

Etterevaluering av statlige investeringsprosjekter i regi av Concept-programmet. Retningslinjer for evaluator.

Versjon 2, 08.03.2015

Bakgrunn

Concept-programmet driver følgeforskning knyttet til statlige investeringsprosjekter som er underlagt ordningen med ekstern kvalitetssikring (KS-ordningen¹). De første prosjektene som ble kvalitetssikret tidlig på 2000-tallet er nå ferdigstilt og kommet i driftsfasen, og i tiden fremover vil Concept etterevaluere mange av disse. De fleste evalueringene settes ut til eksterne evaluatorene, mens enkelte gjennomføres av programmets egne forskere.

Det skal benyttes en felles evalueringsmodell i alle evalueringene. Denne er beskrevet i en egen lærebok, Samset (2014a), og erfaringene med å benytte modellen i fire pilotevalueringer i Concept-rapport nr. 30 (Volden og Samset, 2013).

Dette notatet gir en del anbefalinger og presiseringer basert på erfaringene så langt.

Anbefalingene vil bli løpende oppdatert etter hvert som en oppnår mer erfaring.

Evalueringsmodellen

Det benyttes en målorientert evalueringsmodell i kombinasjon med en samfunnsøkonomisk analyse, som til sammen sikrer at prosjektet blir vurdert på tre ulike nivåer:

- Operasjonell vellykkethet. Om leveransen foreligger som avtalt, og er produsert på tids- og kostnadseffektiv måte.
- Taktisk vellykkethet. Om tiltaket gir den forventete nytten, og innebærer en effektiv måte å nå målet på.
- Strategisk vellykkethet. Om tiltaket samlet sett bidrar til en samfunnsmessig ønsket utvikling over tid, og innebærer en effektiv ressursallokering.

Den målorienterte modellen benytter fem evalueringskriterier, *produktivitet*, *måloppnåelse*, *relevans*, *virkinger* og *levedyktighet*. Prosjektets *produktivitet* er et mål på operasjonell vellykkethet. *Måloppnåelse* er et mål for taktisk vellykkethet, og de tre siste for strategisk vellykkethet.

Den samfunnsøkonomiske analysen handler om å vurdere samfunnsøkonomisk nytte opp mot ressursinnsatsen, det vil si *samfunnsøkonomisk lønnsomhet* som også er et kriterium på strategisk nivå. Dersom det ikke er mulig eller faglig forsvarlig å vurdere nyttesiden i samfunnsøkonomisk forstand, kan alternativet være å vurdere prosjektets *kostnadseffektivitet i forhold til sammenliknbare prosjekter*. Men det blir i så fall en vurdering på taktisk nivå. Ikke-prissatt nytte må i alle tilfeller beskrives og inngå i analysen.

¹ For mer informasjon om ordningen med ekstern kvalitetssikring av store statlige investeringsprosjekter, se www.ntnu.no/web/concept/ks-ordningen1

Generelle føringer

En skiller mellom det konkrete og avgrensede *investeringsprosjektet*, og det mer overordnede og langsiktige *investeringstiltaket*. Prosjektet vurderes i et operasjonelt perspektiv slik det er beskrevet og avgrenset på det tidspunkt det får sin bevilgning av Stortinget. I et taktisk og strategisk perspektiv er det *investeringstiltaket* som skal vurderes. Det betyr at en må vurdere prosjektet i et bredt samfunnsperspektiv, og kanskje også i forhold til andre prosjekter og/eller virkemidler.

Evaluators skal starte med å fortolke de avtalte målene og prosjektlogikken, og sikre at de er konsistente, på et rimelig ambisjonsnivå og egnet som referanseverdier i en evaluering (se temahefte, Samset (2014b), samt Concept-rapport nr. 21). Dersom det er gjennomført KS1 vil kvalitetssikrers vurdering av målene (ref. strategikapittel i KVU/KS1) også være et viktig utgangspunkt. Eventuelle endringer av målene eller målstrukturen må skje i samråd med departement/etat og Concept.

Evaluators skal selv operasjonalisere de seks kriteriene, det vil si å velge ut mer konkrete forhold som skal undersøkes for det enkelte prosjektet, såkalte indikatorer. Deretter må relevante data samles inn og analyseres.

I en overordnet evaluering med begrensede ressurser må en i mange tilfeller avgrense undersøkelsene til hva som er ”godt nok”, men likevel sette tilstrekkelig høye kvalitetskrav. En må redegjøre for metodevalg og begrensninger, dette er drøftet i Concept-rapport nr. 17 (Sunnevåg m.fl., 2007). Triangulering er viktig som et virkemiddel for å sikre informasjonens validitet, ved at en bruker flere informasjonskilder, respondenter, typer analyse etc.

