

Etterevaluering av statlige investerings- prosjekter

Retningslinjer for evaluator

Versjon 4, mai 2020



Innhold

Bakgrunn.....	1
Evalueringsteamet	1
Evalueringsmodellen.....	2
Bakgrunnsdokumentasjon og grunnlaget for evalueringen.....	3
Evalueringsprosessen.....	4
Kvalitet og evalueringsscore	5
1. Produktivitet.....	6
2. Måloppnåelse	8
3. Andre virkninger.....	9
4. Relevans	10
5. Levedyktighet.....	11
6. Samfunnsøkonomisk analyse.....	12
Årsaksforklaringer og læringspunkter.....	13
Kostnadsestimater og -tall som skal registreres	14
Sentrale referanser om evaluering.....	16
Andre referanser	16



Bakgrunn

Concept-programmet driver følgeforskning knyttet prosjekter som er underlagt statens ordning med ekstern kvalitetssikring (Statens prosjektmodell), se [1]. Mange av disse etter-evalueres, de fleste av eksterne fagmiljøer på oppdrag

Evalueringene skal benytte samme evalueringsmodell, se [2] og [3]. Dette notatet gir retningslinjer vedrørende metodikken og bruken av denne, minstekrav til data som skal samles inn, samt scoringssystemet som skal sikre at en kan sammenlikne resultater på tvers av prosjekter og sektorer.



Evalueringsteamet

Teamet bør ha minst tre personer, med relevant kompetanse og erfaring, både evalueringsfaglig, prosjektfaglig (også estimeringskompetanse), samfunnsøkonomisk og sektorfaglig, og bør ledes av den med mest evalueringserfaring.

Habilitet og uavhengighet i forhold til berørte etater og departementer må avklares før kontraktsinngåelse. Evaluatoren må selv sørge for sikkerhetsklarering av personellet ved evaluering av forsvarsprosjekter..

Concept-programmet kan sette inn egen følgeforsker i evalueringsteamet etter avtale. Vedkommende skal bidra i evalueringen på lik linje med de øvrige. Forskerens arbeidsinnsats dekkes i slike tilfeller av Concept, og ikke av evalueringsoppdragets budsjett.



Evalueringsmodellen

Alle evalueringer benytter samme målorienterte evalueringsmodell, som sikrer at prosjektet blir vurdert både på operasjonelt nivå¹, og taktisk/strategisk².

Modellen benytter fem evalueringskriterier, som er beskrevet på side 6 - 12.

1. Produktivitet
2. Måloppnåelse
3. Relevans
4. Andre virkninger
5. Levedyktighet.

Prosjektets *produktivitet* er et mål på operasjonell vellykkethet, *måloppnåelse* for taktisk vellykkethet, og de tre siste for strategisk vellykkethet.

I tillegg gjennomføres en samfunnsøkonomisk analyse, som vurderer samfunnsøkonomisk nytte opp mot ressursinnsatsen.

Det kan om mulig gjøres kvantitativt, det vil si å *beregne* samfunnsøkonomisk lønnsomhet, eventuelt kvalitativt og skjønsmessig basert på en kostnads-virkningsanalyse.

Ikke-prissatt nytte må i alle tilfeller beskrives og inngå i analysen. 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*.

¹ det vil si hvorvidt leveransen foreligger som avtalt, og er produsert på tids- og kostnadseffektiv måte

² det vil si om tiltaket gir den forventete samfunnsmessige nytten, ønsket utvikling over tid, og innebærer en effektiv ressursallokering



Bakgrunnsdokumentasjon og grunnlaget for evalueringen

Evalueringsobjektet er *prosjektet* slik det var beskrevet ved Stortingets bevilgning.

Evaluators skal få frem relevant bakgrunnsdokumentasjon, herunder dokumentere:

- Utviklingen gjennom tidligfasen, dvs. konseptets innhold og avgrensning, og kostnadsestimater, både uformelle og formelle anslag.
- Tidspunktet for når konseptet reelt sett ble valgt³, herunder hvilket kostnadsestimat beslutningen var basert på.
- Analyser og vurderinger som forelå som grunnlag for beslutning, av relevans for hvert av de seks evalueringskriteriene
- Hvorvidt analysene var basert på relevant erfaringsmateriale, som tidligere evalueringer, referanseprosjekter e.l.

