

for
p
e
c
n
o
c

Gro Holst Volden og Knut Samset

Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene

Concept-rapport nr 52





Gro Holst Volden og Knut Samset

Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene

Concept-rapport nr 52

Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene

Gro Holst Volden, Norges teknisk-naturvitenskapelige universitet
Knut Samset, Norges teknisk-naturvitenskapelige universitet

ISSN: 0803-9763 (papirversjon)
ISSN: 0804-5585 (nettversjon)
ISBN: 978-82-93253-61-7 (papirversjon)
ISBN: 978-82-93253-62-4 (nettversjon)

RETTIGHETSHAVER:

© Forskningsprogrammet Concept. Publikasjonen kan siteres fritt med kildeangivelse.

SAMMENDRAG:

Rapporten presenterer en oppsummering av resultatene fra etterevaluering av 20 store statlige investeringsprosjekter, i regi av forskningsprogrammet Concept. Det er benyttet en enkel evalueringsmodell som vurderer seks ulike aspekter ved prosjektenes vellykkethet. Meta-evalueringen viser blant annet at prosjektene kommer godt ut på tids- og kostnadskontroll, noe svakere på samfunnsøkonomisk effektivitet, og at det er forskjeller mellom sektorene. Rapporten omfatter også en begrunnelse og drøfting av selve evalueringsmodellen og en vurdering av hvordan den er anvendt av evaluatorene – dette vil resultere i et opplegg med tydeligere krav og bedre veiledning til gjennomføringen av evalueringene. Evalueringsarbeidet vil fortsette, og vil over tid gi en database med resultater fra et stort antall prosjekter, som skal komme til nytte i departementers og etaters arbeid med å forbedre plan- og analysearbeidet knyttet til fremtidige investeringsprosjekter.

DATO: September 2017

UTGIVER: Ex ante akademisk forlag
Concept-programmet
Norges teknisk- naturvitenskapelige universitet
7491 NTNU – Trondheim
www.ntnu.no/concept

Ansaret for informasjonen i rapportene som produseres på oppdrag fra Concept-programmet ligger hos oppdragstaker. Synspunkter og konklusjoner står for forfatterens regning og er ikke nødvendigvis sammenfallende med Concept-programmets syn. Concept-rapportserien er godkjent som vitenskapelig publiseringskanal på Nivå 1. Alle bidrag kvalitetssikres av uavhengige fagfeller.

Concept-rapportserien

Forskningsprogrammet Concept er forankret ved NTNU og arbeider med forskning knyttet til utviklingen og kvalitetssikringen av store investeringsprosjekter i Norge. Dette er tverrfaglig forskning innenfor fagområdene prosjektledelse, offentlig finansiering, statsvitenskap, samfunnsøkonomisk analyse og evaluering. Rapportserien presenterer forskningsresultater på programmets fagområder og er godkjent som vitenskapelig publiseringskanal på nivå 1. Målgruppen omfatter primært forskere på respektive fagområder og fagpersoner i offentlig forvaltning og utredningsmiljøer.

Redaksjon

Knut Samset, professor, NTNU, redaktør

Gro Holst Volden, forskningssjef Concept

Morten Welde, forsker, NTNU

Redaksjonsråd

Tom Christensen, professor Universitetet i Oslo

Petter Næss, professor, Norges miljø- og biovitenskapelige universitet

Nils Olsson, professor, NTNU

Ingeborg Rasmussen, daglig leder, Vista Analyse

Jørn Rattsø, professor, NTNU

Tore Sager, professor, NTNU

Arvid Strand, forsker 1, Transportøkonomisk institutt

Heidi Ulstein, partner, Menon Business Economics

Vibeke Binz Vallevis, gruppeleder, DNV GL

Bjørn Otto Elvenes, førsteamanuensis, NTNU

Forord

Concept-programmet ved NTNU tok i 2011 initiativ til å få gjennomført fire evalueringer av fire store statlige investeringsprosjekter som hadde gjennomgått ekstern kvalitetssikring (KS-ordningen) og blitt ferdigstilt og kommet i driftsfase. Dette for å bygge kompetanse og erfaring med evaluering og teste ut et generisk opplegg for evaluering av store statlige investeringsprosjekter. Funn, erfaringer og anbefalinger fra disse fire pilotevalueringene ble presentert i Concept-rapport nr. 30 (2013).

Modellen har nå blitt brukt på til sammen 20 prosjekter og metodikken og opplegget for øvrig er i ferd med å konsolideres. I denne rapporten presenterer vi aggregerte funn fra evalueringene og vurderer behovet for å ytterligere spesifisere og justere retningslinjene for evalueringsarbeidet. De oppdaterte retningslinjene vil bli publisert separat.

Alle evalueringene er dokumentert i egne rapporter som kan lastes ned fra Concept-programmets nettside. Nøkkelresultatene er også lett tilgjengelige i en egen database: www.ntnu.no/web/concept/etterevaluering

Vi vil takke alle evaluatorene som har deltatt i de 20 evalueringene, samt departementene, etatene og andre som har stilt opp som informanter, og vitenskapelig assistent Mai Ramdahl som har laget strukturerte sammendrag av dem (se rapportens del 3). En spesiell takk også til de 11 evaluatorene som deltok på en workshop og bidro med nyttige innspill og synspunkter på aggregerte funn og bruken av evalueringsmodellen, og til Erik Whist og Håkon Finne som har gitt nyttige innspill til denne sluttrapporten.

Trondheim, september 2017

Forfatterne

Innhold

Forord	1
Innhold	2
Sammendrag	4
Summary	12
1. Bakgrunn og innledning	18
Kvalitetssikring av statlige investeringsprosjekter	18
Evaluering og læringsløyper	20
Etterevaluering av KS-prosjekter	22
Leserveiledning	24
Del 1 Evaluering av de 20 første KS-prosjektene	
2. Prosjektene og opplegget for denne studien	27
Utvalget av prosjekter	27
Opplegget for evaluering	29
Metaevalueringen	31
3. Hvor vellykkete var de første 20 KS-prosjektene?	33
Operasjonell versus taktisk og strategisk vellykkethet	36
4. Evalueringsresultatene	41
5. Konklusjoner	51
God kostnadskontroll i gjennomføringsfasen	51
Effektene kommer ikke av seg selv	53
Læringspotensialet	54
Del 2. Vurdering av evalueringsopplegget og evalueringene	
6. Nærmere om metodikk for prosjektevaluering	57
Prosjektets årsak-virkningskjede må spesifiseres	57
Et balansert sett av seks evalueringskriterier	62
Gjennomføringen av evalueringene	68
7. Erfaringer med evalueringene	75
Generelle erfaringer	75
Evalueringsobjektet	77
Evalueringsteamet	78
Evalueringskriteriene	81
Datagrunnlaget og triangulering	89

Ambisjonsnivå versus tilgjengelige ressurser	90
Erfaringer med scoresetting	91
Årsaksforklaringer og læringspunkter	93
Involvering av målgruppen	94
8. Konklusjoner	95
Referanser	98

Del 3 Kort sammenfatning av evalueringene

1. Statlig kontrollområde Svinesund	103
2. Dobbeltspor Asker-Sandvika	104
3. E18 Momarken-Sekkelsten	105
4. Skjold-klasse MTBer	106
5. E6 Riksgrensen-Svingenskogen	107
6. Svalbard forskningspark	108
7. Lofoten fastlandsforbindelse LOFAST	109
8. Eiksundsambandet	110
9. NAV IKT Basis	111
10. Høgskolen i Østfold	112
11. E16 Kløfta-Nybakk	113
12. Rv 519 Finnfast	114
13. Dobbeltspor Sandnes-Stavanger	115
14. Regionfelt Østlandet	116
15. Statens pensjonskasse PERFORM	117
16. Halden fengsel	118
17. Nytt operahus i Oslo	119
18. E6 Svingenskogen-Åsgård	120
19. E6 Åsgård-Halmstad	121
20. Gevingåsen jernbanetunnel	122
Vedlegg 1 Populasjon og utvalg av ferdigstilte KS-prosjekter	123
Vedlegg 2 Sjekkliste for rapportgjennomgang. Prosjektenes suksess	126
Vedlegg 3 Sjekkliste for rapportgjennomgang. Gjennomføringen av evalueringene	128
Vedlegg 4 Fokusgruppemøte	130

Sammendrag

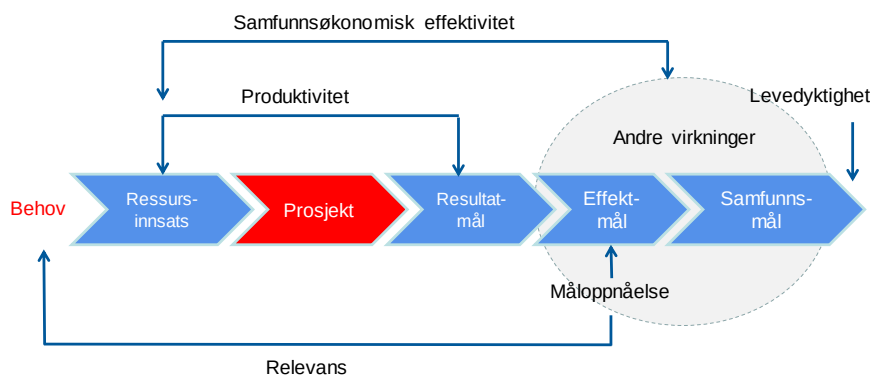
Overordnet etterevaluering av offentlige investeringsprosjekter og tiltak gjennomføres bare i begrenset grad i dag. Etter at kravet om ekstern kvalitetssikring av beslutningsunderlaget i landets største investeringsprosjekter (Statens prosjektmodell) ble innført i 2000, har mer enn 200 prosjekter vært kvalitetssikret under ordningen. I denne rapporten omtales disse som KS-prosjekter. I årene fremover blir stadig flere av dem ferdigstilt og satt i drift, samtidig som nye prosjekter kommer til. For å trekke erfaringer og lære er det nødvendig å følge prosjektene over tid og evaluere både gjennomføringen og hvilke effekter de har gitt for brukere og samfunn. Dette var bakgrunnen for at Concept-programmet for få år siden testet ut en modell for etterevaluering, som til nå har blitt benyttet på 20 KS-prosjekter. Denne metaevalueringen presenterer en oversikt over de aggregerte funnene, samt en vurdering av evalueringene, evalueringsmodellen, og behovet for retningslinjer for fremtidige evalueringer.

Concepts evalueringsmodell

Vi ønsket en bred, og samtidig enkel, vurdering som omfatter resultater langs hele prosjektets årsak-virkningsskjede, både ressursbruk, måloppnåelse på alle nivåer og sidevirkninger. Valget falt på en målorientert evalueringsmodell som anbefales av internasjonale organisasjoner, blant andre OECD og er mye brukt i evaluering av bistandsprosjekter. I tillegg gjøres en vurdering av den samfunnsøkonomiske effektiviteten. Dette gir til sammen seks overordnede evalueringskriterier:

1. *Produktivitet.* Det handler om gjennomføringen av prosjektet, det vil si hvor effektivt prosjektorganisasjonen har omsatt ressurser til leveranser, herunder om resultatmålene for kostnad, tid og kvalitet ble nådd som avtalt.
2. *Måloppnåelse.* Spørsmålet er om de avtalte effektmålene ble nådd, og i hvilken grad prosjektet har bidratt til dette.
3. *Andre virkninger.* Dette omfatter alle virkninger ut over de tilsiktede effektene, både positive og negative, kort- og langsiktige, for ulike berørte parter.
4. *Relevans.* Et prosjekt er relevant dersom det er *behov* for det prosjektet leverer. Det må vurderes i forhold til politiske prioriteringer, men også etterspørsel i markedet og interessegruppers behov.

5. *Levedyktighet*. Et prosjekt er levedyktig dersom de positive effektene av tiltaket vedvarer over hele levetiden. Det må være akseptabelt både økonomisk, miljømessig og fordelingsmessig.
6. *Samfunnsøkonomisk effektivitet*. Dette kan måles ved sammenhengen mellom betalingsvillighet og ressursbruk (lønnsomhet) eller mellom måloppnåelse og ressursbruk (formålseffektivitet).



Sammenhengen mellom evalueringskriteriene sett i forhold til prosjektets årsak-virkningskjede, er illustrert i figuren ovenfor.

I hvert tilfelle engasjeres et evalueringsteam. Teamet starter med å kontrollere og ved behov justere prosjektets årsak-virkningskjede (målstrukturen). Kravet er at det må være en logisk sammenheng mellom leddene, og at målene på hvert nivå er realistisk realiserbare. Deretter operasjonaliseres de seks evalueringskriteriene i konkrete forhold som skal undersøkes, såkalte evalueringsspørsmål eller indikatorer. Etter at relevante data er samlet inn og analysert skal evaluator gi en samlet vurdering av vellykkethet for hvert av de seks evalueringskriteriene. Det benyttes en poengskala fra 1 til 6, hvor 1 er laveste score (helt mislykket) og 6 er høyeste (svært vellykket).

De evaluerte prosjektene

Evalueringen gjelder noen av landets største statlige investeringsprosjekter i sektorene transport, bygg, forsvar og IKT. De bør ha vært i drift i om lag fem år. Denne studien omfatter omtrent halvparten av KS-prosjektene som er kommet så langt per i dag, og gir derfor et rimelig representativt bilde. De 20 evaluerte prosjektene, tidspunkt for evaluering og av hvem, fremgår av tabellen nedenfor.

Nr	Prosjekt	Sektor	Start (år)	Slutt (år)	Eval. (år)	Ansvarlig evaluator.*
1	Kontrollområde Svinesund	Bygg	2004	2005	2012	SINTEF
2	Asker-Sandvika	Jernbane	2001	2005	2012	VTI
3	E18 Momarken-Sekkelsten	Vei	2005	2007	2012	Concept
4	Skjold-klasse MTB-er	Forsvar	2003	2013	2012	Scanteam
5	E6 Riksgr.-Svingenskogen	Vei	2002	2004	2014	COWI
6	Svalbard forskningspark	Bygg	2003	2005	2014	Concept
7	Lofoten fastlandsforbindelse	Vei	2003	2007	2014	UiN
8	Eiksundsambandet	Vei	2003	2008	2014	Menon
9	NAV IKT Basis	IKT	2006	2010	2014	NIBR
10	Høgskolesenteret Østfold	Bygg	2003	2006	2015	SINTEF
11	E16 Kløfta-Nybakk	Vei	2005	2007	2015	Urbanet
12	Rv 519 Finnfast	Vei	2005	2007	2015	Menon
13	Sandnes-Stavanger	Jernbane	2005	2009	2015	Oslo Ec.
14	Regionfelt Østlandet	Forsvar	2002	2012	2015	Prokonsult
15	Perform	IKT	2008	2012	2015	Menon
16	Halden fengsel	Bygg	2006	2010	2016	Oslo Ec.
17	Nytt operahus Oslo	Bygg	2005	2008	2016	HRConsult
18	E6 Svingenskogen-Åsgård	Vei	2005	2008	2016	Menon
19	E6 Åsgard-Halmstad	Vei	2004	2005	2016	Menon
20	Gevingåsen jernbanetunnel	Jernbane	2009	2011	2017	Concept

*I noen tilfeller er evalueringen gjennomført av et konsortium, se rapportens del 3 for mer fullstendige opplysninger

Det dreier seg om de første prosjektene som gjennomgikk ekstern kvalitetssikring av kostnadsestimat og styringsunderlag (KS2) da denne ordningen ble innført i år 2000. Foreløpig har ingen av de som har gjennomgått kvalitetssikring av konseptvalget (KS1) kommet så langt at de er ferdigstilt, og kan derfor ikke evalueres.

Denne metaevalueringen bygger i hovedsak på evalueringsrapportene fra de 20 prosjektene. I tillegg ble det gjennomført et fokusgruppemøte med 11 sentrale ressurspersoner blant de som har gjennomført evalueringene. Disse ble konfrontert med våre foreløpige funn og vurderinger, og fikk anledning til å dele sine erfaringer og vurderinger av metoden, opplegget og behovet for

veiledning. Forfatterne har også trukket på egne erfaringer fra prosessen med å bestille, følge opp og kvalitetssikre evalueringene, samt fra å delta i tre av dem.

Aggregerte resultater

I figuren nedenfor har vi beregnet gjennomsnittlig score på evalueringskriteriene for alle prosjektene innenfor hver sektor. De som ligger i øvre tredjedel av skalaen er markert med grønt, de som ligger i midtre med gult. Det er ingen i nedre tredel.

Et nøkkelresultat er at prosjektene kommer godt ut i et operasjonelt perspektiv (*produktivitet*). Det innebærer at kostnadskontrollen er god og at prosjektene er tids- og kostnadseffektivt gjennomført. Mye tyder på at KS2-ordningen har bidratt til dette. Kanskje mer overraskende var at prosjektene også scorer bra på de øvrige (taktiske og strategiske) kriteriene. Et klart flertall av prosjektene scorer 5 eller 6 på *relevans*. Det er med andre ord stor enighet om at investeringene er i samsvar med prioriteringer og behov i samfunnet. De fleste scorer også godt på *måloppnåelse* og *levedyktighet*. Det vil si at en har oppnådd det en forventet, og at det er grunn til å tro at de positive konsekvensene av tiltaket vil vedvare over tid. Resultatene for *andre virkninger* er noe svakere. Det kan skyldes at en ser negative sidevirkninger, eller at en ikke ser mange positive sådanne. Prosjektene scorer lavest på *samfunnsøkonomisk effektivitet*. Samtidig merker vi oss at mange av prosjektene som vurderes som ulønnsomme samtidig har fått god score på relevans. Dette er en indikasjon på at det ikke er markedet som avgjør i disse offentlig initierte prosjektene, men andre hensyn, slik de kommer til uttrykk gjennom politiske aktører og prosesser.

Sektor	Antall prosjekt	Produktivitet	Mål- oppnåelse	Andre virkninger	Relevans	Levedyktighet	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

Med forbehold om at antall prosjekter er lite, viser tabellen enkelte interessante sektorforskjeller. Når det gjelder produktivitet, kommer byggsektoren jevnt

over godt ut, mens transportsektoren (vei og jernbane) kommer noe svakere ut. Videre skiller bygg og jernbane seg negativt ut hva gjelder måloppnåelse og lønnsomhet – dette i motsetning til veiprosjekter som kommer godt ut på begge disse kriteriene. Til dette må påpekes at de evaluerte veiprosjektene var mer avgrensede og hadde mindre ambisiøse effektmål, mens både bygg- og jernbaneprosjektene har mål som krever større og mer integrert innsats, også ut over selve prosjektet. Det er viktig å være oppmerksom på hva som kreves når en setter ambisiøse effektmål, herunder å kunne vurdere effekten av å kombinere investeringsprosjekter med andre virkemidler.

Erfaringene med evalueringsmodellen

I del 2 av denne rapporten ser vi på selve evalueringsarbeidet og hvordan det er gjennomført. Dette er viktig ettersom evalueringene ble gjennomført både for å teste ut modellen og en rekke aktuelle evalueringsmiljøer.

Erfaringene med modellen er generelt gode. De seks kriteriene beskriver til sammen de viktigste aspektene ved prosjekters vellykkethet. Prosessen med å disaggregere evalueringskriteriene i konkrete indikatorer og deretter aggregere funnene opp for å gi kvalifiserte svar på hvert enkelt av dem, gir en god balanse mellom behovet for standardisering og fleksibilitet. Datagrunnlaget vurderes som akseptabelt i de fleste evalueringene.

Vår vurdering av evalueringene er generelt positiv, tatt i betraktning de begrensede ressursene som er satt av til hver enkelt studie (tilsvarende ca. 3 månedesverk). Flertallet av evalueringene gir god valuta for pengene og leverer gode, overordnede analyser av prosjektene. Samtidig ser vi at enkelte evalueringer er for svake, enten kvalitetsmessig, de fanger ikke opp helt essensielle problemstillinger, de graver ikke dypt nok til å finne årsaken til det som skjer, og tolkningen og operasjonaliseringen av evalueringskriteriene er ikke konsistent.

Vi har registrert et potensial for å forbedre opplegget ytterligere. Det handler blant annet om at:

- Dersom prosjektets kostnadsramme endres etter Stortingets opprinnelige budsjettvedtak, må det være en felles praksis for hvordan dette håndteres. Hovedregelen bør være at sluttkostnaden vurderes mot opprinnelig ramme (i faste kroner). Eventuelle endringer forklares og kommenteres.
- Evalueringene bør få frem mer informasjon om bakgrunnen for og tidligfasen i prosjektene, samt utviklingen i kostnadsestimatet for

budsjettet fastsettes. Det er nyttig for å vurdere prosjektets relevans så vel som kostnadseffektivitet.