Ved avslutningen av evalueringen skal hvert av de seks overordnede evalueringskriteriene gis en poengscore, som angir evalueringsteamets vurdering av investeringstiltakets vellykkethet. Det benyttes en skala fra 1 til 6, der 1 er laveste score (helt mislykket), og 6 er høyeste score (svært vellykket). Score 4 brukes som der resultatet er «helt greit», med andre ord akseptabelt, men ikke noe mer.

Grunnlaget for vurderingen vil være både kvantitativ og kvalitativ informasjon av ulik art. Samlet sett gir dette ikke grunnlag for systematisk gradering av vellykkethet på ordinalnivå, men en vil være henvist til i stor grad å basere poenggivning på skjønnsmessige vurderinger.

For at det skal være en viss stringens i poengfastsettelsen mellom ulike evalueringer, som gjennomføres av forskjellige evalueringsteam, har vi satt opp noen generelle retningslinjer nedenfor.

1. Produktivitet

Dette gjelder leveransen av selve investeringsprosjektet, dvs. realiseringen av resultatmålene og spørsmålet om hvor effektivt ressursene er omsatt til leveranser. Dette har tre dimensjoner: tid, kostnad og kvalitet.

Utgangspunktet for vurderingen er Stortingets vedtak og KS2-rapporten med de anbefalingene som der ble gitt, om blant annet:

- Styrings- og kostnadsramme
- Kontraktstrategi
- Usikkerhet og risikoreduserende tiltak
- Organisering og styring av prosjektet

Score 5-6: forbeholdes prosjekter som leverer innenfor styringsramme, tidsplan og med meget god kvalitet, og som i tillegg kommer godt ut på en referansesjekk.

Score 4 brukes når prosjektet leverer på eller innenfor kostnadsrammen, ikke har større forsinkelser samt med akseptabel kvalitet, på nivå med sammenlignbare prosjekter.

Score 1-3 gis dersom prosjektet overskrider kostnadsrammen og/eller for øvrig leverer på et uakseptabelt nivå i forhold til sammenlignbare prosjekter.

Evaluators skal vurdere i hvilken grad anbefalingene i KS2-rapporten er fulgt, og hva som ble resultatet. Dersom det oppstod utfordringer som ikke var drøftet i KS2-rapporten må dette vurderes særskilt.

Utgangspunktet for fastsettelse av score er resultatmålene som var satt i forkant. Imidlertid kan både priser, teknologi og andre forhold ha endret seg i ettertid. Derfor bør en i tillegg gjøre en referansesjekk mot resultater for andre, liknende prosjekter gjennomført i samme periode.

2. Måloppnåelse

Dette handler om i hvilken grad effektmålet, dvs. førsteordens effekter av prosjektet, realiseres. I vurderingen av måloppnåelse skal en bare legge til grunn mål som er realistisk realiserbare. I prosjekter der formelt avtalte mål er urealistiske eller trivielle, må evalueringen eksplisitt angi hvilke effektmål som legges til grunn og gi begrunnelsen for dette.

I tilfeller hvor evaluators justering av målene har direkte utslag på hvilken score som gis, må det gjøres oppmerksom på dette. I slike tilfeller er det særlig viktig å begrunne, og sikre legitimitet for, de endringene som er gjort.

Score 5-6 gis ved svært god / overoppfyllelse av (de justerte) effektmålene, og der prosjektet fremstår som et treffsikkert virkemiddel for å realisere effektene.

Score 4 gis ved resultater som anses som akseptable, men ikke noe mer.

Score 1-3 gis dersom effektene uteblir eller er klart lavere enn det som er akseptabelt.

3. Virkninger

Vurderingen her gjelder alle andre virkninger utover måloppnåelsen som kan tilbakeføres som resultat av prosjektet. Det gjelder både forventete og ikke-forventete positive og negative konsekvenser, kort- og langsiktige, for målgruppen og for andre berørte parter. Dette er ofte en krevende del av evalueringen. Dersom det er gjennomført KS1 vil alternativanalysen være en viktig referanse (Ref. KVVU og KS1-rapport).

En kan ikke uten videre legge til grunn at negative og positive virkninger oppveier hverandre. En bør være særlig oppmerksom på negative virkninger, og spesielt hvis det er sårbare grupper som rammes.

Score 5-6 anvendes der tiltaket har betydelige positive virkninger (utover måloppnåelsen) og ingen eller marginale negative virkninger.

Score 4 bør en tenke på som «nullpunkt», det vil si at tiltaket ikke har noen andre vesentlige virkninger ut over det som gjelder måloppnåelsen.

Score 1-3 gis dersom det er overvekt av negative virkninger, og laveste score gis dersom de negative virkningene både i omfang og effekt er vesentlig større enn de positive virkningene av tiltaket.