Evaluators skal fortolke og vurdere prosjektets avtalte mål og prosjektlogikken, eventuelt justere disse. En må sikre at de er konsistente og realistisk realiserbare⁴, og egnet som referanseverdier i en evaluering, se temahefte og rapport [9] og [10]. Begrunnelsen for endringene må fremgå i rapporten.

Evaluators skal operasjonalisere de seks evalueringskriteriene, i mer konkrete forhold som skal undersøkes, såkalte indikatorer. Omfanget skal være tilstrekkelig til å dokumentere alle sentrale forhold som skal belyses.

³ Det gjelder også prosjekter som ikke har gjennomgått KS1

⁴ typisk opp mot hva som er oppnådd i andre, sammenlignbare prosjekter



Evalueringsprosessen

Prosessen er trinnvis med følgende hovedaktiviteter:

1. Concept velger hvilke prosjekter som skal evalueres. Det ansvarlige fagdepartementet kan i prinsippet si nei. Vi ber om en kontaktperson i underliggende etat som har hatt ansvaret for prosjektet
2. Evaluator velges etter offentlig utlysning, eller internt. Omfanget har vært i størrelsesorden 3 månedsverk, og tidsrammen omtrent 6 mnd.
3. Ansvarlig departement/etat, involveres tidlig for å få innspill til hva som er spesielt interessant å få belyst, fortrinnsvis gjennom et oppstartsmøte.
4. Evaluator legger frem en oversikt over indikatorer og tilhørende plan for datainnsamling og analyse for eventuelle kommentarer og faglige innspill. Concept kan ikke overstyre evaluators opplegg.
5. Evalueringen gjennomføres. Concept kan bistå med tilgang til sentrale dokumenter som finnes i programmets database.
6. Utkast til evalueringsrapport sendes både Concept og kontaktpersonene i fagdepartement og etat for eventuelle kommentarer. Normalt bør det holdes et møte med alle parter. Det er særlig viktig å fokusere på læringspunktene, som er det siste, viktig kapitlet i evalueringsrapporten..
7. Teamet slutfører og gjør en samlet vurdering av prosjektets vellykkethet for hvert av de seks overordnede kriteriene, som beskrevet nedenfor.
8. Rapporten publiseres på nettsiden til Concept hvor den også kan lastes ned. Det er spesielt viktig at rapporten har et godt og dekkende sammendrag.

Concept og evaluator avklarer i felleskap hvordan resultatene kan formidles. Departement og etat oppfordres til å følge opp resultatene internt. Concept og evaluator er tilgjengelig for å gi en presentasjon eller diskutere resultatene med departement eller etat hvis ønskelig.



Kvalitet og evalueringsscore

I en evaluering med begrensede ressurser må en gjøre en avveining mellom hva som er ”godt nok”, og samtidig holder tilstrekkelig kvalitet. En må redegjøre for metodevalg og begrensninger, dette er drøftet i [11].

Triangulering er et viktig virkemiddel for å sikre informasjonens validitet, det vil si å bruke flere informasjonskilder, respondenter, typer analyse etc.

Evaluators egne observasjoner ved befarings er påkrevet i alle evalueringer. Det gir et helt nødvendig tilskudd til kontekstforståelsen.

Ved avslutningen av evalueringen skal hvert av de seks overordnede evalueringskriteriene gis en poengscore, som angir teamets vurdering av investeringstiltaket vellykkethet.

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

En bør søke å sammenlikne vurderingene mot andre, tilsvarende prosjekter så langt det er mulig. Grunnlaget for vurderingen vil være både kvantitativ og kvalitativ informasjon. Samlet sett gir dette ikke grunnlag for systematisk gradering av vellykkethet på ordinalnivå. En vil derfor være henvist til i stor grad å basere poenggivningen på skjønsmessige vurderinger.

Nedenfor følger retningslinjer for hvert av evalueringskriteriene, som skal sikre en viss stringens i poengfastsettelsen mellom evalueringer som gjennomføres av forskjellige evalueringsteam.



1. Produktivitet

Dette gjelder leveransen av selve investeringsprosjektet. Vurderingen er hvorvidt resultatmålene er realisert, og hvor effektivt ressursene er omsatt til leveranser. Dette har tre dimensjoner: tid, kostnad og kvalitet.