- Formelt avtalte mål må være realistisk oppnåelige, ikke trivielle eller uoppnåelige. Evaluators må vurdere målstrukturen kritisk, og ikke legge til grunn en prosjektlogikk med store feil. En eventuell justert målstruktur må presenteres som evaluators egen tolkning.
- Kvaliteten på evalueringene avhenger av informasjonens validitet og reliabilitet. Evaluator må ha en profesjonell tilnærming og der det er nødvendig å triangulere mellom ulike kilder og metoder. De fleste gjør dette, men noen har et forbedringspotensial.
- En annen utfordring er at evaluatorene tolker de strategiske evalueringskriteriene ulikt og dels overlappende. *Andre virkninger, relevans og levedyktighet* skal behandles hver for seg og i samme detaljeringsgrad, ikke reduseres til å være elementer i en samfunnsøkonomisk analyse.
- Faglig bredde er viktig for å unngå en skjev tilnærming og mangelfull vurdering av prosjektet. Mange av teamene har hatt et flertall av samfunnsøkonomer. Hvert team skal ha tilstrekkelig kompetanse på fire områder: *evalueringsfaglig, samfunnsøkonomisk, prosjektfaglig (herunder estimeringskompetanse) og sektorfaglig.*
- Bruken av score vurderes som nyttig, men det er behov for bedre kalibrering av scoresettingen. En skal unngå, men kan ikke utelukke, positiv skjeving i enkelte av evalueringene.
- Mange av evalueringene har primært hatt fokus på å svare på *hvorvidt* prosjektene lykkes, og i mindre grad forsøkt å forklare årsakene til resultatene og trekke ut lærdommer til nytte for nye prosjekter. I fremtidige evalueringer bør en fokusere enda mer på dette.

Veien videre

Lærdommene etter denne runden vil resultere i et revidert opplegg med tydeligere krav til hvordan evalueringene skal gjennomføres, evaluators kompetanse og erfaring, og nærmere veiledning til hvordan modellen og kriteriene skal forstås. Noen elementer, som kanskje ikke er viktige i den enkelte evaluering, bør allikevel registreres fordi de vil være nyttige i en fremtidig metaevaluering.

Parallelt med å forbedre evalueringsopplegget må vi også jobbe mer med å involvere departementer og etater slik at de ser nytten av og læringspotensialet i evalueringene. Enkel tilgang til nøkkelresultater er viktigst. Dette er sikret ved

hjelp av et interaktivt format på Concept-programmets hjemmesider, der alle evalueringene også er tilgjengelig i sin helhet.

Dette er bare begynnelsen – planen er å bruke evalueringsmodellen på en stor andel av de mange investeringsprosjektene som settes i drift nå og i årene fremover. I neste runde blir det særdeles viktig å evaluere prosjekter som har vært gjennom KS1, hvor disse 20 første evaluerte prosjektene vil utgjøre en kontrollgruppe. 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*.

Summary

English title: A Close-up on Public Investment Cases. Lessons from Ex-post Evaluations of 20 Major Norwegian Projects.

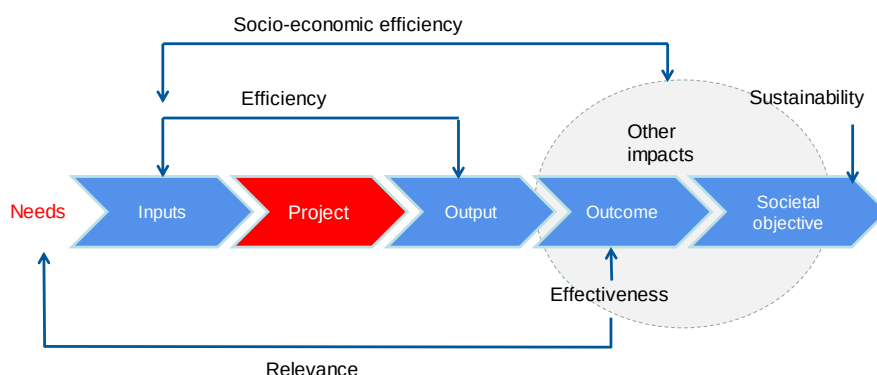
The Norwegian scheme for external quality assurance of the decision basis in major public investment projects– the QA scheme – was introduced year 2000. Since then, more than 200 projects have been processed under the scheme. In the years to come, more and more of these will be completed and put into operation. In order to learn from them, they will be evaluated, with a focus on both implementation and their effects on users and on society as a whole. For this reason, the Concept programme has tested a model for post-evaluation on 20 so-called QA projects. In this meta-evaluation, we present an initial overview of the aggregated findings, as well as an assessment of the evaluations, the evaluation model, and the reasons behind the need for better guidelines.

The Concept evaluation model

The evaluations should cover all relevant aspects of the investment cases, while at the same time being simple to implement. The decision fell on a broad, goal-oriented evaluation model recommended by international organizations, including the OECD, in combination with economic analysis. Our model thus covers six general evaluation criteria:

1. *Efficiency*. This is a measure of the project's implementation: how effectively the project organization has converted resources to deliveries. This includes an assessment of the project's *outputs* in terms of cost, time and quality.
2. *Effectiveness*. This concerns whether the agreed *outcomes* (or first-order effects) were reached and to what extent the project contributed to this outcome.
3. *Other impacts*. This includes all effects beyond the intended effects (the outcome) that can be attributed as the result of the project, positive and negative, short-term and long-term, and for different stakeholders and affected parties.
4. *Relevance*. A project is relevant if there is a *need* for it. Project relevance is measured in relation to needs and priorities as expressed politically and by stakeholders and affected parties.

5. *Sustainability*. A project is sustainable if its positive effects are likely to persist throughout its lifetime.
6. *Socio-economic efficiency*. This can be measured in terms of either the users' willingness to pay in relation to cost (profitability) or the outcome in relation to cost (often denoted as cost-effectiveness).



The evaluation team starts by checking and, if necessary, calibrating the results chain (see figure) and goals on each level. There must be logical consistency, and the goals at each level must be realistic in terms of their level of ambition. Thereafter, the six evaluation criteria are operationalized in terms of the specific issues to be investigated – the ‘evaluation questions’ or ‘indicators’. After the relevant data have been collected and analyzed, the evaluator will provide an overall assessment of the success on each of the six evaluation criteria. The score scale ranges from 1 to 6, where 1 is the lowest score (completely unsuccessful) and 6 is highest (very successful).

The evaluated projects

The evaluated projects are some of the country’s largest public investment projects in the transport, building construction, defense, and ICT sectors. The projects need to have been in operation for about five years. This study examines about half of the QA projects that have been finalized to date and therefore provides a reasonably representative picture. Details of the 20 evaluated projects, the time of their evaluation, and who conducted the evaluation are listed in the table below.

No.	Project	Sector	Start (Yr)	End (Yr)	Eval. (Yr)	Evaluator*
1	Border control Svinesund	Building	2004	2005	2012	SINTEF
2	Asker–Sandvika double track	Railway	2001	2005	2012	VTI
3	E18 Momarken–Sekkelsten	Road	2005	2007	2012	Concept
4	Skjold class MTB	Defence	2003	2013	2012	Scanteam
5	E6 Riksgr.–Svingenskogen	Road	2002	2004	2014	COWI
6	Svalbard Research Park	Building	2003	2005	2014	Concept
7	Lofoten mainland connection	Road	2003	2007	2014	UiN
8	Eiksund mainland connection	Road	2003	2008	2014	Menon
9	NAV ICT Basis	ICT	2006	2010	2014	NIBR
10	University college campus	Building	2003	2006	2015	SINTEF
11	E16 Kløfta–Nybakk	Road	2005	2007	2015	Urbanet
12	Rv 519 Finnfast	Road	2005	2007	2015	Menon
13	Sandnes–Stavanger railway	Railway	2005	2009	2015	Oslo Economics
14	Regional exercise field East	Defence	2002	2012	2015	Prokonsult
15	Perform ICT procurement	ICT	2008	2012	2015	Menon
16	Halden Prison	Building	2006	2010	2016	Oslo Ec.
17	New Opera House, Oslo	Building	2005	2008	2016	HRCConsult
18	E6 Svingenskogen–Åsgård	Road	2005	2008	2016	Menon
19	E6 Åsgård–Halmstad	Road	2004	2005	2016	Menon
20	Gevingsåsen railway tunnel	Railway	2009	2011	2017	Concept

* In some cases, the evaluation was carried out by a consortium (see part 3 of this report for further information)

These are the first projects that underwent external quality assurance of their cost estimates and management basis (QA2) when the scheme was introduced in the year 2000. None of them has undergone quality assurance of the choice of conceptual solution (QA1) which was introduced later, in 2006.

The meta-evaluation presented here is based mainly on the 20 evaluation reports. In addition, we organized a focus group meeting with 11 of the evaluators. We presented our preliminary findings and assessments, and they shared their experiences and assessments of the methodology, the evaluation scheme, and the need for guidance. In addition, the two authors have drawn

on their own experiences of the process, from commissioning, follow-up and quality assurance of the evaluations, as well as participating in three of them.

Aggregated results

Our key finding is that the projects performed well in an operational perspective (*efficiency*), which means that they were finished within budget and on time, and were time- and cost-efficient. There is much to indicate that this can be attributed to the QA2 scheme. Perhaps more surprising is that the projects also scored well on the tactical and strategic criteria. In terms of averages for all projects, the scores were between 4 and 5. A clear majority of the projects scored 5 or 6 on *relevance*. This indicates that there was wide agreement among the evaluators that the investments were in line with the priorities and needs in society. Most projects also scored well on *effectiveness* and *sustainability*, which indicates that expectations had largely been fulfilled and that there is good reason to believe that the positive effects of the projects will last over time. The results regarding the projects' *Other impacts* were somewhat weaker. This was due to the negative external effects of some projects, and the lack of positive external effects in others. The projects scored lowest on *socio-economic efficiency*. At the same time, we noted that a number of projects considered unprofitable scored positively on relevance. This indicates that the market was not the determining factor in these publicly initiated projects, but rather other considerations, as expressed by political actors and processes.

Sector	Projects	Efficiency	Effective- ness	Other impacts	Relevance	Sustain- ability	Socioec. efficiency
Construction	5	5,4	4,2	4,6	4,6	4,8	3,8
Defense	2	4,5	4,5	4,5	4,5	3,5	3,5
ICT	2	5,0	5,5	4,5	4,0	5,5	4,0
Railway	3	4,3	3,3	4,0	4,7	4,7	2,7
Roads	8	4,4	5,3	4,3	4,6	4,5	5,3
Average		4,7	4,7	4,4	4,6	4,6	4,2

Although the number of projects is small, the table above shows some interesting sectoral differences. In terms of efficiency, the building construction sector performed well, while the transport sector (road and rail) performed somewhat weaker. Further, buildings and railway projects scored low on effectiveness and profitability, in contrast to road projects, which

scored better on both criteria. It should be noted, however, that the road projects were more limited in scope and had less ambitious objectives, while the buildings and railway projects had goals that required larger and more integrated efforts, also beyond the projects themselves. This finding illustrates the importance of considering what is required when setting ambitious goals.

Experience from using the evaluation model

Experiences of using the model were generally positive. In combination, the six evaluation criteria provide a comprehensive coverage of the most essential aspects of success in projects. The process of disaggregating the criteria into specific indicators and then aggregating the findings to provide qualified answers to each of them provides a good balance between the need for standardization and the need for flexibility.

However, we have noted the potential for further improvements, with respect to, for example, the following:

- There must be a common practice for how to handle the situation when the budget agreed by Parliament is changed at a later date due to changes in project costs. The principle should be that the final cost is assessed against the original budget in all cases (fixed prices).
- Evaluations should provide more information about the justification and front-end development of the project, as well as the range of cost estimates before the budget is determined. This would be useful when evaluating the project's relevance and cost-effectiveness.
- The evaluator should make a thorough assessment of the project's goal hierarchy and not base their judgment on a flawed project logic. Any adjustments should be explained and presented as the evaluator's own interpretation.
- The quality of the evaluations depend on the validity and reliability of information. Triangulation between different information sources and methods may be necessary. Most evaluators do this, but some have potential for improvement.
- Another challenge is that evaluators interpret the strategic criteria differently and as partly overlapping. Other impacts, relevance and sustainability should be assessed independently and to the same extent, not merely as elements in a socio-economic analysis.
- An interdisciplinary approach is essential to avoid a skewed and inadequate assessment of the project. The majority of members in

many of the evaluation teams were economists. A broader spectrum is thus necessary, and to ensure adequate evaluation experience.

- The use of scores is considered useful, but positive skewed results were suspected in some of the evaluations. There is a need for better calibration of the scoring system.
- Many of the evaluations had primarily focused on whether the projects had succeeded, and to a lesser extent on explaining the reasons and drawing lessons that would benefit new projects. The latter should be emphasized even more in future evaluations.

The way forward

The lessons learned from this meta-evaluation will result in certain amendments to the evaluation scheme, with clearer requirements regarding how evaluations are to be carried out, evaluators' qualifications, and further guidance on how the methodology and evaluation criteria are to be interpreted. Some issues that may be less essential in individual evaluations should nevertheless be included, since they may be useful in future meta-evaluations. In parallel with improving the evaluation scheme, there is a need to involve ministries and agencies to a greater extent, so that they can see the usefulness and learning potential of the evaluations. Easy access to key findings and conclusions is important. The latter is now ensured by, among other things, all evaluation results being available through an interactive presentation format on the Concept programme's website, where also the full reports can be downloaded.

However, this is just the beginning and in the years to come a large proportion of the many investment projects under the Norwegian QA scheme will be evaluated during their operational phase. This should give the answer to whether the patterns we have seen in this study also apply to a larger portfolio of projects. In the next round, the challenge will be to evaluate projects that have been through QA1, and the 20 evaluated projects discussed here will constitute a control group in an assessment of the impacts of the expanded scheme.

1. Bakgrunn og innledning

Forskningsprogrammet Concept utvikler kunnskap som bidrar til bedre konseptvalg, ressursutnytting og effekt av investeringsprosjekter. En sentral oppgave er å drive følgeforskning på de største statlige prosjektene som er underlagt Finansdepartementets krav om ekstern kvalitetssikring av beslutningsgrunnlaget. Siden ordningen ble innført i år 2000, har mer enn 200 prosjekter vært gjenstand for kvalitetssikring. Etter hvert er mange av disse ferdigstilt og i drift. Concept publiserte i 2013 en oversikt over sluttkostnadene i de 40 første ferdigstilte prosjektene som viste at kostnadskontrollen var godt idet 80 prosent av prosjektene holdt kostnadsrammen (Samset og Volden, 2013). Senere oppdateringer av ferdigstilte prosjekter har bekreftet disse funnene (Welde, 2017).

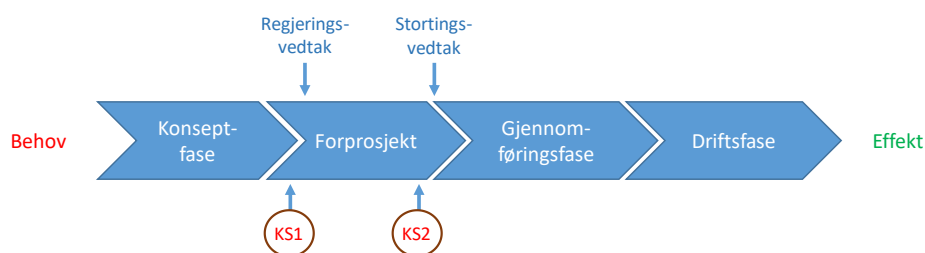
For å lære og trekke erfaringer er det nødvendig å følge opp prosjektene et stykke ut i driftsfasen og evaluere i hvilken grad de planlagte effektene for brukere og samfunn er realisert. Slike overordnede etterevalueringer av statlige tiltak gjennomføres i liten grad i dag. Dette var bakgrunnen for at Concept i 2011-2013 testet ut en modell for evaluering av KS-prosjektene (Volden og Samset, 2013). Til nå er til sammen 20 prosjekter evaluert etter denne modellen, og det er på tide å gi en foreløpig oppsummering av funnene samt en oppdatert vurdering av evalueringsmodellen og anvendelsen av den.

Kvalitetssikring av statlige investeringsprosjekter

Staten investerer hvert år milliardbeløp i store offentlige prosjekter, nesten 50 milliarder kroner i 2015, når en holder bl.a. petroleumsvirksomheten utenfor. Det dreier seg om veier og jernbaneanlegg, signalbygg i kultursektoren og andre formålsbygg for statlig tjenesteproduksjon, forsvarsmateriell, IKT-prosjekter og mye annet. For å redusere fremtidige problemer med store kostnadsoverskridelser, forsinkelser og rene feilinvesteringer, etablerte Finansdepartementet i år 2000 en ordning med ekstern kvalitetssikring av beslutningsunderlaget for de aller største investeringsprosjektene (Statens prosjektmodell)¹. Den gjelder i dag for de største statlige investeringsprosjektene med antatt investeringskostnad over 750 millioner kroner.

¹ Se Berg m.fl. (1999) hvor behovet for bedre styring av statlige prosjekter dokumenteres.

Ordningen innebærer at det stilles krav til omfang og innhold av de utredningene som gjøres av departementer og etater i de enkelte tilfeller. Først en konseptvalgutredning (KvU) der minst tre alternative konseptuelle løsninger skal være utredet (inklusive null-alternativet). Grunnlaget skal være en behovsanalyse, spesifisering av mål og krav, en mulighetsanalyse, og en samfunnsøkonomisk analyse av alternativene. Senere, i forprosjektfasen, skal det utarbeides et styringsdokument for å sikre at kostnadsrammen er realistisk og at styringsmessige utfordringer er identifisert og kan ivaretas. Kvalitetssikringen av disse dokumentene skjer ved hjelp av eksterne kvalitetssikrere som er prekvalifisert av Finansdepartementet. Dette er påkrevet før dokumentene kan legges frem for henholdsvis regjeringen og Stortinget.



Figur 1.1 Finansdepartementets kvalitetssikringsordning

Siden 2005 har det vært to slike kontrollpunkter, KS1 og KS2 som vist i figur 1.1:²

- KS1 Kvalitetssikring av konseptvalg før vedtak i regjeringen om å starte forprosjekt. Konseptet skal være i tråd med samfunnets behov, mål og krav og det skal gjøres en samfunnsøkonomisk analyse.
- KS2 Kvalitetssikring av styringsunderlag samt kostnadsoverslag for fremleggelse for endelig investeringsbeslutning i Stortinget.

Hensikten med de to kontrollpunktene er å sikre at den faglige kvaliteten på beslutningsunderlaget er slik at myndighetene får bedre overordnet styring med store statlige investeringer, herunder at konseptvalget undergis reell politisk styring. Enkelt sagt handler KS1 om å sikre *taktisk* og *strategisk*

² Mer informasjon om Statens prosjektmodell finnes på www.regjeringen.no/no/tema/okonomi-og-budsjett/statlig-okonomistyring/ekstern-kvalitetssikring2/id2523818/

suksess, dvs. at tiltaket gir den forventede nytten, samlet sett bidrar til en samfunnsmessig ønsket utvikling over tid, og innebærer en effektiv ressursallokering. KS2 handler om å sikre *operasjonell* prosjektsuksess, dvs. at leveransen foreligger som avtalt - og er produsert på en tids- og kostnadseffektiv måte (Samset, 2014).

Evaluering og læringssløyfer

Statens prosjektmodell skal sikre at underlaget for de mest kritiske beslutningspunktene i prosjektene er av god kvalitet. Dette skal i sin tur bidra til at både departementene som prosjekteiere og etatene som gjennomførende part blir stadig flinkere til å planlegge, gjennomføre og realisere effektene av store statlige investeringsprosjekter. Men for å undersøke om dette skjer (og hvis ikke, hvorfor ikke) så er det viktig med systematisk læring og tilbakeføring av erfaringer til tidligfasen i nye prosjekter.

Statens økonomiregelverk krever at det gjennomføres evalueringer som et supplement til den løpende mål- og resultatstyringen av statlig virksomhet³. Dette kravet omfatter både investeringsprosjekter og annen virksomhet. Frekvens og omfang av evalueringene må bestemmes ut fra prosjektenes egenart, risiko og vesentlighet. En evaluering kan derfor være alt fra enkle stikkprøver til omfattende undersøkelser som likner forskningsprosjekter.