4. Relevans

At formelt avtalte mål blir realisert er ikke en tilstrekkelig betingelse for at et tiltak er vellykket. Både effekt- og samfunns mål må i tillegg være i samsvar med brukernes og samfunnets behov og prioriteringer. I noen tilfeller ser en at både prioriteringene og behovene da tiltaket ble foreslått kan være annerledes når tiltaket er gjennomført, særlig dersom dette skjer mange år senere. Vurderingen av relevans vil kreve en særskilt vurdering i etterkant av behov, prioriteringer og alternative konsepter/tiltak, der en både drøfter om sentrale behov er dekket og om det valgte tiltaket er det som gir best behovsdekning.

Score 5-6 gis dersom effekt- og samfunns mål er i samsvar med sentrale og høyt prioriterte behov, og tiltaket ikke fører til skjevprioriteringer eller konflikter mellom sentrale interessegrupper som berøres av tiltaket.

Score 4 gis ved akseptabelt samsvar med behov og prioriteringer.

Score 1-3 gis dersom tiltaket ikke er i samsvar med sentrale prioriteringer i samfunnet, og/eller fører til konflikter eller vesentlig skjevfordeling mellom sentrale interessenter.

Behov kan beskrives ut fra tre prinsipielt ulike perspektiv, i) normativt/politisk, ii) etterspørsel/marked og iii) interessegruppers behov, se Concept-rapport nr. 5 (Næss, 2005), dette gjelder også ex post. Dersom det er konflikter mellom de tre, er det viktig å avdekke dette.

Referansepunktet for vurderingen må være de behov som gjelder på evalueringstidspunktet. Det er samtidig interessant å spørre om tiltaket var treffsikkert i forhold til gjeldende behov den gang det ble utformet, og om en har tatt hensyn til eventuelle endringer i behov over tid. For prosjekter som har vært gjennom KS1, vil kvalitetssikrer ha vurdert relevansen i forkant, som del av sin analyse av behov, mål, krav og mulighetsrom (ref. KVVU og KS1 rapport).

Dersom ulike behov er i strid med hverandre, kan tiltaket være relevant i forhold til ett behov, men ikke i forhold til et annet. I slike tilfeller må evaluator forsøke å vekte behovene ut fra beste faglige skjønn, men allikevel slik at Stortingsvedtak rangerer høyest.

5. Levedyktighet

Vurderingen av tiltakets levedyktighet gjelder det langsiktige, strategiske perspektivet. En spør da om de positive effektene av tiltaket kan vedvare over hele levetiden. Dette har både en økonomisk, miljømessig og sosial/fordelingsmessig dimensjon og avhenger blant annet av fremtidige behov og prioriteringer, ressurstilgang, finansieringsevne og -vilje, og fleksibilitet til å tilpasse seg nye rammebetingelser etc. (se Concept-rapport nr. 29, Lædre m.fl., 2012).

Rammen for vurderingen er tiltakets antatte levetid, eller perioden det kan forventes å gi positive virkninger for brukere og samfunn. Vurderingen må derfor baseres på antakelser om den videre utviklingen, og således være usikker.

Score 5-6 gis der en finner at det offentlige og sentrale interessenter både har evne og vilje til å videreføre de prosessene som prosjektet har gitt opphav til over hele levetiden.

Score 4 gis dersom dette er usikkert

Score 1-3 gis dersom det ikke er tilfellet og at sannsynligheten for at det vil skje i fremtiden er lav.

Tiltaket bør ha fleksibilitet til å tilpasses endrede behov over tid. Manglende fleksibilitet bør, alt annet likt, gi en lavere score.

6. Samfunnsøkonomisk analyse

Samfunnsøkonomisk analyse er en metode for å vurdere nytten av et tiltak i forhold til ressursbruken. Grunnlaget for dagens norske praksis finnes i Finansdepartementet (2014) og DFØ (2014). Det skilles mellom tre analysetyper, avhengig av muligheten til å verdsette nytten:

- En nytte-kostnadsanalyse måler de sentrale virkningene i penger, og gjør det mulig å svare på om prosjektet er lønnsomt.
- En kostnadsvirkningsanalyse, er på samme måte som en nytte-kostnadsanalyse er vurdering av både nytte og kostnad, men her verdsetter man kostnadene og beskriver nytten kvalitativt, og spørsmålet om lønnsomhet må vurderes skjønnsmessig til slutt.
- En kostnads-effektivitets-analyse tar derimot nyttesiden for gitt og analyserer kostnadene, og identifisere det mest

Samfunnsøkonomisk analyse:

Score 5-6 forbeholdes lønnsomme prosjekter ($NNV > 0$). Score 4 gis når resultatet er omtrent null, og 1-3 gis ved ulønnsomme prosjekter ($NNV < 0$). Også ved kostnads-virkningsanalysen bør dette være et utgangspunkt.