En må dokumentere hvordan prosjektet er gjennomført og hva resultatene er. Herunder skal evaluator vurdere i hvilken grad anbefalingene i KS2-rapporten er fulgt, med tanke på f.eks. organisering, styring og kontraktstrategi, og hvilke risikofaktorer som ble utløst. Dersom det oppstod utfordringer som ikke var drøftet i KS2-rapporten må dette vurderes særskilt.

Vurderingene av kostnader skal gjøres i forhold til Stortingets opprinnelige vedtak og KS2-rapportens anbefalinger om blant annet styrings- og kostnadsramme, kontraktstrategi, usikkerhet og risikoreduserende tiltak og organisering og styring av prosjektet

I noen tilfeller må Stortinget godkjenne høyere rammer etter første vedtak.

Begrunnelsen for dette skal omtales, men det er de opprinnelige rammene som skal legges til grunn. Valg av indeks for å regne om

Score 5-6

prosjekt har levert innenfor styringsramme, tidsplan og med meget god kvalitet, og kommer i tillegg kommer godt ut på en referansesjekk.

Score 4-3

prosjektet har levert innenfor kostnadsrammen, har ikke større forsinkelser, og med akseptabel kvalitet: på nivå med andre sammenlignbare prosjekter.

Score 1-2

prosjektet har betydelig overskridelse av kostnadsrammen (i størrelsesorden 20% eller mer), og leverer på et uakseptabelt nivå tidsmessig og kvalitetsmessig i forhold til sammenliknbare prosjekter

opprinnelige rammer til dagens verdi må begrunnes, og det skal i så fall gjøres en følsomhetsanalyse med konsumprisindeksen. Evaluatør skal vurdere om kostnadsrammene var realistiske.

Kostnadsutviklingen i tidligfasen skal dokumenteres. Dette er viktig som indikasjon på om den valgte løsningen er kostnadseffektiv, om tiltaket er relevant i forhold til samfunnets behov, og samfunnsøkonomiske effektivitet, som omtales nedenfor.

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 faktiske resultater for liknende prosjekter gjennomført i samme periode.

NB! En oversikt over kostnadsestimater, budsjettall og kostnadstall som skal dokumenteres er tatt med på side 14-15.



2. Måloppnåelse

Dette handler om i hvilken grad effektmålet, det vi si førsteordens effekter av prosjektet, realiseres. Erfaringsmessig vil en i noen tilfeller støte på målstrukturer som inneholder formelt avtalte mål som enten ikke er relevante eller er urealistisk ambisiøse i forhold til hvilke effekter som prosjektet med rimelighet kan gi. I vurderingen av måloppnåelse skal en bare legge til grunn mål som er realistisk realiserbare. Som nevnt over kan det være nødvendig å justere målene og målstrukturen for å sikre dette. Det må i så fall beskrives og begrunnes i rapporten.

Det er viktig å få frem i hvilken grad prosjektet har bidratt til effektene, og om prosjektet har vært optimalt utformet med tanke på å ta ut effektene eller ikke. Spørsmålet om hvorvidt det hadde vært mulig å oppnå større effekter enn det man registrerte, bør også vurderes i evalueringene.

Utgangspunktet ved fastsettelse av score er hvorvidt (de justerte) effektmålene ble nådd, og i hvilken grad prosjektet har bidratt til dette.

Score 5-6

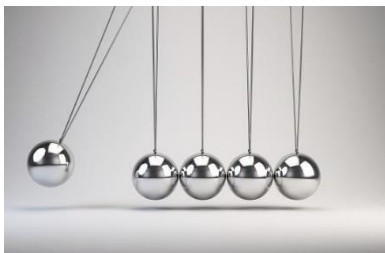
prosjekter med svært god / overoppfyllelse av målene, og hvor prosjektet fremstår som et treffsikkert virkemiddel for å realisere de planlagte effektene.

Score 3-4

prosjekter med effekt og måloppnåelse som anses som akseptable, men ikke noe mer.

Score 1-2

gis dersom effektene uteblir eller er klart lavere enn hva som er forventet.



3. Andre virkninger

Dette 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. I den grad det er mulig, også utslag på andre mål i samfunnet, men som ikke er tatt inn i målhierarkiet for prosjektet. I prosjekter hvor det er gjennomført KS1 vil alternativanalysen være en viktig referanse for evaluator.