Evalueringsvirksomheten i Staten er varierende men omfattende. I perioden 2005-2011 ble det levert omtrent hundre evalueringsrapporter årlig (Askim m.fl., 2013). De omfatter de fleste politikkområder, og det er først og fremst underliggende etater og ikke departementene selv som er oppdragsgiver. Det er adskillig færre *prosjekter* som blir gjenstand for formell evaluering, og enda færre som gjennomføres etter at de har vært i drift noen år for å fastslå effekten av disse. Som et unntak kan nevnes at Samferdselsdepartementet i 2006 tok initiativ til en ordning for etterprøving av de viktigste prissatte virkningene av et utvalg vei- og jernbaneinvesteringer. Dette for å undersøke om prosjektenes lønnsomhet er som forventet og indirekte for å undersøke godheten av nytte-kostnadsmetodikken for transportprosjektene.

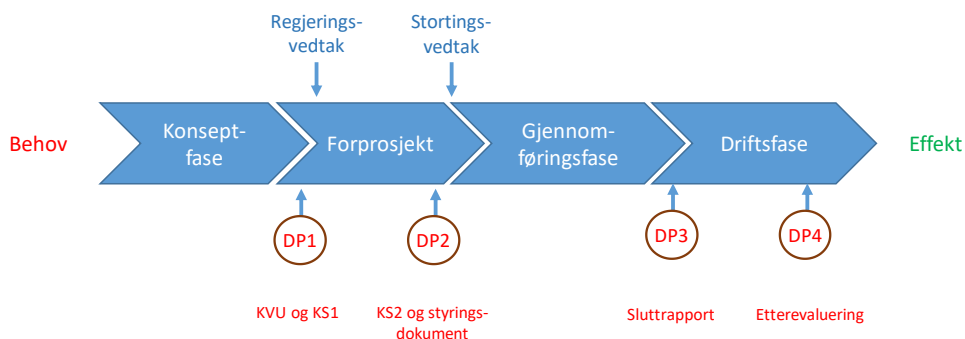
Formelt er det ikke andre regler som gjelder for de store statlige prosjektene som faller inn under kvalitetssikringsordningen, de såkalt KS-prosjektene. Disse kommer allikevel i praksis i en kategori for seg som spesielt interessante evalueringsobjekter, fordi kravene til utforming og analyse i forkant har vært mer omfattende for disse enn for andre prosjekter. Finansdepartementet som

³ Reglement for økonomistyring § 16. Se også Finansdepartementet (2005).

administrerer kvalitetssikringsordningen ønsker at det gjennomføres etterevaluering av prosjektene, og stiller midler til rådighet for dette. For forskningsprogrammet Concept, som er ansvarlig for følgeforskningen, er slik informasjon av avgjørende betydning, og en har derfor tatt på seg ansvaret for å utvikle evalueringsmetodikken, lyse ut evalueringsoppdrag og sikre at evalueringene blir gjennomført.

Etterevaluering av KS-prosjektene gjør at man da får fire evaluerings- eller datapunkter (DP) i sekvens som illustrert i figur 1.2.:

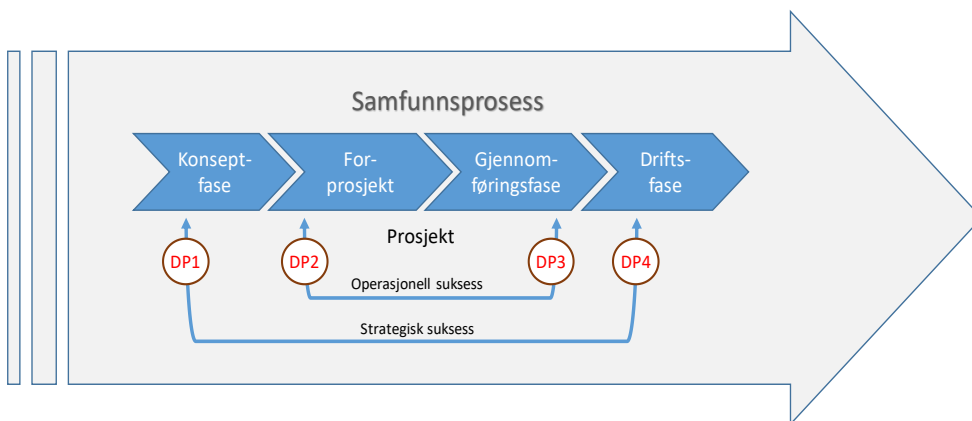
- Kvalitetssikringen KS1 innebærer evaluering av konseptvalget og etabler baseline-informasjon for etterevaluering av effekten på sikt.
- Kvalitetssikringen KS2 er en evaluering av prosjektplanen og gir baseline for evaluering av prosjektet som sådan.
- Prosjektets sluttrapport inklusiv prosjektrekskap gir viktig evalueringsinformasjon om selve prosjektet etter overlevering, her kalt DP3.
- Etterevalueringen noen år inn i driftsfasen gir informasjon om selve investeringstiltaket sett i et bredt samfunnsperspektiv, her kalt DP4.



Figur 1.2 De fire datapunktene i prosjekter som gjennomgår både ekstern kvalitetssikring og etterevaluering

Dette gir grunnlag for to typer læringssløyer, som vist i figur 1.3. Den lille læringssløyfen DP2 – DP3 bidrar til å forbedre en prosess uten å stille spørsmål ved om det er et riktig eller fornuftig prosjekt (prosess), på engelsk kalt *single loop learning* (Argyris og Schön, 1978). Evalueringen blir dermed avgrenset til å vurdere det som betegnes som prosjektets operasjonelle suksess. Dette kan gi nyttige innspill i bestrebelsene for å forbedre det som har å gjøre med prosjektledelsen i fremtidige prosjekter.

Den store læringsløyfen DP1 – DP4 derimot, tar tak i de grunnleggende spørsmålene om hvorfor, og som kan føre til at hele prosessen endres eller erstattes av noe annet. Dette kalles *double loop learning*. Her handler det om tiltakets taktiske og strategiske suksess. For at slik læring skal være mulig må man inn med evaluering i etterkant der man kan fastslå effekten og nytten av tiltaket. Men hvis denne innsikten virkelig skal gi en læringseffekt, må disse erfaringene være tilgjengelig i forkant før planleggingen av nye tiltak, hvor det er mulig å komme inn og korrigere for det mer grunnleggende og prinsipielle.



Figur 1.3 Den store og den lille læringsløyfen i statlige investeringsprosjekter

Etterevaluering av KS-prosjekter

Et prosjekt er per definisjon et tiltak for å nå bestemte mål. Måloppnåelse må derfor stå sentralt når en skal vurdere om det ble vellykket. Det skilles mellom mål på tre nivåer: resultatmål, effektmål og samfunnsmål. En evaluering bør alltid starte med å fortolke prosjektlogikken, det vil si sammenhengen mellom målene på de tre nivåene. Målene skal være konsistente og på riktig nivå i årsak-virkningskjeden, samt med passende ambisjonsnivå for å kunne brukes som referanseverdier i en evaluering. I tillegg til måloppnåelsen må evalueringer undersøke utilsiktede virkninger eller sidevirkninger. Og til slutt bør evaluator vurdere om ressursbruken er effektiv og at målene er nådd uten sløsing.

Concept-programmet har til hensikt å sikre gode evalueringer som gir læring til fremtidige investeringsprosjekter. Vi ønsker en bred vurdering som omfatter de ovennevnte perspektivene på prosjektsuksess, dvs. måloppnåelse på alle nivåer, sidevirkninger og effektiv ressursbruk, og som samtidig er enkel. Valget falt på en målorientert evalueringsmodell som anbefales av blant

andre OECD og er mye brukt i evaluering av bistandsprosjekter internasjonalt. Modellen omfatter fem overordnede kriterier, som på norsk kan oversettes til *produktivitet*, *måloppnåelse*, *andre virkninger*, *relevans* og *levedyktighet*. Prosjektets *produktivitet* er et mål på operasjonell vellykkethet, *måloppnåelse* er et mål for taktisk vellykkethet, og de tre siste for strategisk vellykkethet. I tillegg supplerte vi med *samfunnsøkonomisk effektivitet* som et sjettede kriterium, som står sterkt i KVVU-/KS1-ordningen. Effektivitet kan måles på taktisk eller strategisk nivå, avhengig av hvordan det defineres. Se kapittel 6 for en nærmere diskusjon og begrunnelse for valget av evalueringsmodell. Avgrensningen mellom taktisk og strategisk kan være uklar i praksis. I fortsettelsen vil vi i mange tilfeller snakke om *taktisk og strategisk* vellykkethet samlet, og viser da til alt som går ut over selve prosjektet og handler om effektene av det for brukere og samfunn.

Kriteriene er generiske og er derfor anvendbare på alle typer tiltak. Det gjelder:

- *Produktivitet*. Dette handler om gjennomføringen av investeringsprosjektet, det vil si hvor effektivt prosjektorganisasjonen har omsatt ressurser til leveranser. En må spørre om resultatmålene for henholdsvis kostnad, tid og kvalitet ble nådd som avtalt. I tillegg må en vurdere om arbeidet kunne vært gjort billigere, raskere eller med bedre kvalitet.
- *Måloppnåelse*. Spørsmålet er om de avtalte effektmålene ble nådd, og i hvilken grad prosjektet vurderes å ha bidratt til dette. Dersom det avtalte målet er lite egnet til evaluering, for eksempel fordi det er trivielt eller urealistisk, bør evaluator avklare hvilket ambisjonsnivå man legger til grunn.
- *Andre virkninger*. Prosjekter har ofte utilsiktede virkninger. Disse må dokumenteres og en må da tenke bredere på hvilke konsekvenser tiltaket har hatt, positive og negative, forventede og ikke forventede, kort- og langsiktige, ikke bare for de mest sentrale målgruppene, men også for andre berørte parter.
- *Relevans*. Et prosjekt er relevant dersom det er i samsvar med viktige prioriteringer i samfunnet og for sentrale brukergrupper, altså om det er *behov* for det prosjektet leverer. Det må vurderes i forhold til både politiske prioriteringer, etterspørsel i markedet, og interessegruppers behov.
- *Levedyktighet*. Et prosjekt er levedyktig dersom de positive effektene av tiltaket vedvarer over hele levetiden. Dette har både en økonomisk, miljømessig og sosial/fordelingsmessig side, samt at det er spørsmål om fleksibilitet til å tilpasse seg endrete rammebetingelser.

- *Samfunnsøkonomisk effektivitet.* Et prosjekt er samfunnsøkonomisk *lønnsomt* dersom den samlede betalingsvilligheten for de effektene prosjektet genererer overstiger kostnadene. Dersom det er bestemt at et gitt mål skal nås, kan en i stedet spørre om tiltaket er *formålseffektivt*, dvs. at en har nådd målet med lavest mulig ressursbruk.

Evalueringsprosessen innebærer en prosess med systematisk disaggregering og aggregering av informasjon. Evaluatoren disaggregerer ved å operasjonalisere hvert av de seks kriteriene, det vil si å velge ut mer konkrete forhold som skal undersøkes for prosjektet, såkalte indikatorer. Deretter må relevante data samles inn og analyseres. Til slutt aggregeres dette opp til øverste nivå ved å gi score for hvert kriterium på en skala fra 1 til 6, der 1 er laveste score (helt mislykket) og 6 er høyeste score (svært vellykket).

Modellen ble testet i pilotevalueringer på fire prosjekter i 2012⁴. Erfaringene var i hovedsak positive (Volden og Samset, 2013). Senere er ytterligere 16 prosjekter evaluert, hovedsakelig ved hjelp av eksterne evalueringsfaglige miljøer, men også ved å bruke egne forskere. Et første utkast til retningslinjer som angir nærmere hvordan kriteriene skal tolkes og hvordan score bør fastsettes ble utarbeidet (Concept, 2015). Dette for å sikre at evaluatorene vurderer kriteriene på samme måte og at like prosjekter vurderes likt. Resultatet av denne studien er en ytterligere konsolidering og forbedring av retningslinjene.

Leserveiledning

Rapporten er delt i tre deler.

Del 1 handler om hvor vellykkete de 20 prosjektene er, basert på hva evalueringene har funnet og kommet frem til. Kapittel 2 presenterer prosjektene og opplegget for studien. Kapitlene 3 - 5 presenterer funn fra evalueringene – først resultatene på et overordnet nivå som gjelder henholdsvis operasjonell og taktisk og strategisk vellykkethet (kapittel 3), og deretter mer spesifikt det som gjelder resultatene for hvert evalueringskriterium (kapittel 4). Kapittel 5 konkluderer.

Del 2 drøfter erfaringene med evalueringene og godheten av disse, som bakgrunn for retningslinjer og praksis for dette arbeidet i årene som kommer.

⁴ Disse var Svinesund kontrollområde, dobbeltspor Asker-Sandvika, E 18 Momarken-Sekkelsten og Skjold-klasse MTBer til Sjøforsvaret, senere omtalt som «pilotevalueringer» (Volden og Samset, 2013).

Kapitlene 6 - 8 drøfter bruken av evalueringsmodellen. Kapittel 6 i forhold til erfaringer fra evalueringslitteraturen, mens kapittel 7 presenterer og drøfter erfaringer med å anvende modellen i de 20 evalueringene. Kapittel 8 konkluderer og gir anbefalinger for evalueringsvirksomheten videre.

Del 3 gir en kort sammenfatning av prosjektene og evaluators vurderinger i hver av de 20 evalueringsrapportene.

Del 1 Evaluering av de 20 første KS-prosjektene

2. Prosjektene og opplegget for denne studien

Utvalget av prosjekter

Det som etter hvert skal evalueres er noen av landets største statlige investeringsprosjekter i sektorene transport, bygg, forsvar og IKT. Det er mange av dem, men utviklings- og gjennomføringsperioden er lang, og per i dag er bare noen få kommet så langt at de kan evalueres. Av disse er 20 evaluert og omfattes av denne studien. De øvrige vil bli evaluert fortløpende. Situasjonen er beskrevet i tabell 2.1

Tabell 2.1 Prosjekter som har vært kvalitetssikret under Statens ordning. Status per februar 2017.

Kvalitetssikrede og evaluerte prosjekter	Antall kvalitets-sikret	Hvorav ferdigstilt	Evaluerbare*	Evaluert
Totalt antall KS-prosjekter per sept. 2016	252	92	40	20
Prosjekter som bare har vært gjennom KS2	177	92	40	20
Prosjekter som bare har vært gjennom KS1	93	0	0	0
Prosjekter som har vært gjennom KS1 og KS2	22	0	0	0

*Prosjekter som er kommet 5 år inn i driftsfase, dvs. ferdigstilt i 2011 eller tidligere

Den statlige ordningen for kvalitetssikring, startet som nevnt i år 2000 og per februar 2017 hadde den omfattet 252 prosjekter. Dette gjelder bare prosjekter med en estimert kostnadsramme høyere enn 750 millioner kroner og det dreier seg om i størrelsesorden 20 nye hvert år.⁵ Tidsspennet for de enkelte prosjektene fra utredningen startet til de er gjennomført er mange år, i de aller fleste tilfellene flere regjeringsperioder. De første prosjektene frem til 2006 ble ikke berørt av KS1 (dvs. kvalitetssikring av konseptvalget) men bare av KS2

⁵ Frem til 2011 var terskelverdien 500 mill. kroner. En del av de prosjektene som er etterevaluert så langt er derfor noe mindre enn dagens terskelverdi tilsier, men det er fortsatt de største prosjektene i landet. Unntaksvis er også mindre prosjekter inkludert.

(kostnader og styringsgrunnlag). Etter 2006 skal alle prosjektene gjennom både KS1 og KS2.

Tabellen viser at bare 93 prosjekter har gjennomgått KS1 så langt. Mange av disse har ennå ikke kommet så langt som til KS2, slik at det totale antallet som har vært gjennom begge deler bare er 22. Imidlertid er ingen av disse ferdigstilt så langt.

Det totale antall prosjekter som er ferdigstilt er ifølge tabellen 92. Av disse har Concept-programmet samlet inn sluttkostnader for 79 prosjekter, se vedlegg 1 for en oversikt over disse⁶. Disse har imidlertid bare vært gjennom KS2.

Ettrevaluering kan tidligst skje når de har kommet noen år inn i driftsfasen. Dette for å kunne registrere hvilke effekter, virkninger og nytte osv. de har hatt i samfunnet. Per i dag har 40 av prosjektene (dvs. veiene, jernbanestrekningene, bygningene osv.) vært i drift i minst fem år.

Denne studien omfatter halvparten disse, og vi mener at de gir et rimelig representativt bilde av situasjonen for de prosjektene som kan evalueres per i dag. I snitt har prosjektene vært vel 6 år i drift, med en variasjon fra 11 år i tilfellet E6 Åsgård-Halmstad, til MTB Skjold hvor det siste fartøyet ennå ikke var kommet i drift på evalueringstidspunktet. Man kan ikke snakke om et utvalg med de begrensningene det gir for representativitet, men vi har sørget for god spredning på sektorer, inkludert både prosjekter som er kjent som vellykkede og slike som har kjente problemer, og for øvrig forsøkt å ta med prosjektene kronologisk etter når de var ferdigstilt.

Det er åtte veiprosjekter, fem byggeprosjekter, tre jernbaneprosjekter, to IKT- og to forsvarsprosjekter, se tabell 2.2. Dette gjenspeiler at de to første sektorene er representert med flest prosjekter innenfor KS-ordningen. Som vist i vedlegg 1 fordeler de ferdigstilte prosjektene seg med 53 % veiprosjekter, 18 % bygg, 9 % jernbane, 11 % IKT, og 9 % forsvarsprosjekter. Vårt utvalg samsvarer altså rimelig godt med dette.

⁶ Se også Welde (2017) for mer informasjon om årsaker til at sluttkostnad ikke foreligger i 13 prosjekter.

Tabell 2.2 Nøkkelinformasjon. Prosjektene er nummerert etter evalueringsåret

Nr	Prosjekt	Sektor	Start (år)	Slutt (år)	Eval. (år)	Ansvar. eval.*	Forskere/eksterne
1	Kontrollområde Svinesund	Bygg	2004	2005	2012	SINTEF	E/F
2	Asker-Sandvika	Bane	2001	2005	2012	VTI	E
3	E18 Momarken-Sekkelsten	Vei	2005	2007	2012	Concept	F
4	Skjold-klasse MTB-er	Forsva	2003	2013	2012	Scantea	E/F
5	E6 Riksgr.-Svingenskogen	Vei	2002	2004	2014	COWI	E
6	Svalbard forskningspark	Bygg	2003	2005	2014	Concept	F
7	Lofoten fastlandsforb.	Vei	2003	2007	2014	UiN	E
8	Eiksundsambandet	Vei	2003	2008	2014	Menon	E
9	NAV IKT Basis	IKT	2006	2010	2014	NIBR	E
10	Høgskolesenteret Østfold	Bygg	2003	2006	2015	SINTEF	E/F
11	E16 Kløfta-Nybakk	Vei	2005	2007	2015	Urbanet	E
12	Rv 519 Finnfast	Vei	2005	2007	2015	Menon	E
13	Sandnes-Stavanger	Bane	2005	2009	2015	Oslo Ec.	E
14	Regionfelt Østlandet	Forsva	2002	2012	2015	Prokons	E
15	Perform	IKT	2008	2012	2015	Menon	E
16	Halden fengsel	Bygg	2006	2010	2016	Oslo Ec.	E/F
17	Nytt operahus Oslo	Bygg	2005	2008	2016	HRCCons	E
18	E6 Svingenskogen-Åsgård	Vei	2005	2008	2016	Menon	E/F
19	E6 Åsgard-Halmstad	Vei	2004	2005	2016	Menon	E/F
20	Gevingåsen	Bane	2009	2011	2017	Concept	F

*I noen tilfeller er evalueringen gjennomført av et konsortium, se rapportens del 3 for mer fullstendige opplysninger

Opplegget for evaluering

Evaluering av KS-prosjektene er en samfunnsoppgave som i prinsippet skal tilligge de ansvarlige departementene på hvert sitt ansvarsområde.

Finansdepartementet har imidlertid ønsket at ansvar for metodevalg, tilrettelegging, kontrahering og formidling av resultatene er delegert til NTNU ved forskningsprogrammet Concept, som har en særskilt interesse av å benytte funnene i sin forskning.

Evalueringene skal gjennomføres av eksterne, uavhengige miljøer eller Concept-programmets forskere. De første fire evalueringene var piloter der uttesting av metodikken stod sentralt. Resultatene fra disse er beskrevet i Concept-rapport nr. 30, som sammenfatter erfaringer og råd for fremtidig evalueringsvirksomhet, se Volden og Samset (2013). De resterende 16 er gjennomført i perioden 2014 – 2017. I alt 17 konsulentselskaper eller forskningsinstitutter har vært involvert, i tillegg til Concepts egne forskere. Tre evalueringer ble gjennomført utelukkende av Concepts egne forskere, og i seks tilfeller var det med en Concept-forsker i eksterne team. Begrunnelsen for dette er forskernes behov for uttesting av metodikken og intern læring.