En må i tillegg se hen til resultater fra andre tilsvarende prosjekter. Dersom de fleste prosjekter av en viss type er ulønnsomme, f.eks. fordi det er grunn til å tro at nyttesiden undervurderes, kan resultatkravet reduseres noe, og vice versa.

Kostnadseffektivitetsanalyse:

Score 5-6 forbeholdes prosjekter som er mer effektive enn gjennomsnittet for sektoren, 4 gis ved resultater på akseptabelt nivå eller på gjennomsnittet, og 1-3 ved resultater klart under det normale og akseptable.

kostnadseffektive tiltaket for å nå et gitt mål.

I en overordnet evaluering med begrensede ressurser, og gitt erfaringer som tilsier at det er svært krevende å kvantifisere og verdsette alle virkninger ex post, bør en ikke legge for mye ressurser i dette. Det kan i mange tilfeller være bedre med en kvalifisert vurdering av *sannsynligheten* for at tiltaket er lønnsomt, enn en detaljert beregning som ikke er komplett.

I de tilfellene en samfunnsøkonomisk analyse var gjennomført i forkant av prosjektet skal en vurdere om resultatene avviker fra det som var lagt til grunn. Men det er faktisk lønnsomhet ex post som skal ligge til grunn for fastsettelse av score.

Årsaksforklaringer og læringspunkter

Et sentralt formål med evalueringene er å samle lærdom som kan komme til nytte i senere prosjekter. I tillegg til å vurdere prosjektets og investeringstiltakets vellykkethet må evaluator derfor også beskrive viktige årsaker til et godt eller dårlig resultat. Stikkord kan være: behovsvurdering i tidligfasen, målformulering, prognoser og metoder, kompetanse, konsepter som ble vurdert/ikke vurdert, involvering/ikke involvering av bestemte aktører, valg av gjennomføringsmodell, styring, ansvar for gevinstrealisering, etc.

En skal også påpeke hvorvidt anbefalingene fra KS-rapporten ble fulgt, og i hvilken grad kvalitetssikringen kan antas å ha betydning for eller påvirket resultatet.

Krav til evaluator

Evalueringsteamet bør bestå av minst tre personer som til sammen har god kompetanse og erfaring både på målorientert evaluering og samfunnsøkonomisk analyse. For øvrig bør teamet være tverrfaglig sammensatt, med både prosjektfaglig kompetanse og sektorkompetanse representert. Sektorkompetansen bør ikke dominere. Teamleder bør være den med mest kompetanse og erfaring fra evaluering.

Evaluators habilitet og uavhengighet i forhold til berørte etater og departementer er alltid viktig.

Ved evaluering av forsvarsprosjekter forutsetter vi at evaluator selv sørger for sikkerhetsklarering av personellet som skal jobbe med evalueringen.

Concept-programmet forbeholder seg retten til å sette inn en egen følgeforsker i evalueringsteamet. Vedkommende vil bidra i evalueringen på lik linje med øvrige deltakere, i tillegg til å samle data og erfaring som del av Concept-programmets følgeforskningsarbeid. Forskerens arbeidsinnsats belastes i slike tilfeller ikke av evalueringsoppdragets budsjett.

Sentrale referanser om evaluering

Volden, G. og K. Samset (2013). *Etterevaluering av statlige investeringsprosjekter. Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter*. Concept-rapport nr. 30

Samset, K. (2014a). *Evaluering av prosjekter. Vurdering av suksess*. Fagbokforlaget

Finansdepartementet (2014). *Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv*. Rundskriv R 109/14

Nettside med resultater fra tidligere evalueringer: www.ntnu.no/concept/etterevaluering-av-prosjekter

Andre referanser

Lædre, O. m.fl. (2012): *Levedyktighet og investeringstiltak*, Concept-rapport nr. 29

Direktoratet for økonomistyring (2014). *Veileder i samfunnsøkonomiske analyser*

Næss, P. (2005). *Bedre behovsanalyser. Erfaringer og anbefalinger om behovsanalyser i store offentlige investeringsprosjekt*, Concept-rapport nr. 5

Samset, K. (2014b). *I riktig retning. Prosjekters mål og målstruktur*. Concept temahefte nr. 5

Samset, K. m.fl. (2009). *Sykebus, fregatter og skipstunnel: Logisk minimalisme, rasjonalitet – og de avgjørende valg*, Concept-rapport nr. 21

Sunnevåg, K.J., red. (2007). *Beslutninger på svakt informasjonsgrunnlag. Tilnærminger og utfordringer i prosjekters tidlige fase*, Concept rapport nr. 17