhetsstyring

Økonomiregelverket krever at alle statlige virksomheter sørger for at det gjennomføres evalueringer, men tar ikke stilling til hvordan dette kan formaliseres. Det er behov for en særskilt instruks for evalueringsvirksomheten. Den ansvarlige virksomheten bør utarbeide en oppfølgingsplan basert på evalueringsrapporten, og ansvarlig departement bør følge opp på grunnlag av denne.

4. Å bruke evalueringer der de allerede finnes

Virksomheter bør vurdere bruk av referanseprosjekter ved planleggingen av nye prosjekter. Både med tanke på tids- og kostnadsmål, og bruker- og samfunnseffekter, utilsiktede virkninger, og samfunnsøkonomisk lønnsomhet. Det kan for eksempel stilles krav til å bruke funn fra relevante etterevalueringer i forbindelse med konseptvalgutredninger.

5. Utarbeide standardiserte sluttrapporter innen bestemte frister

Studien fant at det gjennomgående er et etterslep for utarbeidelse av sluttrapporter fra prosjektene, og at innholdet og kvaliteten varierer. Dette har også blitt påpekt i flere av Concepts evalueringer. Å ha tilgang til gode erfaringstall, som er

systematisert på en slik måte at de er tilgjengelige for hele organisasjonen, er nødvendig for blant annet å utarbeide gode kostnadsestimater i ulike prosjektfaser.

6. Systemer for å spre relevant kunnskap internt i organisasjonen

Selv om en evaluering avdekker viktige funn, er det ingen automatikk i at dette kommer den aktuelle virksomheten til nytte. Enkeltpersoner med et begrenset ansvar er neppe de rette for å spre relevant kunnskap. Her har ledelsen i virksomhetene, men også Concept, et ansvar for bedre formidling om resultatet av etterevalueringer. Men både etatene og departementene må selv ta ansvar for å tilgjengeliggjøre relevant informasjon og benytte seg av denne.

Referanser

Bukkestein I, Welde M, Volden GH, 2020, *Bruk og nytte av etterevalueringer av prosjekter*, Arbeidsrapport 2020-3, Concept-programmet, NTNU
<https://www.ntnu.no/documents/1261860271/1262021752/Bruk+og+nytte+av+ett+erevaluering+endelig+versjon.pdf/9b8f2505-efba-a7fd-3785-a8df94253ff2?t=1589629656518>

Concept-programmet, 2020, *Ettrevaluering av statlige investeringsprosjekter. Retningslinjer for evaluator*. Versjon 4,
<https://www.ntnu.no/documents/1261860271/1263838555/Template++for+ett+erevaluering++version+4+mai+2020.pdf/3528f930-db4e-5a26-3c48-92ac3c4bf6f8?t=1592299741955>

Direktoratet for økonomistyring, 2018, *Veileder i samfunnsøkonomisk analyse*,
<https://dfo.no/filer/Fagomr%C3%A5der/Utreddinger/Veileder-i-samfunnsokonomiske-analyser.pdf>

Direktoratet for økonomistyring, 2019, *Veiledere til evaluering*,
<https://dfo.no/fagomrader/evaluering/veiledere-til-evaluering#DFOs-veiledere:10174>

Finansdepartementet, 2005, *Veileder til gjennomføring av evalueringer*,
<https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/veileder+til+gjennomforing+av+evalueringer.pdf>

Finansdepartementet, 2010, *Konseptvalg og detaljeringsgrad*, Veileder nr. 11.
<https://www.ntnu.no/documents/1261860271/1263838555/Veileder%20nr%2011%20Konseptvalg%20og%20detaljeringsgrad.pdf>

Lervåg L-E, 2017, *Teoribasert evaluering av intelligente transportsystemer*, rapport, SINTEF Teknologi og samfunn
<https://www.ntnu.no/documents/1261860271/1262021752/2017+00248+Teoribasert+evaluering+av+intelligente+transportsystemer.pdf/9fc59064-e304-42b9-90f7-69a22fafb757?version=1.0>