Evalueringsprosessen har foregått i følgende trinn:

1. Concept tildeles årlige budsjett med et tilhørende krav om antall prosjekter som skal evalueres det året⁷. Forskerne velger selv hvilke prosjekter som skal evalueres, basert på at de må være et antall år inne i driftsfasen, samt hensynet til spredning på sektorer og en rimelig representativ fordeling. Valget avklares med Concepts kontaktperson i det ansvarlige fagdepartementet, som i prinsippet kan si nei, men det har ikke skjedd. En ber også om en kontaktperson i den underliggende etaten som har hatt ansvaret for prosjektet.
2. Det etableres et evalueringsteam. Prinsippet har vært at alle involverte skal være upartiske og uavhengige av prosjektet, og teamet skal ha både god evalueringskompetanse og fag- og sektorkompetanse. I de fleste tilfellene har tildeling av oppdrag skjedd etter offentlig utlysning, eller invitasjon til et antall potensielle evaluatore. Omfanget har vært i størrelsesorden 3 månedesverk, noe avhengig av prosjektets kompleksitet og eksisterende datagrunnlag, og tidsrammen har typisk vært satt til rundt 6 mnd.
3. Evalueringsteamet kontrollerer og justerer ved behov prosjektets målstruktur. Deretter operasjonaliseres de seks evalueringskriteriene i mer konkrete forhold som skal undersøkes, såkalte evalueringsspørsmål eller indikatorer. Teamet står fritt til dette, men blir oppfordret til å triangulere mellom kvalitativ og kvantitativ informasjon. Concept får oversendt oversikten over indikatorer og tilhørende plan for datainnsamling og

⁷ Programmet får såkalte gjennomføringsbrev (tilsvarende tildelingsbrev) fra Finansdepartementet, hvor et av prestasjonskravene er knyttet til etterevalueringer. For eksempel var kravet for 2016 «*Oppstart av minst tre etterevalueringer av store statlige investeringsprosjekter som er ferdigstilt og kommet i driftsfasen.*» Gjennomføringsbrevene er offentlig tilgjengelige og kan lastes ned fra Concept-programmets nettside.

analyse for eventuelle kommentarer og faglige innspill, men kan ikke overstyre evaluators opplegg.

4. Deretter gjennomføres evalueringen, ved at data samles inn og analyseres, og evalueringsspørsmålene besvares. Concept bistår med tilgang til sentrale dokumenter som finnes i programmets database. Evaluator står for øvrig ansvarlig for å samle inn supplerende data. I praksis er man avhengig av noe tilrettelegging fra kontaktpersonen i etaten til å skaffe tilgang til relevante interne data.
5. Teamet gjør til slutt en samlet vurdering av prosjektets vellykkethet for hvert av de seks overordnede kriteriene. Evaluator skal oppsummere sin vurdering for hvert kriterium ved å sette en score mellom 1 og 6 hvor 1 er helt mislykket og 6 er svært vellykket.
6. Resultatene oppsummeres i en rapport. Utkast til evalueringsrapport sendes både Concept og kontaktpersonene i fagdepartement og etat for eventuelle kommentarer, fortrinnsvis før man setter tallscore (da det ellers er lett at leseren henger seg opp i dette, som skal være evalueringsteamets uavhengige vurdering).
7. Rapporten publiseres på nettsiden til Concept i et eget interaktivt program hvor dokumentet også kan lastes ned, samt omtales i forskningsprogrammets nyhetsbrev. En oppfordrer evaluator til å spre informasjon om resultatene, for eksempel ved å skrive faglige artikler. Departementet og etaten oppfordres til å følge opp resultatene internt. Concepts forskere har ikke hatt andre faste rutiner for oppfølging, ut over å være tilgjengelig for å gi en presentasjon eller diskutere resultatene med departement eller etat hvis ønskelig.

Metaevalueringen

OECD definerer en metaevaluering som følger:

The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators.

(OECD, 2000. Glossary of evaluation and results based management)

Begrepet har altså to betydninger, og i denne studien bruker vi begge: vi foretar først en oppsummering og aggregering av funnene fra de 20 evalueringene (rapportens del 1), og deretter gjør vi en vurdering av evalueringene selv, med

fokus på hvordan de har anvendt evalueringsmodellen – det omfatter også en vurdering av evalueringskonseptets egnethet for oppgaven (del 2).

Aggregering av funn

Kapittel 3 sammenstiller funnene fra de 20 evalueringene, med tanke på å kunne trekke noen foreløpige konklusjoner om norske KS-prosjekters vellykkethet. Vi undersøker om det er systematiske forskjeller mellom sektorer med hensyn på score, og om det er samvariasjon mellom resultatene for de seks overordnede kriteriene. I tillegg til å se på tallscore, som er øverste aggregeringsnivå, gjennomgår vi også rapportene basert på en sjekkliste med mer detaljerte spørsmål og generiske indikatorer på suksess [vedlegg 2]. Analysen består i å telle opp antall prosjekter som ifølge evalueringene kommer godt eller dårlig ut på disse indikatorene. Resultatene er selvfølgelig basert på en god del skjønn. Vi tar også forbehold om at antallet prosjekter i dette tilfellet er lavt.

For å illustrere suksess på et enda mer aggregert nivå – henholdsvis *operasjonell* og *taktisk og strategisk* suksess – kan en definere det første ved score for «produktivitet» og det siste som et ikke veid gjennomsnitt av score for de fem resterende kriteriene, se boksen under. For å illustrere resultatene visuelt har vi også brukt trafikklys hvor fargene rød, gul og grønn dekker hver sin tredjedel av måleskalaen.⁸

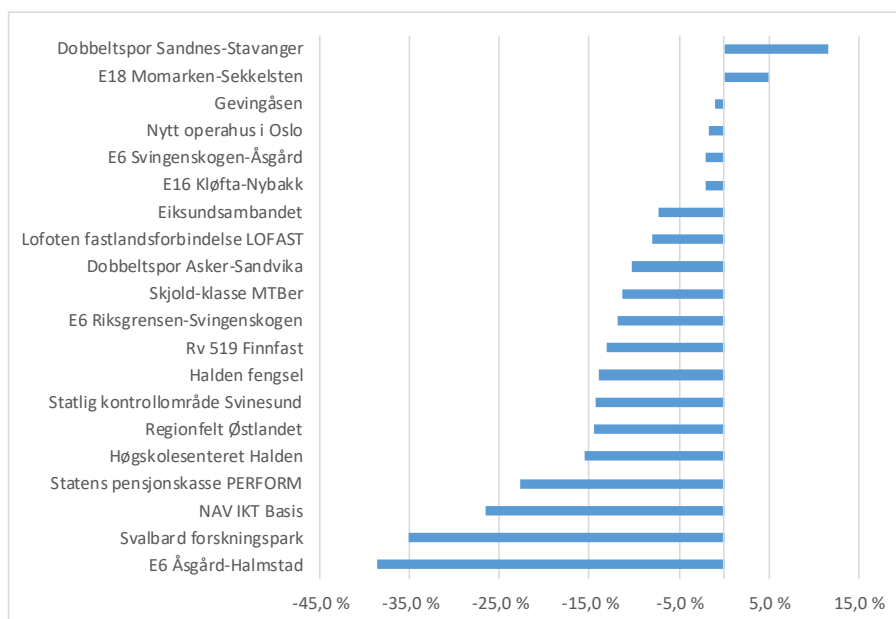
Operasjonell suksess = Score for «produktivitet»

Taktisk og strategisk suksess = $\sum \frac{1}{5} (\text{Score for «måloppnåelse»} + \text{«andre virkninger»} + \text{«relevans»} + \text{«levedyktighet»} + \text{«samf.øk. effektivitet»})$

⁸ Score mellom 1 og 2,7 gir da rødt, score mellom 2,7 og 4,3 gir gult og score mellom 4,3 og 6 gir grønt.

3. Hvor vellykkete var de første 20 KS-prosjektene?

De 20 prosjektene representerer en samlet investering på rundt 30 milliarder kroner i en periode på 16 år (fra den første Stortingsbevilgningen til alle prosjektene var gjennomført). Evalueringene konkluderer med at så mye som 18 prosjekter, eller 90 prosent ble ferdigstilt med en sluttkostnad som lå på eller under Stortingets kostnadsramme, se figur 3.1. De siste to hadde kostnadsoverskridelser. Resultatet er godt i forhold til hva en har sett i en rekke større internasjonale undersøkelser.



Figur 3.1 Prosjektene sluttkostnad i forhold til Stortingets kostnadsramme, målt i prosent av egen ramme (rammene er prisjustert)

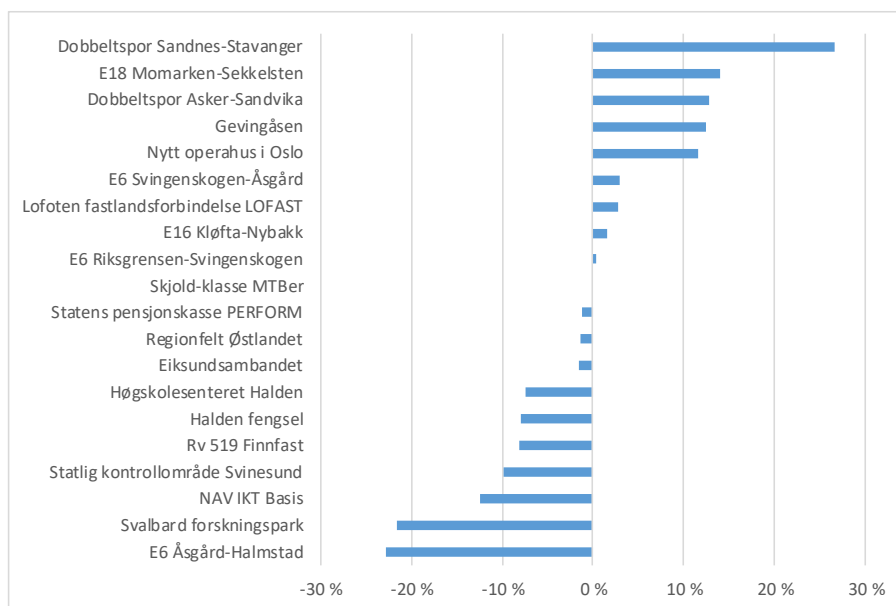
Det må imidlertid presiseres at det kan forekomme avvik mellom tallene i denne rapporten, som er basert på de 20 evalueringsrapportene, og tallene for de samme prosjektene i en undersøkelse av et større antall KS-prosjekter, Welde (2017). Det skyldes at evalueringsteamenes vurderinger av hvilken ramme de sammenligner sluttkostnaden med, ikke er helt konsistente. Vi kommer tilbake til dette i kapittel 7. Welde (2017) finner at av 78 avsluttede prosjekter hadde 60 en sluttkostnad på eller innenfor kostnadsrammen. Det

utgjør 77 prosent. Den samlede sluttkostnaden ble om lag 7,5 milliarder kroner lavere enn kostnadsrammen, tilsvarende mer enn 6 prosent av hele porteføljen.

En viktig begrepsavklaring: Stortingets vedtatte *kostnadsramme* er det som det ansvarlige departement har fullmakt til å disponere. Departementet på sin side gir utførende etat en noe innskrenket fullmakt som tilsvarende kostnadsrammen fratrasket en såkalt *usikkerhetsavsetning*. Dette kalles for *styringsrammen*. Dersom etaten står i fare for å overskride denne skal departementet varsles. På utøvende nivå gir etaten den prosjektansvarlige et *styringsmål* som er noe lavere enn styringsrammen, for å sikre forutsigbarhet. Praxis på dette området varierer mellom sektorer. Enkelte departementer delegerer fullmakten over hele eller deler av usikkerhetsavsetningen til etaten, men det er et mål for etaten å holde seg innenfor styringsrammen, se Andersen m.fl. (2016).

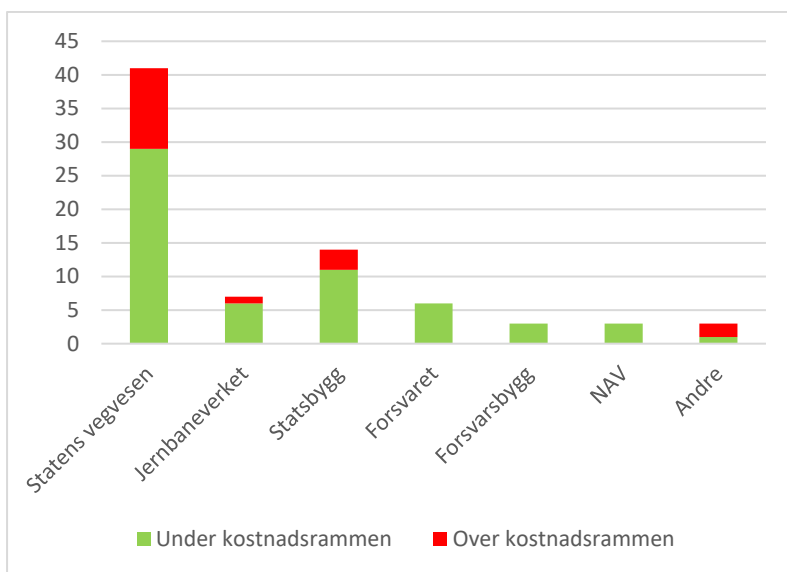
Figur 3.2 viser prosjektenes sluttkostnad i forhold til *styringsrammen*. Den viser at 10 prosjekter kommer under og 10 over styringsrammen. Ettersom fordelingen er nær symmetrisk betyr det at summen for hele porteføljen går omtrent i null. Eller sagt på en annen måte, at porteføljen av prosjekter samlet sett holder seg innenfor den samlede styringsrammen. Det betyr at en ikke har brukt av den samlede usikkerhetsavsetningen, og altså ender med en samlet sluttkostnad som ligger godt under samlet kostnadsramme. Figuren viser imidlertid også at så mange som syv av porteføljens 11 transportprosjekter har kostnadsoverskridelser i forhold til styringsrammen. Spørsmålet som krever en avklaring er da om styringsrammene er for snevre eller om usikkerheten i gjennomføringen er for høy. Eller om det er andre forhold som gir systematiske avvik fra det forventede. Welde (2017) drøfter noen mulige årsaker til at spredningen i en del tilfeller er estimert for lavt i forkant, spesielt i veiprojekter. En peker blant annet på at det er uheldig å operere med krav til kalkylenøyaktighet slik Statens vegvesen tradisjonelt har gjort⁹. Det kan føre til at usikkerheten underestimeres for å komme innenfor kravet.

⁹ Statens vegvesen har operert med mål for kalkylenøyaktighet, og dette har gjerne blitt oppfattet som krav. Målet er at estimatet skal ha minimum 70 prosent sannsynlighet for å ligge innenfor det intervallet som bestemmes av nøyaktighetsgrensene – og for reguleringsplannivået er det +/- 10 prosent standardavvik som gjelder (høyere for tidligere planfaser). Som påpekt i Welde (2017) kan det tenkes at 10 % har vært urealistisk lavt og ført til at gjenværende usikkerhet i virkeligheten er høyere enn det anslagene gir uttrykk for.



Figur 3.2 Prosjektene sluttkostnad i forhold til etatenes styringsramme, målt i prosent av egen ramme (rammene er prisjustert)

De fire prosjektene som kommer best ut i forhold til kostnadsrammen har en innsparing på mellom 20 og 40 prosent. Spørsmålet blir da om kostnadsrammen var for romslig. To av disse prosjektene er IKT-prosjekter, der forklaringen kan være at det erfaringsmessig er vanskelig å forutsi omfanget av slike. De to andre er et byggeprosjekt på Svalbard og et veiprojekt i Østfold som kom ut mellom 30 og 40 prosent under kostnadsrammen. Det er lett å forstå at forholdene på Svalbard er slik at det er vanskelig å forutsi kostnadene. Her viste evalueringen at man hadde lagt inn en romslig usikkerhetsavsetning for å sikre seg mot problemer knyttet til klima og vanskelig logistikk. Prosjektet med størst innsparing gjelder byggingen av motorvei på en strekning mellom svenskegrensen og Oslo i et landskap som er relativt lite kupert. Gitt etatens betydelige erfaring med slike prosjekter ville det være rimelig å anta at det store avviket skyldes selve kostnadsestimeringen. I følge evalueringen var forklaringen dels at grunnforholdene viste seg å være bedre enn antatt, dels at selve prosjektgjennomføringen var svært vellykket. I tillegg har markedssituasjonen nesten alltid stor innflytelse på omfanget av under- eller overskridelser. I det aktuelle veiprojektet var markedsforholdene svært gunstige.



Figur 3.3. Antall prosjekter med overskridelse/innsparing i forhold til kostnadsrammen, fordelt på etat (av totalt 78 ferdigstilte prosjekter). Kilde: Welde (2017)

Begge de to prosjektene som hadde overskridelser i forhold til kostnadsrammen var transportprosjekter, ett vei- og ett jernbaneprojekt. Og som nevnt over har et flertall av veiprosjektene overskridelse av styringsrammen. Den mer omfattende undersøkelsen (Welde, 2017), illustrert i figur 3.3, fant at nesten 30 prosent av veiprosjektene endte med overskridelse.

Operasjonell versus taktisk og strategisk vellykkethet

Operasjonell vellykkethet bygger i sin helhet på vurderingen av gjennomføringen av selve prosjektet (i tid, kostnad, kostnadseffektivitet, kvalitet, etc.). Taktisk og strategisk vellykkethet er resultatet av den samlede vurderingen av hvordan en scorer på de fem andre evalueringskriteriene etter at prosjektet er avsluttet. Tabell 3.1 viser resultatene for de 20 prosjektene, som gjennomsnittlig scorer noe høyere operasjonelt enn taktisk-strategisk:

- Mer enn halvparten av prosjektene scorer høyt (grønt trafikklys) både operasjonelt og taktisk-strategisk.
- Et klart flertall av prosjektene (13 av 20) får grønt lys på det operasjonelle. Dette til tross for at det forekommer kostnadsoverskridelser som vist i forrige avsnitt, til dels store overskridelser av styringsrammen i to

prosjekter (henholdsvis vei og jernbane). I begge prosjektene kommer imidlertid evaluator til at gjennomføringen samlet sett var akseptabel, i det ene tilfellet antydes det at rammen var undervurdert til tross for at den var kvalitetssikret.