Det presiseres at «andre virkninger» ikke er synonymt med samfunnsøkonomiske virkninger, og de skal ikke prissettes i kroner under dette kriteriet.

Evaluator bør være forsiktig med å anslå annen- og tredjeordens ringvirkninger og omtale disse som virkninger av prosjektet. Det gjelder blant annet såkalt mernytte av transportinfrastruktur. En bør i så fall påpeke usikkerheter og innvirkning fra andre faktorer.

En bør være særlig oppmerksom på eventuelle negative virkninger dersom det er sårbare grupper som berøres.

Score 5-6

prosjektet har betydelige andre positive virkninger (utover måloppnåelsen) og ingen eller bare marginale negative virkninger.

Score 3-4

prosjektet har få virkninger ut over det som gjelder måloppnåelsen, og få eller ingen vesentlige negative virkninger.

Score 1-2

prosjektet har vesentlige negative virkninger. 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 kan disse ha endret seg, etter at tiltaket er gjennomført..

Politiske/samfunnsmessige prioriteringer og behov er overordnet i forhold til markedsmessige (etterspørsel), og interessegruppers behov, se [8]. Interessekonflikter mellom aktører skal omtales. En skal skille mellom behov og ønsker, og vurdere bestandighet av behov, for eksempel opp mot situasjonen andre steder i landet.

I analysen skal en vurdere relevans opp mot nullalternativet og hvorvidt en annen løsning på behovet kunne vært *mer relevant* enn den som ble realisert. En skal legge til grunn at de færreste behov er så viktige at de må tilfredsstilles for enhver pris.

I tilfeller hvor lokale brukergrupper selv betaler en del av kostnaden, vil deres betalingsvillighet være en indikator på markedsmessig relevans. I tilfeller hvor bruken er «gratis», gjelder ikke nødvendigvis dette.

Score 5-6

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 3-4

det er akseptabelt samsvar med behov og prioriteringer.

Score 1-3

målene er ikke i samsvar med sentrale prioriteringer i samfunnet, og/eller fører til konflikter eller vesentlig skjevfordeling mellom sentrale interesser.



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 [6].

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

I KS-prosjekter som ikke er brukerfinansiert og Staten finansiere driften over hele levetiden, må en spørre om så vil skje, herunder om de politiske prioriteringene er stabile og om de samlede virkningene av tiltaket er akseptable for staten på lang sikt.

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

Score 5-6

prosjekter 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

prosjekter der dette er usikkert.

Score 1-3

prosjekter der det ikke er tilfellet, og hvor sannsynligheten for at det vil skje i fremtiden er lav.



6. Samfunnsøkonomisk analyse

Samfunnsøkonomisk analyse vurderer nytten av et tiltak i forhold til ressursbruken. Grunnlaget for dagens norske praksis finnes i [4] og [7]. En skiller mellom:

- nytte-kostnadsanalyse, som gir et tallmessig svar på om prosjektet er lønnsomt i kroner.
- kostnads-virkningsanalyse, tallfester kostnadene, men beskriver nytten kvalitativt. Vurderingen blir skjønnsmessig.
- kostnads-effektivitetsanalyse, der en tar nyttesiden for gitt, men bare analyserer kostnadene og vurderer om prosjektet er kostnadseffektivt

Der 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.

En må i tillegg se hen til resultater fra andre tilsvarende prosjekter. Dersom de fleste prosjekter av en viss type er ulønnsomme, kan resultatkravet reduseres noe, og vice versa.

I nytte-kostnadsanalyse og kostnads-virkningsanalyse:

Score 5-6 lønnsomme prosjekter ($NNV > 0$)

Score 3-4 lønnsomheten er lav eller nærmere null.

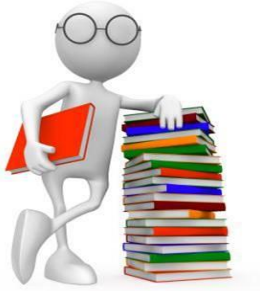
Score 1-2 ulønnsomme prosjekter ($NNV < 0$).

I kostnadseffektivitetsanalyse:

Score 5-6 prosjekter som er mer effektive enn gjennomsnittet for sektoren

Score 3-4 resultater på akseptabelt nivå eller på gjennomsnittet

Score 1-2 resultater klart under det akseptable.