Samset K, 2014, *Evaluering av prosjekter: Vurdering av suksess*, Fagbokforlaget, Bergen, <https://www.fagbokforlaget.no/Evaluering-av-prosjekter/I9788245017489>

Tornes K, 2013, *Evaluering i teori og praksis*, Fagbokforlaget,
<https://www.fagbokforlaget.no/Evaluering-i-teori-og-praksis/I9788232100040>

Ulstein H, Gulbrandsen M, Wifstad K og Holmen R, 2014, Etterevaluering av Rv 653 Eiksundsambandet. Menon-publikasjon nr. 4/2014
https://www.ntnu.no/documents/1261860271/1262012574/221_Evaluering%20Eiksundsambandet%20final%2031012014.pdf

Volden G H, og Samset K, 2017, *Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene*, Concept-rapport nr. 52, Ex ante akademisk forlag,
https://www.ntnu.no/documents/1261860271/1262010703/CONCEPT_52_etterjustert_norsk_web.pdf/4ef69cf3-e34a-4761-a244-06a833998d0c?version=1.0

Volden G H, og Samset K, 2013, *Etterevaluering av statlige investeringsprosjekter Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter*, Concept-rapport nr. 30, Ex ante akademisk forlag,
https://www.ntnu.no/documents/1261860271/1262010703/Concept_rapport_nr_30.pdf

Temahefter fra Concept-programmet

Nummer	Tittel	Forfatter	Utgitt
1	Fleksibilitet i prosjekter – et tveegget sverd	Nils Olsson	2009
2	På sporet av relevans og levedyktighet	Ole Jonny Klakegg	2010
3	Valg av kontraktstrategi i bygg- og anleggsprosjekt	Ola Lædre	2012
4	Kostnadsestimering under usikkerhet	Frode Drevland	2013
5	I riktig retning. Prosjekters mål og målstruktur	Knut Samset	2014
6	Prosess for kostnadsestimering under usikkerhet	Olav Torp, Frode Drevland, Kjell Austeng	2015
7	Mulighetsrommet. Utgangspunktet for et godt konseptvalg	Knut Samset	2016
8	Økonomisk verdsetting av liv og helse	Rune Elvik	2017
9	Fremsynsmetoder	Tore Sager, Knut Samset	2018
10	Mandatet. Utgangspunktet for et godt konseptvalg	Knut Samset	2019
11	Problemanalysen. Det første steget i konseptvalg-utredninger	Knut Samset	2020
12	Tidlig involvering av entreprenør	Paulos Abebe Wondimu	2020
13	Prosjektevaluering	Knut Samset	2021

Temahefter og andre publikasjoner fra Concept-programmet kan lastes ned fra programmets nettsider: www.ntnu.no/concept/publikasjoner

Forskningsprogrammet Concept skal utvikle kunnskap som sikrer bedre ressurs-utnyttning og effekt av store statlige investeringer. Programmet driver følgeforskning knyttet til de største statlige investeringsprosjektene over en rekke år. En skal trekke erfaringer fra disse som kan bedre utformingen og kvalitetssikringen av nye investeringsprosjekter før de settes i gang.

Concept er lokalisert ved Norges teknisk-naturvitenskapelige universitet i Trondheim (NTNU), ved Institutt for bygg- og miljøteknikk. Programmet samarbeider med ledende norske og internasjonale fagmiljøer og universiteter, og er finansiert av Finansdepartementet.



Concept-programmet
Institutt for bygg og miljøteknikk, NTNU, Høgskoleringen 7A, 7491 TRONDHEIM

Informasjon om Concept-programmet: www.ntnu.no/concept

ISBN: 978-82-8433-002-0 (papirversjon)
ISBN: 978-82-8433-003-7 (nettversjon)