Tabell 3.1. Prosjektene operasjonelle og taktiske-strategiske vellykkethet, rangert etter sistnevnte (Sektor: V=vei, J=jernbane, F = forsvar, B = bygg)

Prosjekt	Sektor	Operasjonell suksess	Taktisk og strategisk suksess
E6 Riksgrensen-Svingensbogen	V	5	5,3
Rv 519 Finnfast	V	5	5,2
Regionfelt Østlandet	F	5	5,2
Statens pensjonskasse PERFORM	IKT	5	5,2
Eiksundsambandet	V	5	5,0
Nytt operahus i Oslo	B	5	4,8
E6 Åsgård-Halmstad	V	5	4,8
Statlig kontrollområde Svinestund	B	6	4,8
E18 Momarken-Sekkelsten	V	3	4,7
Svalbard forskningspark	B	6	4,7
E6 Svingensbogen-Åsgård	V	4	4,7
Halden fengsel	B	5	4,5
NAV IKT Basis	IKT	5	4,3
Dobbeltspor Asker-Sandvika	J	5	4,2
Høgskolesenteret Halden	B	5	4,2
E16 Kløfta-Nybakk	V	4	4,2
Dobbeltspor Sandnes-Stavanger	J	4	4,2
Lofoten fastlandsforbindelse LOFAST	V	4	3,8
Gevingsåsen	J	4	3,2
Skjold-klasse MTBer	F	4	3,2

- Byggsektoren kommer jevnt over godt ut. Fire av de fem evaluerte prosjektene får grønt lys både operasjonelt og taktisk-strategisk
- Transportsektoren kommer dårligere ut, der syv av elleve prosjekter scorer lavere enn det som tilsvarer grønt trafikklis, enten taktisk-strategisk, operasjonelt eller begge deler

Tabell 3.2 Samlet oversikt over hvordan prosjektene scorer på taktiske og strategiske evalueringskriterier

	Prosjekt	Sektor	Mål- oppnåelse	Andre virkninger	Relevans	Leve- dyktighet	Samf.øk. effektivitet
E6 Riksgrensen-Svingenskogen	V		6	5	5	5	6
Rv 519 Finnfast	V		5	5	5	5	6
Regionfelt Østlandet	F		5	5	6	5	5
Statens pensjonskasse PERFORM	IKT		5	5	5	6	5
Eiksundsambandet	V		6	5	4	5	5
Nytt operahus i Oslo	B		5	6	5	4	4
E6 Åsgård-Halmstad	V		5	4	5	4	6
Statlig kontrollområde Svinesund	B		4	4	5	5	
E18 Momarken-Sekkelsten	V		5	4	5	5	6
Svalbard forskningspark	B		4	5	5	5	3
E6 Svingenskogen-Åsgård	V		5	4	5	4	6
Halden fengsel	B		4	5	4	5	4
NAV IKT Basis	IKT		6	4	3	5	3
Dobbeltspor Asker-Sandvika	J		3	4	5	5	3
Høgskolesenteret Halden	B		4	3	4	5	4
E16 Kløfta-Nybakk	V		5	3	5	4	4
Dobbeltspor Sandnes-Stavanger	J		4	4	5	5	3
Lofoten fastlandsforbindelse LOFAST	V		5	4	3	4	3
Gjevingåsen jernbanetunnel	J		3	4	4	4	2
Skjold-klasse MTBer	F		4	4	3	2	2

Når det gjelder den overordnede vurderingen av prosjektene, det vil si hvordan de lykkes taktisk og strategisk viser evalueringsresultatene at:

- Hele 85 prosent av prosjektene har en samlet score på fire eller bedre, mens bare tre prosjekter har lavere score (Lofast, Gevingåsen og Skjold)
- Hele 60 prosent ligger innenfor øverste tredjedelen av skalaen (> 4,3) som vi anser som vellykket, dvs. grønn farge.
- Resultatene viser at 65 prosent av prosjektene (13 av 20) får grønt lys på *relevans*, og det samme for *levedyktighet*, mens nesten like mange (12) får grønt på *måloppnåelse*. Dette betyr at det er stor enighet om at de fleste investeringene er nyttige i den forstand at de er i samsvar med prioriteringer og behov i samfunnet. Videre at en har oppnådd det en

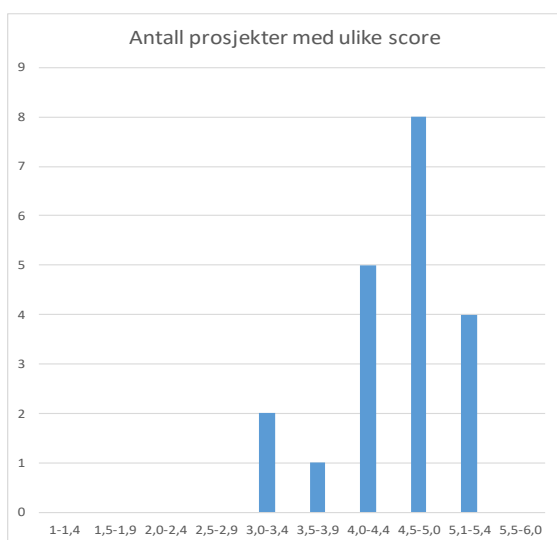
forventet, og at det er god grunn til å tro at de positive effektene av tiltaket vil vedvare over tid.

- Prosjektene scorer noe lavere på *samfunnsøkonomisk effektivitet*. Ni av 20 prosjekter får grønt lys, noe som betyr at de vurderes positivt som lønnsomme sett i et større samfunnsperspektiv. De fem mest lønnsomme er alle veiprosjekter. De som vurderes minst lønnsomme er tre jernbaneprosjekter, ett forsvarsprosjekt hvor det er uenighet om behovet, signalbygget på Svalbard, og fastlandsforbindelsen til Lofoten som konkurrerer med flere alternative ruter i et tynt befolket område.
- Det er åtte prosjekter som får grønt lys når det gjelder prosjektenes *virksomheter* ut over det som gjelder den planlagte måloppnåelsen. Lav score kan skyldes at en ser negative konsekvenser av investeringstiltaket, eller at en ikke ser mange positive ringvirkninger. Det prosjektet som scorer høyest på andre virkninger er Nytt operabygg i Oslo, hvor evaluator ser positive virkninger for byutviklingen samt en symboleffekt som går ut over de kulturpolitiske målene for prosjektet. Ikke alle vil være enige i vurderingen. Et annet eksempel på prosjekter med positive virkninger er forskningsbygget på Svalbard som er blitt et signalbygg for norsk tilstedeværelse, og et samlingspunkt for ny virksomhet og turisme på stedet.

Ser man nærmere på sammenhengen mellom hvordan prosjektene scorer på de fem taktisk-strategiske evalueringskriteriene kan vi notere følgende observasjoner:

- Det later til å være god samvariasjon mellom relevans og levedyktighet, som forventet. Det betyr at dersom det er behov for noe, kan vi også regne med at det blir satt pris på og ivaretatt over tid.
- Det er også samsvar mellom måloppnåelse og lønnsomhet. Der en måler god måloppnåelse er altså netto betalingsvilligheten også høy. Det må tas forbehold om at det er mange veiprosjekter i utvalget, disse har ofte effektmål som handler om tidsbesparelser, noe som verdsettes høyt i en samfunnsøkonomisk analyse. I andre sektorer kan målene være knyttet til andre endringer som det ikke nødvendigvis er lett å måle betalingsvilligheten for.
- Det er også samvariasjon mellom relevans og samfunnsøkonomisk effektivitet, men kanskje ikke så stor som en kunne forvente. Dette kan tas som en indikasjon på at det ikke er markedet som avgjør i disse offentlig initierte prosjektene, men andre hensyn, slik de kommer til uttrykk gjennom politiske aktører og prosesser.

- Det kan også være en indikasjon på pengeriklighetens konsekvenser, politikere som overbyr hverandre i å bruke penger for ulike, ikke alltid lønnsomme formål.
- Sortert etter sektor scorer byggsektoren høyt på produktivitet og levedyktighet, men svakere på lønnsomhet og måloppnåelse. Veisektoren scorer (ikke uventet) høyt på måloppnåelse og lønnsomhet. Forsvarets to prosjekter er både det dårligste og et av de beste, og tilsvarende er det vanskelig å trekke noen systematikk ut av de to IKT-prosjektene. Jernbaneprosjektene scorer lavt på samfunnsøkonomisk effektivitet men høyt på relevans og levedyktighet.

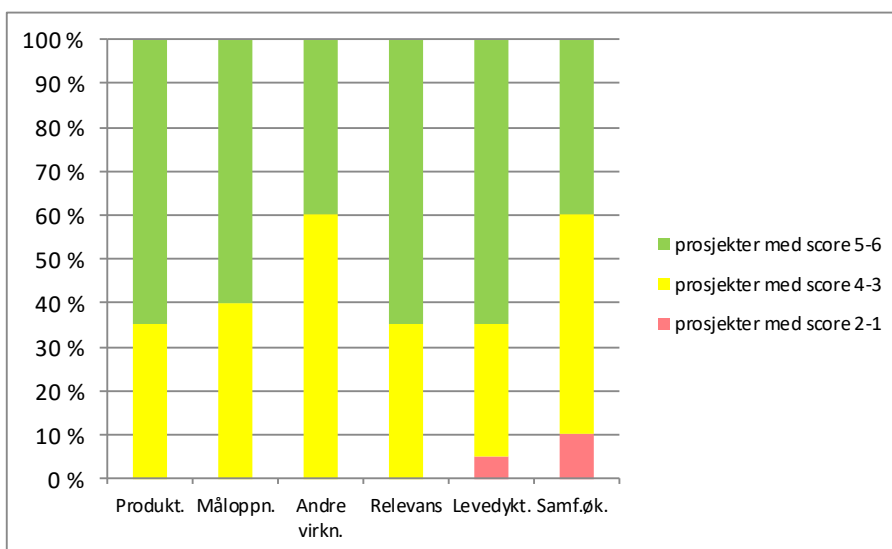


Figur 3.4. Antall prosjekter med ulik gjennomsnittsscore. Positiv skjeving, eller svært vellykkete prosjekter?

Det er her på sin plass å igjen påpeke at evalueringene gjelder prosjekter som bare har vært gjennom KS2. De er sannsynligvis ikke representative for senere prosjekter hvor også selve konseptvalget har vært grundig utredet etter Statens prosedyre og deretter kvalitetssikret eksternt. Det er grunn til å anta at en vil se et løft i strategisk ytelse for senere prosjekter som også har vært gjennom KS1. Evalueringene i denne studien vil være nyttige som referansemateriale for senere evalueringer. Dette forutsetter imidlertid at vurderingen av det taktisk-strategiske er sammenlignbar (herunder like streng) der referansegrunnlaget er KS1 og der referansegrunnlaget er svakt eller mangler som i disse 20 evalueringene (mer om dette i kapittel 7).

4. Evalueringsresultatene

Tabell 3.1 i foregående kapittel viste eksepsjonelt god samvariasjon mellom hvordan prosjektene scorer på produktivitet og taktisk-strategisk suksess - i den grad at man kan være fristet til å si at evalueringen av det strategiske og samfunnsmessige ikke er nødvendig så lenge produktiviteten i gjennomføringen er avklart. Men dette ville være en meningsløs generalisering, for det er selvsagt helt forskjellige faktorer som påvirker gjennomføringen av prosjektet, og hva som avgjør de samfunnsmessige konsekvensene av det. Her dreier det seg både i prinsippet og i praksis om å evaluere to helt forskjellige men koplete fenomener (ikke nødvendigvis sterkt koplete), det vil si *selve byggeprosjektet* på den ene siden og *investeringstiltaket* på den andre.

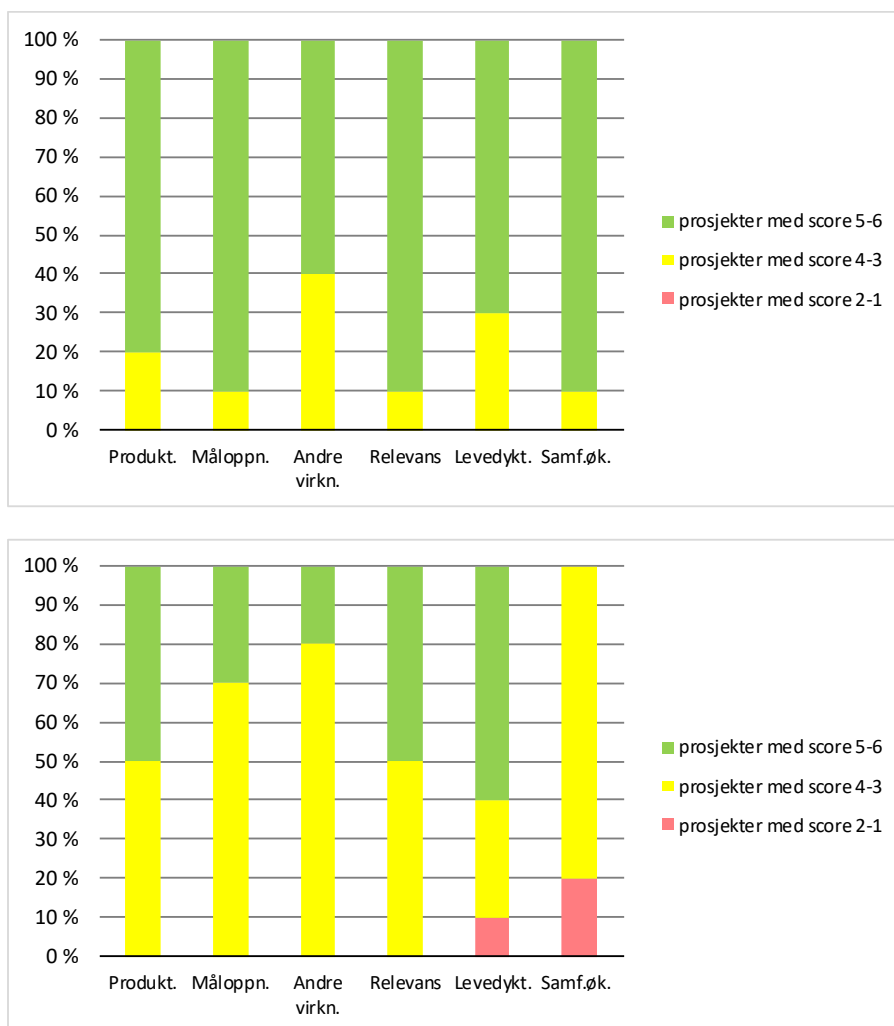


Figur 4.1 Sammenstilling av hvor mange prosjekter som scorer innenfor tre nivåer av vellykkethet på de seks evalueringskriteriene

Figur 4.1 viser en aggregert fremstilling av hvordan prosjektene scorer på de seks evalueringskriteriene. Gruppen med prosjekter kommer sterkest ut når det gjelder produktivitet og måloppnåelse, men også relativt godt på relevans og levedyktighet, det vil si at de både er godt gjennomført og at det er behov for

tiltakene i samfunnet. Det bør ikke overraske noen at prosjekter som er resultatet av en langvarig faglig og politisk prosess kommer ut som nyttige tiltak og at de også er godt gjennomført. Vi kan likevel ikke se bort fra at scoresettingen kan være positivt skjevet, noe som diskuteres nærmere i kapittel 7.

I det etterfølgende har vi sett nærmere på hva som karakteriserer de ti minst vellykkete prosjektene i forhold til de ti mest vellykkete. Vellykkethet er her et uttrykk for uveid gjennomsnittscore for de seks evalueringskriteriene.



Figur 4.2 Sammenstilling av hvor mange prosjekter som scorer innenfor tre nivåer av vellykkethet på de seks evalueringskriteriene, for henholdsvis de ti mest vellykkete (øverst) og de ti minst vellykkete prosjektene (nederst)

Prosjektene fordeler seg som følger:

- De ti mest vellykkede prosjektene er: E6 Riksgrensen-Svingenskogen, Rv 519 Finnfast, PERFORM, Regionfelt Østlandet, Eiksundsambandet, Statlig kontrollområde Svinesund, E6 Åsgård-Halmstad, Nytt operahus Oslo, E18 Momarken-Sekkelsten og E6 Svingenskogen-Åsgård
- De ti minst vellykkede prosjektene er: Svalbard forskningspark, Halden fengsel, NAV IKT Basis, E16 Kløfte-Nybakk, Dobbeltspor Sandnes-Stavanger, Dobbeltspor Asker-Sandvika, Høgskolen i Østfold, LOFAST, Skjold klasse MTB-er og Gevingåsen jernbanetunnel.

Fargene som følger av scoresettingen for de to gruppene hver for seg er vist i figuren, de ti mest vellykkede øverst og de ti minst vellykkede nederst.

En overordnet vurdering av prosjektene er vist i tabell 4.1. Det mest iøynefallende er at det i liten grad er snakk om kontroversielle prosjekter. Det gjelder alle de mest vellykkete, men også syv av de ti minst vellykkete. Enda viktigere er det at evalueringen finner at det er et klart behov for prosjektene i samfunnet, det gjelder også for halvparten av de minst vellykkete prosjektene. En interessant observasjon er at de mest vellykkete i liten grad var rettet mot en bestemt interessegruppe, i noe større grad gjaldt det de mindre vellykkete prosjektene. Dette bidrar til å bekrefte en tendens man har sett i andre prosjekter, nemlig at der sterke interesser lykkes i å få finansiert prosjekter med offentlige midler, kan det i noen tilfeller vise seg at nytten ikke står i forhold til kostnadene. Dette fenomenet kan i enkelte tilfeller forklares med begrepet *perverse insentiver*. Det kan oppstå der det ikke følger økonomiske eller andre forpliktelser for mottakeren av en ytelse fra det offentlige. Insentivet til å vurdere kostnadene opp mot den potensielle samfunnsnyten er dermed svekket. Fenomenet er grundig drøftet i en tidligere Concept-rapport (Samset m.fl., 2014).

Tabell 4.1 Forklaringsvariable som gjelder den overordnede vurderingen av prosjektene

Generelt	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Vurdert som vellykket	70	10	4
Var tiltaket ukontroversielt	85	10	7
Var det et klart behov	65	8	5
Gjaldt det en bestemt interessegruppe	30	2	4

Produktivitet

Produktiviteten gjelder gjennomføringen av prosjektet. Tabell 4.2 viser at 90 prosent av prosjektene ble gjennomført innenfor kostnadsrammen. Men også at bare omtrent halvparten av prosjektene hadde realistiske kostnadsrammer. I fem prosjekter mente man at kostnadsrammen var for høy og i fire prosjekter for lav. Dette reiser et viktig spørsmål som en må se nærmere på i fremtidige evalueringer: at sluttkostnaden ikke bare skal vurderes i forhold til vedtatt budsjett, men at en må se nærmere på rimeligheten av dette.

Tabell 4.2 Forklaringsvariable som gjelder vurderingen av prosjektenes produktivitet

Produktivitet	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Vurdert med høy produktivitet	65	8	4
Holdt styringsrammen	50	7	3
Holdt kostnadsrammen	90	9	9
Realistisk kostnadsramme	45	4	5
Ble det ferdigstilt innen tidsmålet	75	7	8
Ble kvaliteten vurdert som god	80	8	8

Samtidig som prosjektene scorer høyt på produktivitet, noterer vi oss at det i minst ett av prosjektene har vært foretatt omfattende endringer og tilpasninger i de første driftsårene, delvis på grunn av at det var valgt for enkle/billige løsninger i gjennomføringen. I et annet prosjekt ble det gjort en forenkling som potensielt medfører lavere nytte på lang sikt (en tapt realopsjon). Generelt er det viktig at en trekker inn et levetidsperspektiv også når en evaluerer prosjektgjennomføringen. En annen bemerkning gjelder tre av samferdselsprosjektene, som kun var parseller i en større utbygging. Selv om det enkelte prosjekt er effektivt gjennomført, kan en hevde at denne type oppsplitting er lite kostnadseffektivt – det har i liten grad slått ut på scoresettingen her.

Et annet relevant spørsmål er hvorvidt prosjektene fulgte KS2-rapportens anbefalinger, som nettopp skal bidra til god gjennomføring (riktig kontraktstrategi, oppfølging av sentrale usikkerhetsfaktorer, og god organisering og styring mer generelt). Det later til å være stor variasjon, fra de prosjektene som har fulgt kvalitetssikrers råd i detalj, til de som overhodet ikke har forholdt seg til dem. De fleste evalueringsrapportene sier imidlertid lite om dette.

I tillegg kan man notere enkelte andre observasjoner som gjelder samvariasjon mellom variable:

- Sortert etter størrelse (kostnadsrammen) viser evalueringene at de største prosjektene kommer dårligst ut (unntak Regionfelt Østlandet spesielt mht. samfunnsøkonomisk effektivitet).
- Sortert etter overskridelse av kostnadsrammen viser evalueringene at IKT-prosjektene kommer godt ut (imidlertid har en her kun to prosjekter), og at veiprosjektene scorer gjennomgående lavt.
- Sortert etter overskridelse av styringsrammen viser evalueringene at de som holder seg innenfor styringsrammen også scorer høyt på de fleste andre parameterne.

Det er blant veiprosjektene man finner både de største overskridelsene og innsparingene, og det er kanskje grunn til å stille spørsmål med om dette kan ha sammenheng med tradisjonen med å sette mål for kalkylenøyaktighet i denne sektoren.

Måloppnåelse

Tabell 4.3 viser at de aller fleste prosjektene hadde god eller akseptabel måloppnåelse. Dette står i kontrast til evalueringens konklusjoner om at målene var realistiske i bare halvparten av prosjektene. Det skyldtes enten at målene var trivielle og ikke pekte videre enn mot prosjektets leveranse, eller at målene var urealistiske i forhold til hva en kunne forvente at prosjektene kunne føre til.

I veiprosjekter er dette mindre av et problem fordi prosjektene i stor grad er forholdsvis ensartete. Der handler det ofte om reisetid, reisekostnader og trafikkssikkerhet. I mer unike prosjekter er det vanskeligere å sammenligne måloppnåelse på tvers. Et gjennomgående problem viser seg å være at samfunnsmålet mangler i prosjektdokumentasjonen. Mange av prosjektene hadde heller ikke relevante effektmål. På dette området ligger det et forbedringspotensial både hos departementer og etater. Manglende eller urealistiske mål gjør at evaluator må vurdere helheten og legge til grunn et mål som er relevant og på et ambisjonsnivå som er realiserbart.

I syv av de ti mest vellykkete prosjektene mente evaluator at målene var realisert som forventet, mens bare i fire av de ti minst vellykkete prosjektene var dette tilfellet. Det mest tydelige skillet gjelder spørsmålet om hvorvidt prosjektet som sådan var det vesentligste bidraget til realiseringen av målet. Bare i tre av de ti minst vellykkete prosjektene var det tilfellet, mot åtte av ti for

de mest vellykkete. Dette er i seg selv en indikasjon på kompleksiteten i den samfunnsmessige settingen prosjektet inngår i. Det understreker dermed viktigheten av å se investeringstiltakene i et tilstrekkelig bredt perspektiv når de evalueres – det gjelder både ex post og ikke minst ex ante.