Årsaksforklaringer og læringspunkter

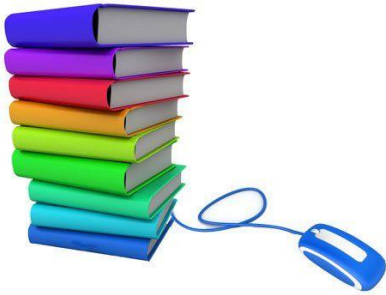
Både for berørte departementer og etater, og Concept-programmet er hovedformålet med evalueringer å samle lærdom som kan komme til nytte i senere prosjekter. Spørsmålet er altså ikke bare hvorvidt et prosjekt har vært vellykket, men også hvorfor man lyktes eller ikke, og hva man kan lære av det.

Et eget kapittel om årsaksforklaringer og læringspunkter skal derfor alltid inngå i rapporten, typisk som siste kapittel.

Stikkord kan være om det dreide seg om:

- realistisk kostnadsestimering,
- problemanalyse og behovsvurdering i tidligfasen,
- realistiske mål,
- bruk av prognoser, modeller og metoder,
- kompetanse hos aktører eller institusjoner,
- flere eller færre konsepter som ble vurdert,
- involvering/ikke involvering av bestemte aktører,
- medfinansiering og insentiver hos initiativtakere og andre,
- valg av gjennomføringsmodell,
- valg av kontraktstrategi
- usikkerhetsfaktorer som slo ut/ikke slo ut
- ansvar for gevinstrealisering, etc.

En bør skille mellom læringspunkter for departementet (som typisk gjelder konseptvalget) og læringspunkter for etaten (som typisk gjelder prosjektgjennomføringen), for eksempel ved å dele kapitlet i to.

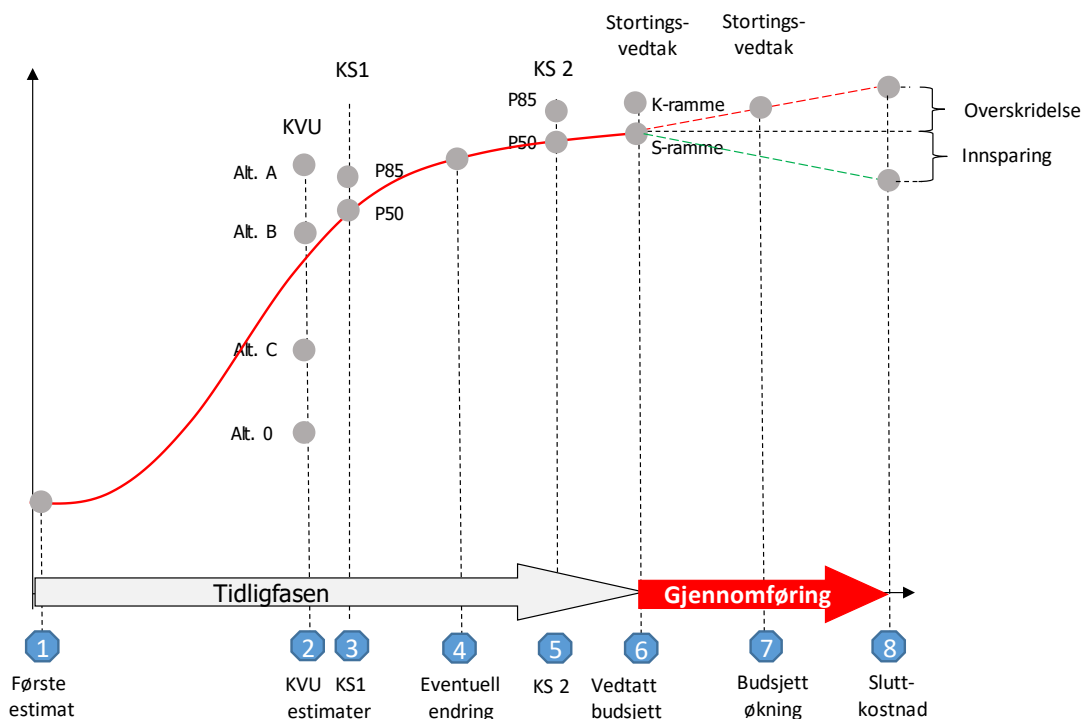


Kostnadsestimater og -tall som skal registreres

Figuren på neste side gir en oversikt over alle relevante kostnadsestimater og kostnadstall (nummerert).