Spørsmålet om hvorvidt det hadde vært mulig å oppnå større effekter enn det man registrerte, ble bare i begrenset grad drøftet i evalueringene.

Tabell 4.3 Forklaringsvariable som gjelder vurderingen av prosjektenes måloppnåelse

Måloppnåelse	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Vurdert med høy måloppnåelse	60	7	4
Var målene realistiske	50	5	5
Var det nødvendig å endre målene	40	3	5
Var prosjektet et vesentlig bidrag til måloppnåelsen	55	8	3

Når det gjelder måloppnåelsen er sektorforskjellene nokså tydelige (med forbehold om at utvalget er lite). Veiprosjektene og de to IKT-prosjektene skiller seg positivt ut, mens både bygg og jernbane har svakere måloppnåelse. Samtidig merker vi oss at veiprosjektene var mer avgrensede og hadde mindre ambisiøse effektmål. Målformuleringene gjaldt effekter som vanligvis vil realiseres umiddelbart, så snart den nye veistandarden er på plass og veien åpnet. Bygge- og jernbaneprosjektene hadde mer ambisiøse effektmål som krever en større og mer integrert innsats, også ut over selve prosjektet. Jernbaneprosjektene skulle for eksempel gi økt kollektivandel, noe som forutsetter at operatørene bedrer tilbudet, og som også er avhengig av blant annet kvaliteten på konkurrerende tilbud av vei, parkeringsplasser, osv. De tre byggeprosjektene skulle alle gi faglige og/eller økonomiske synergier for brukerne av bygget, noe som antas å kreve mye mer enn kun et nytt bygg, herunder endringer i organisasjon og arbeidsprosesser. I både bygge- og jernbaneprosjektene var effektene begrensede selv etter flere års drift.

Det er mye man kan lære av dette. Ikke minst viktigheten av å tydelig plassere ansvaret for å ta ut de ulike effektene. I flere av byggeprosjektene lot målet om synergieffekter til å være «glemt» idet staten bevilget penger til nytt bygg. Ved tydelig plassering av ansvar for å ta ut gevinstene, og oppfølging av dette ut gjennom driftsfasen, kan en motvirke at dette skjer. Det er også viktig at en, der prosjektets måloppnåelse avhenger av mer enn prosjektet, faktisk tenker

tiltakspakker mer enn enkeltprosjekter allerede i planleggingsfasen. Herunder bør en styrke kompetansen om hvordan økonomiske, juridiske, informative og andre virkemidler og tiltak fungerer sammen med fysiske investeringstiltak.

Andre virkninger

Når det gjelder evaluering av prosjektenes andre virkninger, det vil si utover det som er tatt med under måloppnåelse, varierer vurderingene. Mange av prosjektene vurderes å ha få eller ingen utilsiktede virkninger av nevneverdig omfang.

Tabell 4.4 Forklaringsvariable som gjelder vurderingen av prosjektenes andre virkninger

Andre virkninger	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Betydelige virkninger utover måloppnåelsen	85	10	7
Var de fleste virkningene positive	85	10	7

Der andre virkninger beskrives er de oftere positive enn negative. Spesielt er evaluatorene opptatt av regionale økonomiske ringvirkninger, med særlig vekt på såkalt mernytte (produktivitetsgevinster) av transportinfrastruktur. Mernytte har blitt tillagt vekt i økende grad de senere år. Ofte kan det føre til dels svært optimistiske betraktninger om annen- og tredjeordens ringvirkninger som kan bidra til å rettferdiggjøre store investeringstiltak. Byggingen av fergefri E39 på Vestlandet er et eksempel. I evalueringene vi her ser på finner de fleste evaluatorene at mernytten er begrenset. Bare i tilfellet Eiksundsambandet er mernytten av en viss betydning, dette var da også en vesentlig hensikt med prosjektet som gir landfast tilknytning for et øysamfunn med en betydelig klyngedannelse innen skipsbygging og maritim service. Negative virkninger, herunder i form av naturinngrep, nevnes også i en del tilfeller. Flere av prosjektene later til å ha et forbedringspotensial når det gjelder å kartlegge sidevirkninger i forkant, og designe prosjektet med tanke på å maksimere de positive og minimere de negative.

Relevans

Tabell 4.5 viser at alle de 10 mest vellykkete prosjektene ble vurdert som relevante i forhold til prioriteringer og behov i samfunnet, men også åtte av de ti minst vellykkete. De fleste prosjektene i utvalget kan altså vises å være svar på reelle problemer som samfunnet mener at en må løse.

Samtidig ser en at bare i halvparten av prosjektene i den første gruppen forekom det vesentlige konflikter eller interessenmotsetninger, mot i hele åtte av ti tilfeller i de minst vellykkete prosjektene. Dette er som forventet idet interessenmotsetninger de facto betyr at relevansen er omstridt. Samtidig er ikke eksistensen av interessekonflikter alltid grundig omtalt i evalueringsrapportene, noe vi kommer tilbake til i kapittel 7. I omtalen av prosjektene dreier det seg om konflikter i veiprosjekter i forhold til klima- og miljømål, og at forskjellige grupper drar nytte av prosjekter i ulik grad. Eller at det mellom eller innenfor våpengrenene er ulik oppfatning om hvilket forsvarsmateriell som skal prioriteres. Et annet eksempel på målkonflikt er når ulike geografiske regioner har ulike behov. Vi ser dette blant annet i prosjektet LOFAST hvor den valgte traseen var best for Lofoten, men ikke for Vesterålen og heller ikke for et regionalt mål om å integrere Lofoten og Vesterålen til én region. Det sistnevnte var ikke formulert eksplisitt som mål for prosjektet, men var like fullt erkjent som et viktig hensyn i prosjektets tidligfase. I dette tilfellet valgte evaluator å vektlegge behovet for en integrert region ved relevansvurderingen.

Et annet spørsmål, som ikke alltid er godt besvart av evaluator, er hvorvidt en annen løsning på behovet kunne vært *mer relevant* enn den som ble realisert.

Tabell 4.5 Forklaringsvariable som gjelder vurderingen av prosjektenes relevans

Relevans	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Var prosjektet relevant	95	10	8
Vesentlige konflikter/ interesse-motsetninger	65	5	8

Levedyktighet

I samtlige av de mest vellykkete prosjektene kommer man til at de vil gi nytteeffekt i mange år fremover. Forutsetningene varierer. Levedyktigheten for Svalbard forskningspark er avhengig av internasjonal politikk og at Norge opprettholder sine økonomiske forpliktelser på øygruppen. For kontrollstasjonen på Svinesund forutsetter man naturlig nok en opprettholdelse av fortollingsregimet mellom EU- og EØS-land. I veiprosjekter bygger resonnementet på forventninger om fortsatt trafikkvekst, det gjelder også bompengefinansierte prosjekter, og akseptabel utvikling når det gjelder miljø- og klimaeffekter og andre sidevirkninger. I tillegg må de finansielle konsekvensene knyttet til drift og vedlikehold være akseptable – der dette ikke er tilfelle trekker det scoren ned, det gjelder blant annet Nytt

Operabygg hvor en avhengighet av svært store statlige driftstilskudd, samtidig som store pensjonsforpliktelser truer driften på sikt. Men det har også å gjøre med hvilken levetid en legger til grunn, og fremtidig behov. Vurderingen av levedyktighet er nødvendigvis mer kortsiktig for et IKT-prosjekt (10-15 år) enn for et jernbaneprosjekt som er ment å vare i generasjoner. I tilfellet Lofast dreier dette seg om hvorvidt fremtidig investering i en annen trase over Vesterålen vil redusere behovet for den veien som nå er bygget.

Spørsmålet om fleksibilitet i forhold til fremtidige behov er også relevant for vurderingen av levedyktighet. I de tilfellene der dette er drøftet i evalueringen handler det om utvidelser av (eller alternativt bruk av) bygg og endringer i IKT-systemer etter at de er satt i drift.

Tabell 4.6 Forklaringsvariable som gjelder vurderingen av prosjektenes forventete levedyktighet

Levedyktighet	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Vurdert levedyktig i lang tid fremover	80	10	6

Samfunnsøkonomisk effektivitet

Kvantifisering av nytten er notorisk problematisk i mange statlige investeringstiltak, i motsetning til kommersielle prosjekter der nytte er synonymt med inntekt, som kan prissettes direkte og ofte med stor presisjon. En tilleggsutfordring i våre evalueringer, som altså ikke har gjennomgått KVVU/KS1, er at det sjelden foreligger en tilstrekkelig grundig for-analyse.

Evaluatorene er bedt om å gjøre en *vurdering* av prosjektets samfunnsøkonomiske effektivitet, oppsummert ved en tallscore, på samme måte som for de øvrige kriteriene, men ikke nødvendigvis beregne lønnsomheten i kroner og øre. Som påpekt i Volden og Samset (2013) er det viktig å ha et nøkternt ambisjonsnivå når det gjelder kvantifisering av samfunnsøkonomisk effektivitet. I halvparten av prosjektene har man da heller ikke greid å kvantifisere nytten. I fem av veiprosjektene er nytten basert på etatens modellberegninger, med den usikkerheten og forutsetningene som da er langt til grunn, og som en ikke nødvendigvis gjør godt rede for i rapportene.

Den største forskjellen mellom de ti minst og de ti mest vellykkete prosjektene er nettopp her. I den ene gruppen er bare to prosjekter vurdert som lønnsomme, i den andre gruppen så mange som åtte av ti.

Spørsmålet om lønnsomheten er større eller mindre enn forutsatt er noe som bare omtales i vel halvparten av rapportene, og da med positivt resultat. Dette gjelder særlig veiprojektene, hvor det i flere tilfeller faktisk forelå før-analyser. En hovedårsak til at prosjektene ble mer lønnsomme enn forutsatt, er at trafikken ble større. Evaluatorene trekker imidlertid også frem at retningslinjene for nytte-kostnadsanalyser har endret seg i «fordelaktig retning» for veiprojekter (se sist oppdaterte rundskriv, Finansdepartementet, 2014). Det nevnes blant annet at kalkulasjonsrenten er lavere, analyseperioden lengre og en har begynt å realprisjustere tidsgevinster, samtidig som for eksempel negative miljøvirkninger fremdeles har lav pris i analysene. Dermed fremstår lønnsomheten som bedre i dag enn da prosjektet ble utredet tidlig på 2000-tallet.

Tabell 4.7 Forklaringsvariable som gjelder vurderingen av prosjektenes samfunnsøkonomiske effektivitet

Samfunnsøkonomisk effektivitet	Samlet for alle prosjektene (%)	Av de 10 mest vellykkete	Av de 10 minst vellykkete
Ble tiltaket vurdert som samfunnsøkonomisk effektivt	50	8	2
Var det mulig å kvantifisere nytten	40	5	3

5. Konklusjoner

Finansdepartementets kvalitetssikringsordning innebærer at det stilles strenge krav til kvaliteten på beslutningsgrunnlaget før store statlige investeringsprosjekter kan behandles politisk. Etter at ordningen ble innført i år 2000 og utvidet i 2006, har departementer og etater forbedret sin praksis når det gjelder utredning og planlegging av disse (som vist blant annet i Kvalheim m.fl., 2015). Vi vet også at den samme metodikken etter hvert og i økende grad benyttes på mindre prosjekter og i sektorer utenfor KS-ordningen.

Nå er mange av de første prosjektene kommet i driftsfasen og et relevant spørsmål er om prosjektene, som altså er godt utredet i forkant, også blir godt gjennomført og når sine mål. Tradisjonelt har statlige investeringsprosjekter sjelden vært gjenstand for formell etterevaluering. Det er de heller ikke under KS-ordningen. Men som nevnt har Concept som følgeforskningsprogram påtatt seg å sørge for etterevaluering av en andel av KS-prosjektene etter at de har kommet noen år ut i driftsfasen. Med den evalueringsmodellen som er valgt, vurderes prosjektenes vellykkethet både operasjonelt, taktisk og strategisk. Alle evalueres etter samme modell, med seks overordnede evalueringskriterier. Vurderingen for hver av dem oppsummeres ved en score mellom 1 og 6. Prosjektene evalueres på omtrent samme tidspunkt i sin livssyklus. Dette gir et godt grunnlag for å sammenligne resultatene mellom prosjekter og sektorer.

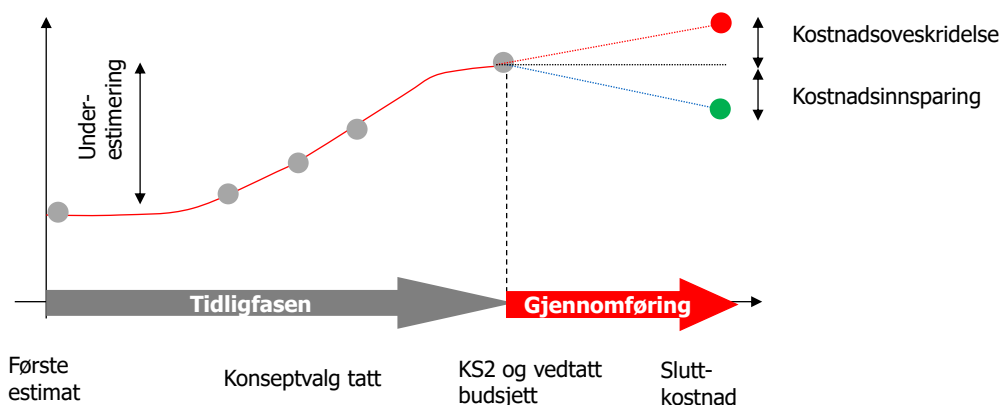
God kostnadskontroll i gjennomføringsfasen

Et nøkkelresultat fra metaevalueringen er at prosjektene kommer godt ut i et operasjonelt perspektiv. Kostnadskontrollen er god i de fleste prosjektene og evaluatorene konkluderer med at de er effektivt gjennomført.

Byggeprosjektene fremstår som overveiende vellykkete målt mot de avtalte resultatmålene. Men også veisektoren, som kommer svakest ut, scorer høyt i halvparten av prosjektene. Det er mye som tyder på at KS2-ordningen har bidratt til dette. Prosjektene i denne studien var de første som gjennomgikk KS2 tidlig på 2000-tallet.

Et forhold som imidlertid bør problematiseres, er at evaluatorenes vurdering gjelder overholdelse av de rammene som er satt av Stortinget *etter* KS2, og ikke kostnadsestimatet da den politiske forpliktelsen om å gjennomføre prosjektet, ble tatt, som vil være på et tidligere tidspunkt. Dette estimatet er ikke tidfestet i

de 20 prosjektene, som ikke har gjennomgått KS1. Figur 5.1 viser hvordan utviklingen for kostnadsestimatet typisk kan se ut, fra det første estimatet blir lansert sammen med prosjektideen, via konseptvalget og frem til budsjettvedtak og endelig sluttkostnad. Concept har i en tidligere studie vist at økningen gjennom hele denne perioden kan være betydelig og i prosent langt overstige eventuelle overskridelser i gjennomføringsfasen (Welde m.fl., 2014). I en annen studie av 20 prosjekter som har vært gjennom både KS1 og KS2 fant man at veksten mellom disse to kontrollpunktene selvsagt hadde vært mindre, men allikevel på hele 41 % i gjennomsnitt (Welde, 2016). Dette er en type kostnadsøkning som beslutningstaker er kjent med når prosjektet endelig vedtas. Men blant annet på grunn av den politiske kostnaden det vil medføre, er det ofte svært vanskelig (nesten umulig) å ikke skulle vedta prosjektet allikevel. Det vet interessentene, som i konseptfasen kan ha incentiver til å øke sine krav til løsningen.



Figur 5.1 Mulig utvikling av kostnadsestimatet basert på Samset (2008)

Dette er en problematikk som en bør se på i fremtidige evalueringer. Hvorvidt prosjektet har hatt en overskridelse eller ikke må vurderes i forhold til det vedtatte budsjettet, slik en her har gjort. Men kostnadsveksten i tidligfasen er relevant blant annet for å svare på om den valgte løsningen er kostnadseffektiv, noe som vil kunne slå ut ikke bare på vurderingen av produktiviteten, men også på relevans og ikke minst samfunnsøkonomisk effektivitet.

Effektene kommer ikke av seg selv

At prosjektene som har gjennomgått KS2 har god kostnadskontroll, er som forventet ut fra hva andre studier har vist (Samset og Volden, 2013; Welde, 2017). Noe mer overraskende er det at porteføljen også kommer godt ut taktisk og strategisk. Om dette er reelt er fremdeles et åpens spørsmål, fordi underlaget for de taktiske og strategiske vurderingene til dels har vært mangelfullt ettersom ingen av prosjektene har vært gjennom KVVU og KS1. Vi kan ikke se helt bort fra at evaluator har blitt litt påvirket av de gode resultatene som gjelder det operasjonelle. Det burde ikke overraske noen at disse største investeringsprosjektene i regi av staten er godt utredet og at konseptvalgene er rimelig fornuftige. Men det kan likevel være grunn til å forvente et løft i taktisk og strategisk ytelse for fremtidige KS1-prosjekter sammenlignet med disse.

Denne metaevalueringen har avdekket forbedringspotensialer på flere punkter. Mange av prosjektene har ambisiøse effektmål, som åpenbart krever en større og mer integrert innsats ut over selve prosjektet for å realiseres. Det gjelder for eksempel jernbaneprosjekter som har mål om bedre togtilbud og økt markedsandel for jernbanen, og byggeprosjekter som har mål om faglige og økonomiske synergier for brukerne av bygget. I mange av prosjektene med ambisiøse effektmål er måloppnåelsen svak selv etter flere års drift, og det later til å kunne forklares med at en nettopp *ikke* har satt inn noen integrert innsats. Det kan tyde på at staten bør styrke kompetansen om hvordan økonomiske, juridiske, informative og andre virkemidler og tiltak fungerer sammen med investeringsprosjekter. Videre må effektmålene være realistisk oppnåelige, og at ansvaret for å realisere tiltenkte effekter må plasseres tydelig, fordi de ellers fort kan bli glemt så snart prosjektet har fått sin bevilgning. Heldigvis ser vi at staten nå begynner å bli opptatt av dette. Direktoratet for økonomistyring har publisert en veileder om gevinstrealisering (DFØ, 2014b) hvor et hovedbudskap er at mål ikke realiseres av seg selv, men at det krever god planlegging, oppfølging og ansvarliggjøring. I IKT-sektoren har man lenge vært bevisst dette. Nå ser vi at man i jernbanesektoren snakker om effektpakker fremfor kun utbyggingsprosjekter, det vil si at utviklingen av infrastrukturen ses i sammenheng med endret rutemodell og andre elementer for å forbedre togtilbudet for passasjerene (Jernbaneverket, 2015).

Våre resultater viser videre at lønnsomheten er noe svak i flere av prosjektene (dog ikke nødvendigvis svakere enn anslått i forkant). Flere av de ulønnsomme prosjektene kommer likevel godt ut på de øvrige strategiske kriteriene, noe som kan skyldes at det ikke er verdiskaping som begrunner dem, men andre

politiske mål, som for eksempel klimamål, kulturpolitiske eller forsvarspolitiske mål. Det er uansett viktig at samfunnsøkonomiske analyser gjennomføres ex ante, i tillegg til det som gjelder de målorienterte evalueringskriteriene, og at analysene er transparente nok til å kunne brukes som sammenligningsgrunnlag ex post. Det har ikke vært tilfelle så langt.

Som nevnt er referansegrunnlaget for den taktiske og strategiske delen av evalueringen generelt mangelfull i disse 20 første prosjektene. Vurderingene av andre virkninger, relevans, levedyktighet og samfunnsøkonomisk effektivitet vil alltid måtte ha innslag av skjønn, men dersom grunnlaget for konseptvalget hadde vært godt dokumentert i forkant ville evaluators jobb i etterkant ha vært langt enklere. Problemet er trolig av midlertidig art og vil bedre seg i senere evalueringer av KS1-prosjekter. Concept har imidlertid påvist i andre studier at også KVVU-er og KS1-rapporter har et forbedringspotensial når det gjelder dokumentasjon av forutsetninger og begrensninger (Kvalheim, 2015), så det er all grunn til å fortsette å presisere viktigheten av transparente analyser og vurderinger.