1. Første estimat er det første tallet som er registrert i offentlig debatt eller hos utredere, og som derfor kan sies å være utløsende for prosessen som har ført frem til det aktuelle prosjektet. Kilder: dokumenter, møterefater, omtale i pressen, e.l.
2. Konseptvalgutredningen (KVU) skal komme frem til og vurdere minst to ulike konseptuelle løsninger i tillegg til nullalternativet, og et kostnadsestimat for hvert av disse, her angitt som A, B, C og 0. Estimatenes sannsynlighetsbasert og angitt med forventningsverdiene P85 og P50. KVUen skal anbefale hvilket konsept som bør velges.
3. Ved KS1 skal kvalitetssikrer foreta egen sannsynlighetsbasert estimering av de foreslåtte konseptene i KVUen, og anbefaling av hvilket konsept som bør velges. Kostnadsestimatene er angitt med forventningsverdiene P85 og P50. Kilder: KVU-rapport og KS1 rapport.
4. Departementet skal innen 6 måneder etter KS1 angi et styringsmål for prosjektet. Dette og eventuelle endringer i estimatene mellom KS1 og KS2 skal registreres, og årsaken til slike endringer skal forklares. Kilder: prosjektdokumenter, endringslogg og intervju med etat/departement
5. Ved KS2 skal kvalitetssikrer foreta egen sannsynlighetsbasert estimering styringsmålet for det valgte prosjektalternativet, tatt i betraktning eventuelle endringer som er kommet til etter KS1. Forventningsverdiene P85 og P50 skal angis. Kilde: KS2-rapport

6. Regjeringen fremmer forslag til prosjekt med kostnadsestimater for Stortinget. Det vedtatte budsjett oppgis ved kostnadsramme og styringsramme. Eventuelle avvik i forhold til forventningsverdiene ved KS2 (P85 og P50) forklares. Kilder: Stortingsproposisjon
7. I noen tilfeller gir Stortinget tilleggsbevilgninger. Dette registreres og forklares. Kilde: Stortingsproposisjon
8. Sluttkostnad er den samlede kostnaden etter avsluttet prosjekt. Dersom det er utestående fordringer eller uavklarte tvister angis anslag på beløpene dette gjelder. Kilde: intervju, sluttrapport, dokumentasjon



Sentrale referanser om evaluering

1. For mer informasjon om ordningen med ekstern kvalitetssikring av store statlige investeringsprosjekter, se www.ntnu.no/web/concept/ks-ordningen1
2. Volden, G. og K. Samset (2017). *Statlige investeringstiltak under lupen. Erfaringer med evalueringer av de første 20 KS-prosjektene*. Concept-rapport nr. 52
3. Samset, K. (2014a). *Evaluering av prosjekter. Vurdering av suksess*. Fagbokforlaget Vigmostad & Bjørke AS
4. Finansdepartementet (2014). *Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv*. Rundskriv R 109/14
5. Nettside med resultater fra tidligere evalueringer: www.ntnu.no/concept/etterevaluering-av-prosjekter

Andre referanser

6. Lædre, O. m.fl. (2012): *Levedyktighet og investeringstiltak*, Concept-rapport nr. 29
7. Direktoratet for økonomistyring (2014). *Veileder i samfunnsøkonomiske analyser*
8. Næss, P. (2005). *Bedre behovsanalyser. Erfaringer og anbefalinger om behovsanalyser i store offentlige investeringsprosjekt*, Concept-rapport nr. 5
9. Samset, K. (2014b). *I riktig retning. Prosjekters mål og målstruktur*. Concept temahefte nr. 5
10. Samset, K. m.fl. (2009). *Sykebus, fregatter og skipstunnel: Logisk minimalisme, rasjonalitet – og de avgjørende valg*, Concept-rapport nr. 21
11. Sunnevåg, K.J., red. (2007). *Beslutninger på svakt informasjonsgrunnlag. Tilnærminger og utfordringer i prosjekters tidlige fase*, Concept rapport nr. 17