Læringspotensialet

En annen utfordring handler ikke om kvaliteten eller innretningen på evalueringene, men om hvordan en kan sikre god formidling og bruk av dem. Det er når prosjektenes eiere i fagdepartement og etat tar lærdom av resultatene og forbedrer sin praksis, at vi vil se nytteverdien av evalueringsaktiviteten. Erfaringen så langt er at selv om interessen har vært stor hos enkelte, har vi generelt et forbedringspotensial når det gjelder å engasjere disse aktørene. Aktuelle tiltak kan handle om bedre formidling av nøkkelresultater og tettere involvering av prosjekteier i evalueringsprosessen på ulike måter, eventuelt å trekke inn ressurser fra departement eller etat i selve evalueringsteamet der dette er mulig. Det må presiseres at vi ikke har snakket med brukerne av evalueringene som del av denne metaevalueringen – dette vil vi følge opp på andre måter i etterkant.

Evaluering har både et kontroll- og et læringsformål. Det første er særlig viktig når det gjelder prosjektenes operasjonelle vellykkethet, fordi ansvaret for leveransene er tydelig plassert hos de prosjektansvarlige. Som nevnt over bør en også være mer opptatt av å plassere ansvar for at brukereffektene oppnås – i det minste der disse i hovedsak skal realiseres innenfor en bestemt virksomhet. Likevel er det læringsperspektivet vi har vært mest opptatt av, og å fremskaffe informasjon som er relevant for selve gjennomføringen av

prosjekter (den lille læringssløyfen) og typen av prosjekt, det vil si konseptvalget (den store læringssløyfen).

Dette er bare begynnelsen – planen er å bruke evalueringsmodellen til å etterevaluere en stor andel av de mange investeringsprosjektene som settes i drift nå og i årene fremover. Over tid vil en få svar på om de mønstrene vi har sett også gjelder i en større portefølje. En vil også ha et bedre grunnlag for å si noe om hvilke tiltak som kan bidra til bedre analyser og beslutningsprosesser i forkant. Ikke minst blir det viktig å evaluere prosjekter som har vært gjennom KS1, hvor disse 20 første evalueringene vil utgjøre en kontrollgruppe for å vurdere effekten av den utvidete ordningen.

Del 2. Vurdering av evalueringsopplegget og evalueringene

6. Nærmere om metodikk for prosjektevaluering

I dette kapitlet går vi nærmere inn på det metodiske opplegget som evalueringene er bygget på. Prinsippet er at alle evalueringer må ta utgangspunkt i prosjektets målstruktur, og vurdere denne kritisk. Videre har det vært ønskelig fra vår side å evaluere prosjektene i et bredt perspektiv, som dekker både operasjonelle, taktiske og strategiske aspekter. Vi har derfor tatt utgangspunkt i de fem såkalte OECD-kriteriene, kombinert med kriteriet samfunnsøkonomisk effektivitet. Hvordan disse operasjonaliseres i konkrete evalueringsspørsmål eller indikatorer er essensielt for å sikre tilstrekkelig og etterrettelig grunnlag for vurderinger og konklusjoner. Det er viktig at valg av indikatorer, datainnsamlings- og analysemetoder tilpasses det enkelte prosjekt. Men generelt bør en ikke bruke for mye ressurser på detaljer – evalueringene skal først og fremst gi et overordnet bilde av prosjektenes vellykkethet. Vi kommer også inn på utfordringer ved bruk av tallscore samt krav til evalueringsteamets kompetanse.

Prosjektets årsak-virkningskjede må spesifiseres

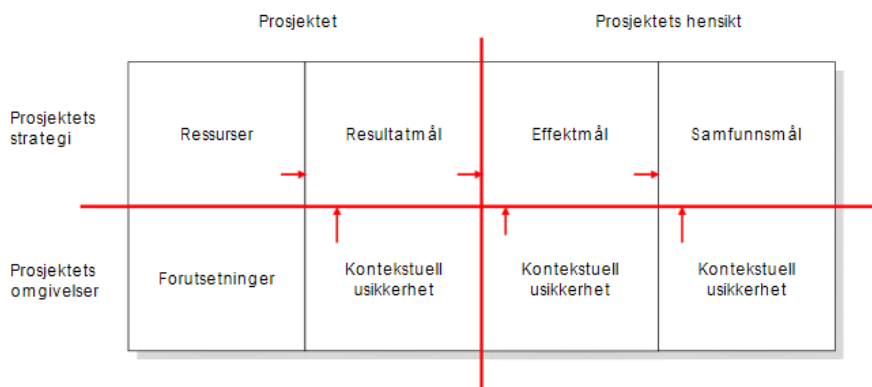
Et hvert prosjekt bygger på en idé om en underliggende kausal sammenheng mellom de viktigste hendelsene - fra ressursinnsatsen, via prosjektaktivitetene, til leveransen, brukereffektene og samfunnseffektene. Denne kjeden har flere navn, herunder prosjektets logikk, målstruktur, strategi, årsak-virkningskjede og programteori. I de statlige KS-prosjektene skal det være satt *mål* for de ulike nivåene: resultatmål for selve leveransen, effektmål for brukereffektene og samfunnsmål for samfunnseffektene. Da blir den kausale kjeden synonymt med en målstruktur.

Finansdepartementets veileder i gjennomføring av evalueringer (2005) krever at årsak-virkningskjeden for det aktuelle tiltaket skal legges til grunn ved evalueringer. Men erfaring viser at denne ofte er uklar (Karlsen og Jentoft, 2013) og at målstrukturer er preget av feil som gjelder selve logikken og sannsynlighet for realisering av mål på ulike nivåer (Samset m.fl., 2013). I slike tilfeller må evaluator justere eller omdefinere målene, for at det skal gi mening å evaluere prosjektet i forhold til dem.

Det finnes ulike tilnærminger til hvordan prosjektets målstruktur fremkommer og beskrives. Spesielt er logisk rammeverk og programteori mye brukt.

Logisk rammeverk

Logisk prosjektanalyse, på engelsk best kjent som LFA¹⁰, har en lang tradisjon innenfor utforming av utviklingsprosjekter. Teknikken tar utgangspunkt i et logisk rammeverk og knyttes opprinnelig til Leon Rosenberg og konsulentfirmaet Practical Concepts Inc. som utarbeidet en modell for USAID (se Rosenberg et al., 1970, og senere veileder, PCI, 1975). Dette spredte seg raskt, både til FN-systemet, bistandsadministrasjonen i en rekke land (herunder de skandinaviske) og senere OECD og EU-kommisjonen. Etter at disse begynte å kreve bruk av logisk rammeverk ved søknader om prosjektfinansiering, har en hel generasjon av organisasjoner og konsulenter tatt teknikken i bruk.



Figur 6.1 Eksempel på logisk rammeverk (Samset, 2008)

Rammeverket anvendes i forskjellige varianter¹¹, men består i prinsippet av en matrise definert ved målhierarkiet i fire nivåer og spesifisering av forutsetninger/risiko på hvert nivå, se figur 6.1. Hensikten er at en vil forankre den detaljerte beskrivelsen av det påtenkte prosjektet i overordnede mål, samtidig som analyse og håndtering av usikkerhet og risiko i gjennomføringen

¹⁰ Logical Framework Approach (LFA)

¹¹ I mer detaljerte versjoner inngår spesifisering av målbare indikatorer og datakilder på hvert målnivå i matrisen.

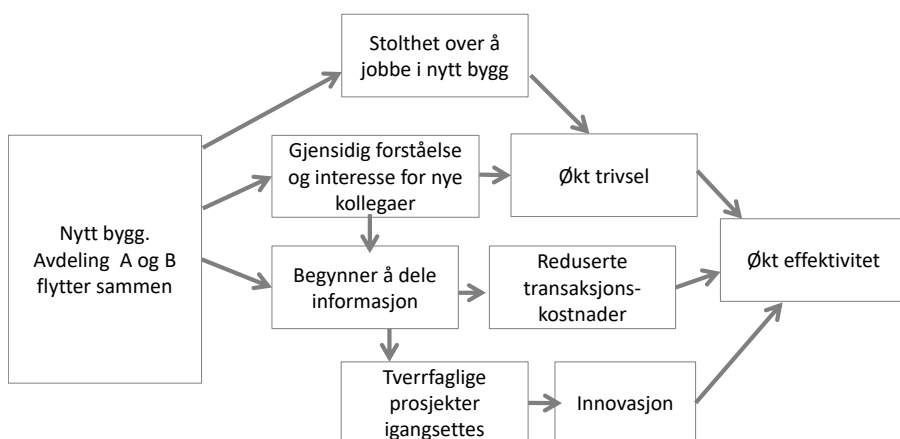
står sentralt. Prosjektleder skal dermed få en samlet oversikt over hva som forventes og hvilke eksterne faktorer som trolig vil påvirke resultatet.

Teknikken brukes til å løfte blikket fra selve prosjektet til formålet med det. Logikken har samtidig vært et viktig utgangspunkt for et evalueringsformat, nå kjent som de fem OECD-kriteriene som omtales i neste avsnitt. En kritikk mot rammeverket var i starten at en ble for opptatt av tilsiktede effekter og for å bare representere prosjekteiers interesser (Sartorius, 1991; Gasper, 2000; Bakewell og Garbutt, 2005). Over tid har praksis utviklet seg mot å legge større vekt på involvering av interessentenes perspektiver, i bistandssektoren ikke minst etter Paris-deklarasjonen (2003) som legger til grunn at mottakerlandene selv skal formulere sine utviklingsplaner og at ekstern bistand skal støtte opp om disse. Men teknikken er fortsatt utbredt, både i planlegging, oppfølging og evaluering av prosjekter. Bruken varierer, fra detaljorienterte varianter med store matriser som skal fylles ut, til den enkle logiske kjeden som vist over (Bakewell og Garbutt, 2005). Noen har hevdet at LFA-teknikken er for enkel, fordi effektkjedene i virkeligheten ikke er lineære og endimensjonale, men omfatter en rekke ulike prosesser og aktører (det siste er bl.a. brukt som argument for teoribasert evaluering, se under).

Programteori

Programteori er et annet begrep for å beskrive de kausale mekanismene som skal gi en bestemt effekt (Chen, 1990; Weiss, 1997; Rogers m.fl., 2000). Deler av denne litteraturen, som kan sies å ligge nærmere forskning enn evaluering, er opptatt av de samfunnsvitenskapelige teoriene som tiltaket er basert på (konsumentteori, motivasjonsteori, etc.).¹² Teoriene gir nærmere angitte hypoteser om hvordan målgruppen og andre vil respondere på tiltaket. Det er ofte snakk om flere mekanismer som virker sammen, se figur 6.2 for en enkel illustrasjon av en mulig programteori for et byggeprosjekt. Begrepet programteori knyttes opprinnelig til Edward Suchman som på 1960-tallet snakket om at fravær av ønskede effekter kunne ha to årsaker: enten at tiltaket var dårlig implementert (*implementation failure*) eller at man hadde en dårlig teori (*theory failure*). For å undersøke det siste er en innledende litteraturgjennomgang og ekspertvurdering ofte nyttig, og vil kunne avsløre en dårlig teori lenge før man har samlet inn noe empiri om effektene (Brousselle og Champagne, 2011).

¹² Det vanlige er likevel at programteorier har betydelig innslag av legfolks teorier og bare i begrenset grad er vitenskapelige, noe som også skyldes at det er vanskelig å spesifisere relevante teorier for de komplekse situasjonene det ofte er snakk om.



Figur 6.2 Enkel illustrasjon av en programteori

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 variant som er mye omtalt er Realistisk Evaluering som knyttes til Pawson og Tilley (1997). Det nye er at de bruker et generativt kausalitetsbegrep og er opptatt av at resultater ikke er noe som «kommer av seg selv», men skapes i møte med brukerne/målgruppen. En må derfor fokusere på hva som motiverer folk. Effekten vil være et sosialt produkt av konteksten og de mekanismene som oppstår i denne konteksten. Hvis evalueringsresultater skal kunne ha overføringsverdi til andre, lignende tiltak, er det følgelig ikke nok å beskrive en effekt, men også hvilken kontekst og mekanisme som utløste den.¹³

Praktisk tilnærming

I et operasjonelt perspektiv er det selve prosjektet som er gjenstand for evaluering, men i et taktisk og strategisk perspektiv må prosjektet vurderes i et bredt samfunnsperspektiv og ofte i forhold til andre prosjekter eller tiltak. I Concept-rapport nr. 30 (Volden og Samset, 2013) og Concept-programmets

¹³ En annen variant er Theory of Change (ToC) eller endringsteori, ofte brukt for å fremme sosial endring på et komplekst felt. Bred deltakelse står sentralt i ToC-tilnærmingen – godheten av programteorien vurderes langt på vei ut fra om det er konsensus om den blant alle involverte. Se Blamey og MacKenzie (2007) om forskjellen på Realistisk Evaluering og ToC.

lærebok om tidligfasevurdering (Samset, 2014) legger vi til grunn en enkel logisk modell hvor man spesifiserer ressurser, selve prosjektgjennomføringen, leveransene, brukereffekter og samfunnseffekter, med mål for de ulike nivåene, og sentrale usikkerheter knyttet til overgangen mellom dem¹⁴.

I mange tilfeller foreligger godt dokumenterte beskrivelser av målstrukturen og da er det naturlig å ta utgangspunkt i dette. Men en bør unngå at målstrukturen bare blir en ad hoc beskrivelse av hva prosjekteier tror eller ønsker.

Sammenhengen mellom leddene i kjeden må vurderes kritisk. En annen kompliserende faktor er at målene og logikken kan endres underveis, enten formelt eller i forståelse mellom aktørene¹⁵. En gjennomgang, vurdering og eventuell revisjon av målstrukturen er ofte avgjørende for å kunne utforme et hensiktsmessig design for evalueringen og velge fornuftige indikatorer. Dette vil være en viktig innledende oppgave for evaluator før evalueringen settes i gang.

Der det er relevant bør evaluator også, ved hjelp av eksisterende litteratur og samtale med eksperter, undersøke godheten av de underliggende teoriene nærmere. En full evaluering av programteorien forventes ikke, men det vil 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 ofte godt dokumentert (regelmessige reisevaneundersøkelser etc.). Det betyr at en observert trafikkøkning i etterkant av en kapasitetsforbedring kan antas – med stor grad av sikkerhet – å være en effekt av prosjektet. Det samme kan ikke uten videre sies om regionale endringer i økonomi og sysselsetting – mange transportprosjekter har ambisjoner om å bidra til økt økonomisk vekst, men at dette skulle være en effekt av prosjektet er langt fra sannsynliggjort i samme grad.¹⁶ Første ordens effekter er naturlig nok lettere å finne støtte for enn komplekse annen- og tredjeordens effekter.

¹⁴ Det kan også vises til en mye referert veileder om bruk av logisk analyse, W. K. Kelloggs Foundation (2004), som viser at både logiske rammeverk og programteorier kan brukes enkelt og fleksibelt.

¹⁵ Uansett, målene må reflektere hva som er realistisk oppnåelig. Dersom de avtalte målene er urealistiske vil prosjektet måtte vurderes som mislykket. Om målet er trivielt vil det formelt sett måtte vurderes som en suksess.

¹⁶ Som påpekt i Næss m.fl. (2017) er såkalt mernytte ikke godt dokumentert empirisk i Norge – derimot har flere studier forsøkt å *simulere* dem basert på til dels sviktende forutsetninger (som for eksempel at alt areal sentralt er brukt opp, slik at agglomerasjon kun kan skje ved pendling fra regionene utenfra). Det er tvilsomt om en evaluator skal ta denne type forutsetninger for gitt.

Tilsvarende har mange byggeprosjekter mål om å skape ulike typer samlokaliseringsevinster (ref. figur 5.2), men Hatling m.fl. (2016) viser at heller ikke dette er godt forankret i litteraturen.

Ifølge Karlsen og Jentoft (2013) er en teoribasert tilnærming først og fremst nyttig for tiltak som griper inn i folks atferd og holdninger. Dessuten for innovative og komplekse tiltak, hvor det er stor usikkerhet om hvorvidt og hvordan de vil virke. Kaplan og Garrett (2005) nevner at et vanlig eksempel på sviktende programteori er å anta at folk, uten veiledning eller andre incentivordninger, vil endre sine vaner bare fordi en bestemt teknologi eller infrastruktur er gjort tilgjengelig. En gjennomgang av programteorien vil kunne peke på andre supplerende virkemidler som eventuelt mangler – eller omvendt, prosjektkomponenter som er overflødige.

Et balansert sett av seks evalueringskriterier

Prosjektets årsak-virkningskjede, eller målstruktur, er utgangspunktet for å identifisere evalueringsspørsmål og databehov. Finansdepartementets veileder i evaluering (2005) krever at en skal vurdere resultater langs hele kjeden. Dette har også stått sentralt ved bruken av logisk rammeverk – det handler om å «løfte blikket» ut over prosjektet i seg selv. Baccarini (1999) bruker for eksempel dette til å definere to typer eller nivåer av prosjektsuksess, nemlig prosjektledersuksess (som handler om gjennomføringen) og produktsuksess (realisering av mål på bruker og samfunnsnivå). Samset (2014) definerer prosjekters vellykkethet på tre nivåer, operasjonelt, taktisk og strategisk, relatert til de tre målnivåene resultatmål, effektmål og samfunns mål.

I tillegg til å vurdere måloppnåelse på ulike nivåer, bør en også se på hvilke, gjerne utilsiktede, sidevirkninger prosjektet har hatt, for målgruppen og andre grupper. Videre bør en se på tiltaket i et økonomisk perspektiv. Samfunnet har ikke ubegrenset med ressurser og det er viktig å vurdere hva en har fått for pengene.

Når det gjelder Concept-programmets evaluering av KS-prosjekter, ønsker vi en overordnet vurdering av prosjektenes vellykkethet i et bredt perspektiv, som dekker alle disse sentrale aspektene ved vellykkethet. Ved å definere et sett av felles kriterier har man både en nyttig sjekkliste for hva som bør vurderes, og et grunnlag for å sammenligne resultater.

De fem «OECD-kriteriene»

OECD's Development Assistance Committee gjorde på 1990-tallet en bred gjennomgang av evalueringsmetodikk og begrepsbruk hos medlemslandene og internasjonale bistandsorganisasjoner. Utgangspunktet var organisasjonens definisjon av evaluering som inneholder nettopp de fem kriteriene (OECD, 1991, vår understreking):

An evaluation is an assessment, as systematic and objective as possible, of an ongoing or completed project, programme or policy, its design, implementation and result. The aim is to determine the relevance and fulfillment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.

Kriteriene er nærmere definert og forklart i en flerspråklig ordliste som et bidrag til å harmonisere begrepsbruken internasjonalt (OECD, 2002).

Status i dag er at denne definisjonen, og de fem overordnede evalueringskriteriene, har blitt gjeldende standard når det gjelder evaluering av utviklingsprosjekter. NORAD viser for eksempel til de fem kriteriene på sin nettside, det samme gjør de andre skandinaviske bistandsorganisasjonene, men også OECD, EU-kommisjonen (2004), UNDP og de fleste særorganisasjonene. I andre sektorer benyttes også de fem kriteriene eller varianter av dem. Det er for eksempel utarbeidet syv evalueringskriterier for nødhjelpsprosjekter, de såkalte ALNAP-kriteriene, som er en utvidelse av de fem (OECD, 1999, og ALNAP, 2006)¹⁷. EU sin veileder i evaluering av sosial utvikling benytter i dag en litt mer fleksibel variant av de fem kriteriene (EU, 2013)¹⁸ og for evaluering av reguleringer brukes en variant bestående av

¹⁷ Ettersom nødhjelpsprosjekter står overfor litt andre utfordringer enn utviklingsprosjekter, er *sustainability* erstattet med *connectedness* (sikre at de kortsiktige aktivitetene tar hensyn til langsiktige utfordringer). Videre har man supplert med tilleggskriteriene *coverage* (hvorvidt tiltaket når frem til utsatte grupper) og *coherence* (at nødhjelpstiltaket er konsistent med det som gjøres innen utvikling, handel og andre politikkområder).

¹⁸ EU hadde tidligere en detaljert veileder tilpasset strukturfondets tiltak (de såkalte MEANS-håndbøkene), men sier i veilederen fra 2013 at det var behov for en mer fleksibel tilnærming. Det fremgår at evalueringskriteriene kan være, for eksempel, de fem OECD-kriteriene, men at også andre kriterier nevnes, som *equity*, *coherence*, *synergy* og *reproducibility* (EU, 2013).

kriteriene *efficiency*, *effectiveness*, *relevance*, *coherence* og *EU added value* (EU, udatert).¹⁹

Selv om det er gjort enkelte lokale tolkninger, tilpasninger og åpnet for større fleksibilitet, har de fem kjerne-kriteriene stått støtt. Blant annet skriver OECD (1999) at de syv ALNAP-kriteriene strengt tatt kunne vært dekket av de fem eksisterende. En grundig gjennomgang av de fem kriteriene gjort av en gruppe av profesjonelle evaluatorene (Chianza, 2008) konkluderte med at de fungerer godt – herunder at de i all hovedsak tilfredsstiller en anerkjent sjekkliste av kriterier for programevaluering laget av forskeren Michael Scriven. En peker imidlertid på at *relevance* i praksis ofte er mer fokusert på myndighetenes mål enn befolkningens, at *sustainability* tolkes noe snevert (omfatter ikke det som gjelder kultur, politisk støtte, teknologi, institusjonell kapasitet, etc.), og at *efficiency* ofte avgrenses til betalbare kostnader, men burde dekke formålseffektivitet og samfunnsøkonomisk effektivitet også.

Concept-programmet har tilpasset bruken av de fem kriteriene til det som gjelder evaluering av store offentlige investeringstiltak med utgangspunkt i vår lærebok på dette området (Samset, 2014).

- Vi avgrenser *produktivitet* til det som gjelder gjennomføringen og spørsmålet om hvor effektivt prosjektorganisasjonen har omsatt ressurser til leveranser. Til sammenligning omfatter tolkningen av *efficiency* i prinsippet flere nivåer av effektivitet (se for eksempel Sida, 2007, og Scriven, 2013). Vi har løftet samfunnsøkonomisk effektivitet ut som eget kriterium. Derimot må produktivitet vurdere alle sider ved gjennomføringen, også HMS og i hvilken grad KS2-rapportens anbefalinger er fulgt. Slik sett strekker vi produktivitetsbegrepet noe i forhold til slik det vanligvis brukes.
- *Måloppnåelse* handler i hovedsak om realiseringen av prosjektets effektmål. Der formelt avtalte mål er urealistiske eller trivielle må de justeres. Slike endringer skal begrunnes og bør i prinsippet sikres legitimitet for. Dette er i stor grad i tråd med både OECDs definisjon og med gjeldende praksis andre steder.

¹⁹ Vi kan også legge til at det finnes andre sett av kriterier som ligner. Et eksempel er RE-AIM modellen (*reach*, *effectiveness*, *adoption*, *implementation* og *maintenance*) for evaluering av helsetiltak. Rammeverket ble først presentert i en forskningsartikkel, Glasgow (1999) som kritiserte det kortsiktige og snevre perspektivet i typiske studier av effekten av opplæringsprogrammer for diabetikere, hvor man kun spurte om helseeffekten på enkeltindivider under ideelle betingelser. Glasgow foreslo en utvidet modell som også fanget effekten på folkehelsen på lang sikt. Vår vurdering er at de fem kriteriene i RE-AIM langt på vei dekker det samme som de fem OECD-kriteriene (for eksempel er *adoption* langt på vei det samme som relevans). RE-AIM-rammeverket er i dag utbredt for helsetiltak i USA, se www.re-aim.org

- *Andre virkninger* tolkes snevrere i vår modell ved at vi velger å fokusere kun på utilsiktede virkninger, mens internasjonal praksis er å gjøre en samlet vurdering av alle virkninger (dette tydeliggjøres ved at vi kaller kriteriet andre virkninger). Vi mener at vår tolkning gir en mer fornuftig avgrensning mot *måloppnåelse*, som er avgrenset til de positive virkningene en forventer å oppnå som resultat av prosjektet. Det kan også nevnes at slik det praktiseres andre steder er dette kriteriet ofte uklart, blant annet fordi *impact* på engelsk er et flertydig begrep, som både kan bety «samlete virkninger», men også «langsiktige virkninger».
- Et prosjekt er *relevant* dersom det er behov for det prosjektet leverer. Relevans-kriteriet er krevende fordi en typisk må vurdere tiltaket opp mot behov og prioriteringer hos ulike interessenter og berørte. Det gjelder selvsagt lokalt, men ved tiltak som finansieres av staten vil det være viktig å minne om at også nasjonale hensyn skal tas i betraktning (Samset m.fl., 2014). I tråd med behovsanalysen i konseptvalgutredninger skal etterevalueringen se spesielt på de tre perspektivene nevnt i Næss (2005), nemlig overordnede politiske behov, etterspørsel og interessegruppebaserte behov.
- Innen bistand handler *levedyktighet* om hvorvidt det er sannsynlig at aktiviteten, og dermed nyttestrømmene, vil fortsette etter at en ekstern giver har trukket seg ut. I norske KS-prosjekter som ikke er brukerfinansiert er det normalt lagt opp til at staten fortsetter å finansiere driften over hele levetiden. Da må en spørre om så vil skje, herunder om behovet er stabilt og om de samlede virkningene av tiltaket er akseptable for staten på lang sikt. Concept gjorde en egen utredning om dette kriteriet som anbefalte et både bredt og langsiktig perspektiv på levedyktighet, som inkluderer miljø-, økonomi- og samfunnspilaren, og spesielt tiltakets evne til å tilpasse seg endrede behov over tid (Lædre m.fl., 2012).

De fem kriteriene ble testet ut og viste seg robuste i en tidligere evaluering av et utvalg store historiske investeringsprosjekter, der datagrunnlaget er til dels svakt (Samset, 2012).

Samfunnsøkonomisk effektivitet

Det er mulig å bruke indikatorer for samfunnsøkonomisk effektivitet under ett eller flere av OECD-kriteriene. Vi har imidlertid valgt å løfte dette opp som et eget evalueringskriterium. Samfunnets ressurser er knappe, og effektiv bruk av dem er et overordnet hensyn. I Norge er kravet om å evaluere forankret i Økonomiregelverket og nært relatert til kravet om resultatoppfølging av midler bevilget over statsbudsjettet (Finansdepartementet, 2015). Det er ikke uvanlig

at evalueringskrav som kommer fra et finansdepartement tolkes som et krav om å vurdere samfunnsøkonomisk effektivitet²⁰.

Etter at KS1 ble innført har den samfunnsøkonomiske analysen vært en viktig del av utredningen av store statlige investeringsprosjekter²¹. Det legges ofte mye arbeid i disse analysene, det har vært to større utredninger om metodikken (NOU 1997:27 og NOU 2012:16) og både Finansdepartementets retningslinjer (2014) og DFØs veileder til gjennomføringen av samfunnsøkonomiske analyser (2014) er oppdatert og i tråd med beste praksis internasjonalt. En rekke etater har også egne veiledere.

Hensikten med analysen er å svare på om prosjektet er *samfunnsøkonomisk lønnsomt*, og rangere alternative konsepter etter hvilke som er mest lønnsomme. Bruken av penger som felles målestokk gjør at en kan sammenligne lønnsomheten før og etter og mellom prosjekter. Analysen skal omfatte både positive og negative, og betalbare og ikke-betalbare virkninger, for alle grupper over hele tiltakets levetid. En er ikke opptatt av formelt avtalte mål i prosjektet, og skiller derfor ikke mellom måloppnåelse og andre virkninger.

Identifiseringen av virkninger vil bygge på en resultatkjede for tiltaket, og sammenligningsgrunnlaget er det man tror ville skjedd i et kontrafaktisk ikke-investeringsalternativ (null-alternativet). Alle virkningene kvantifiseres og verdsettes så langt det er mulig og faglig forsvarlig. Nytte måles som summen av befolkningens betalingsvillighet for de godene som fremskaffes, mens kostnader verdsettes til alternativkostnad. Virkninger på ulike tidspunkt sammenstilles ved bruk av en diskonteringsrente.

Samfunnsøkonomisk analyse brukes først og fremst i prosjektanalysen *ex ante*, mens bare et fåtall land gjør regelmessige beregninger av lønnsomheten *ex*

²⁰ Det blir for eksempel ofte nevnt at det britiske Finansdepartementets veileder i evaluering, *The Magenta Book* (HM Treasury, 2011a) er sterkt tuftet på økonomiske prinsipper, og nært knyttet til *The Green Book* som er veilederen i samfunnsøkonomisk analyse (HM Treasury, 2011b). Av *Magenta book* fremgår at en evaluering bør stille spørsmål om tre temaer: hvordan tiltaket ble gjennomført, hvilke samlede virkninger det fikk, og om det var lønnsomt (HM Treasury, 2011a).

²¹ Samfunnsøkonomiske analyser har vært i bruk siden 1960-tallet i Norge, spesielt for veiprosjekter. Finansdepartementet utarbeidet en veileder i Programanalyser, herunder nytte-kostnadsanalyse, allerede i 1977, men forankringen i formelt regelverk kom først med Plan- og bygningslovens bestemmelser om konsekvensutredning (1990), Utredningsinstruksen (1994) og ikke minst Økonomiregelverket (1996). Med KS1 (2005) ble det et ufravikelig krav å gjøre slik analyse allerede i forstudiefasen for de største investeringsprosjektene.

post, og knapt noen har sektoruavhengige krav om dette²². Ifølge Worsley (2014) er mangelen på ex post analyser «det svake ledde» i analyse- og vurderingsprosessen for prosjekter og tiltak i transportsektor. I Norge har transportetatene gjennomført rutinemessige etterprøvinger av de viktigste prissatte virkningene siden 2006 (se egen veiledning, Statens vegvesen, 2006). Der dette er gjort for prosjekter som etterevalueres av Concept vil resultatet være nyttig input for oss.

En fullstendig nytte-kostnadsanalyse er neppe realistisk innenfor de rammene som gjelder Concepts evalueringer. De direkte effektene (brukereffekter) kan være krevende nok, men kvantifisering og verdsetting av sidevirkninger og ringvirkninger er svært ambisiøst. Mange har påpekt hvordan langsiktige indirekte virkninger nokså gjennomgående utelates i samfunnsøkonomiske analyser nettopp fordi de er usikre og ligger for langt ut i årsak-virkningskjeden (se f.eks. Hansen, 2011; Næss m.fl., 2017). Men inntil en har gode metoder for å kvantifisere dem kan det være bedre å omtale slike virkninger kvalitativt, samt oppgi begrensninger, forbehold og usikkerhet. Måten fremtidige virkninger neddiskonteres på gjør at langsiktige virkninger uansett ikke har særlig stor vekt når netto nåverdi beregnes. Med andre ord, velferden til senere generasjoner ignoreres langt på vei. Dette, i tillegg til at betalingsvillighet som mål på nytte favoriserer pengesterke grupper, og at en ikke tar hensyn til mål, standarder, kulturelle verdier eller etikk (Scriven, 2013), gjør at metoden er mye kritisert som normativt mål på hva samfunnet bør prioritere²³. Som *eneste* suksesskriterium for prosjekter er kriteriet derfor tvilsomt, men sammen med de fem OECD-kriteriene kan det gi nyttig informasjon om effektiviteten av tiltaket.

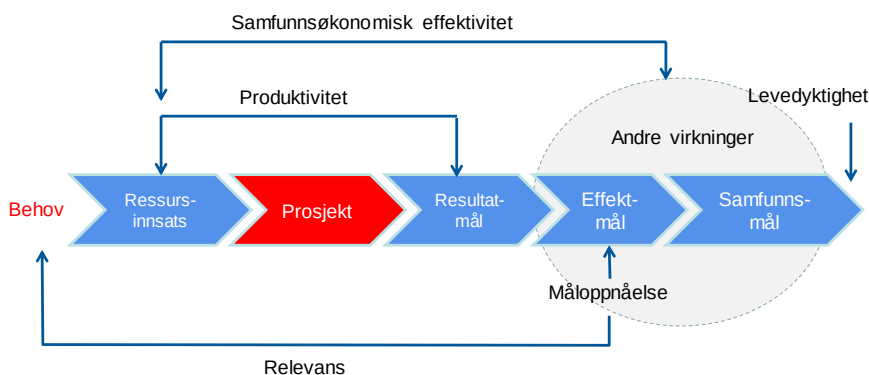
Det som behøves er en faglig vurdering, basert på kvantitativ og kvalitativ analyse samt evaluators eget skjønn, av i hvilken grad tiltaket er lønnsomt og viktige betingelser for dette. Ikke nødvendigvis et nøyaktig tall. Noen ganger er det ikke en gang faglig forsvarlig, eller interessant, å vite hva nyttesiden er, målt i kroner, for eksempel når tiltaket rettes mot ressurssvake grupper eller gjennomføres for å innfri en internasjonal forpliktelse. Da kan en heller

²² Den britiske ordningen med overordnet prosjektstyring (sammenlignbar med den norske KS-ordningen) har krav om såkalte gateway reviews ved en rekke faseoverganger. Herunder omfatter ordningen en review i driftsfase som spør om hvorvidt prosjektet ga *Value-for-money*, se www.gov.uk. Videre har man i England et eget system for etterevaluering av transportprosjekter (den såkalte POPE-modellen) som er en kombinasjon av lønnsomhet og målorientert evaluering, men med mest vekt på det første (Highways England, 2015).

²³ Se også Heinzerling og Ackerman (2002), Næss (2004) og Nyborg (2012).

beregne effektivitet på taktisk nivå (omtales gjerne som kost-effektanalyser eller formålseffektivitetsanalyse). Her finnes ulike analysetyper hvor en enten tar nytten for gitt eller anvender en felles fysisk måleenhet for effekt, som sammenholdes med kostnaden.

Modellen med de seks overordnede evalueringskriteriene er illustrert i figur 6.3.



Figur 6.3. Oversikt over de seks evalueringskriteriene i Concepts evalueringsmodell

Gjennomføringen av evalueringene

Datainnsamling og analyse – med små ressurser

Det er evalueringsteamets ansvar å disaggregere de overordnede evalueringskriteriene i mer konkrete og underliggende indikatorer - og tilhørende data og analysemetoder.

Det finnes et bredt spekter av metoder å velge mellom, se tekstboksen nedenfor for en grov oversikt.

En skiller mellom de kvantitative og de kvalitative. *Kvantitative* data er tallbasert, og vil typisk være mengder og frekvenser, resultater av modellberegninger, parametrisk statistikk etc. Kvantitativ analyse brukes til å belyse sammenhenger mellom variabler og til å generalisere funn på statistisk grunnlag. For å skaffe slik informasjon må evaluator følge klart definerte prosedyrer, for eksempel tilstrebe at et utvalg er statistisk representativt. Resultatet vil være avhengig av at dataene har en viss kvalitet, hvis ikke svekkes påliteligheten av konklusjonene. Det er lett å gå i den fellen at en får funnene til å virke mer

objektive og av høyere kvalitet enn de i virkeligheten er. Dataene må også være *valide*, det vil si at de er indikatorer på det de er ment å belyse.

Oversikt over viktige datainnsamlings- og analysemetoder:

Kvantitative

- Registerbaserte undersøkelser (for eksempel SSB)
- Formell spørreskjemaundersøkelse med statistisk representativt utvalg
- Andre egne målinger/registreringer
- Etterprøving av prissatte effekter i en nytte-kostnadsanalyse

Kvalitative

- Dokumentanalyse (prosjektdokumenter, tidligere evalueringer etc.)
- Direkte observasjon (for eksempel befaring, teste bruken av infrastruktur)
- Personlig intervju med nøkkelinformanter (ustrukturert eller semistrukturert), enkeltvis eller fokusgruppeintervju
- Case-studie

Kvalitative data er gjerne tekstlig informasjon i form av beskrivelser, observasjoner, sitater fra åpne spørreundersøkelser, intervjuer etc. Kvalitativ analyse kan for eksempel omfatte innholdsanalyse, analyse av casestudier, induktiv analyse og logisk og sekvensiell analyse. Resultatet er beskrivelser (mønstre, trender etc.) og fortolkninger og forklaringer av dette. Kvalitativ analyse bidrar til å gi bredde og dybde i det som evalueres og er helt nødvendig i tillegg til kvantitative funn. Utfordringen kan være at slike data er påvirket av evaluators subjektive vurdering og dermed i mindre grad etterprøvbare. Kvaliteten hviler i stor grad på datautvelgelse, perspektiv og metoder, samt på evaluators troverdighet mer generelt.

Samset (2014) gir en god oversikt over metoder²⁴. Mange lærebøker skisserer ambisiøse design, hvorav det ideelle er en før-og-etterstudie med kontrollgruppe. Dette er sjelden aktuelt i evaluering av prosjekter. Ulike typer

²⁴ Det finnes utallige lærebøker som går i detalj på datainnsamlings- og analysemetoder, ikke minst for å dokumentere kausale effekter, se for eksempel. Scriven (1991) eller Rossi m.fl. (2003). Boardman m.fl. (2011) er en mye brukt lærebok i samfunnsøkonomisk analyse, i tillegg til offentlig veiledning (Finansdepartementet, 2014, DFØ, 2014a).

kvasi-eksperimentelle opplegg kan være aktuelle hvor man «simulerer» en kontrollgruppe fremfor å trekke en tilfeldig²⁵. Knapphet på tid og ressurser gjør imidlertid at en ofte velger det mist ambisiøse, det vil si etterstudie uten kontrollgruppe, i noen tilfeller før-og-etter målinger. Evaluators egne observasjoner ved befaring, gjerne i samband med intervjuer, gir et helt nødvendig tilskudd til kontekstforståelsen, og er påkrevet i alle evalueringer.

Analysen som hovedsakelig bygger på kvalitativ informasjon kan også gi valide resultater. De bør bygge på et solid teoretisk grunnlag (Yin, 2013). Som nevnt: dersom tiltaket er fundert på en teori som allerede er godt dokumentert i tidligere studier, kan man lagt på vei bruke teorien til å argumentere for at det som observeres er en effekt (forutsatt at de kontekstuelle forholdene er sammenlignbare). Det er uansett viktig at kvalitativ informasjon suppleres med kvantitativ informasjon i den grad det er mulig.

Concepts evalueringer lyses typisk ut med et budsjett som gir begrensninger med hensyn til omfang og metodebruk. Å dokumentere tilsiktede effekter er krevende nok, i tillegg skal en beregne samfunnsøkonomisk effektivitet og gå inn i prosjektets gjennomføringsfase og vurdere prosesser, aktører og kostnader. Er det realistisk å gi en troverdig vurdering av alt dette? Erfaringene så langt tyder på at svaret er ja (se neste kapittel). Men det er viktig å påpeke at evalueringene skal gi en *overordnet* vurdering av prosjektets vellykkethet som går inn i en større database som *til sammen* genererer lærdommer til senere prosjekter. Det er derfor ikke detaljene i det enkelte prosjektet som er det sentrale.

Evaluering er ikke det samme som forskning og en kan ikke stille de samme kravene til vitenskapelighet (i betydningen bruk av forskningsbaserte metoder). I disse evalueringene må en ofte bruke *mindre formelle* metoder som avgrensede spørreundersøkelser med små utvalg, gruppeintervju fremfor individuelle intervju, og omfattende bruk av sekundærkilder²⁶. Det handler om å fokusere på hva som er ”godt nok” innenfor realistiske rammer, men likevel tilfredsstillende akseptable kvalitetskrav. Dette utelukker selvfølgelig ikke at vi i enkelte tilfeller kan utvide rammene for evalueringen eller igangsette et parallelt

²⁵ Et eksempel er Difference-in-difference metoden (se for eksempel HM Treasury, 2011a for en innføring i denne og andre kvasi-eksperimentelle design).

²⁶ Om formelle og uformelle metoder, se Samset (2014). Bruk av enkle og lite ressurskrevende evalueringsmetoder er også diskutert i blant andre Sunnevåg (2007) («beslutninger på svakt informasjonsgrunnlag»), World Bank (2004) («rapid appraisal methods») og Bamberger m.fl. (2004) («shoestring evaluation»). Veilederen til Statens vegvesen (2006) i etterprøving av lønnsomhet omtaler også forenklete analysevarianter.

forskningsprosjekt for å kartlegge bestemte prosesser, effekter, nye datatyper, etc. nærmere.

Uansett skal det stilles krav til å opprettholde en akseptabel faglig standard. Det gjelder for det første kravet til etterprøvbarehet (*reliabilitet*) som innebærer at informasjonen skal være så pålitelig at en uavhengig undersøkelse av samme fenomen ville kunne gi tilnærmet identisk resultat. For det andre kravet til informasjonens gyldighet (*validitet*). Det innebærer at informasjonen må være slik at den svarer på nettopp den problemstillingen som skal belyses. Det viktigste hjelpemiddelet en har for å sikre validitet er ved å anvende triangulering. Det betyr å studere et fenomen fra ulike vinkler, og bruke flere kilder eller teknikker for å få svar på de mest essensielle spørsmålene. Da kan man sikre et mer komplett datagrunnlag og trekke mer robuste konklusjoner.

Når det gjelder samfunnsøkonomisk effektivitet er metoden for sammenstilling gitt, men det er allikevel mange måter å utlede de ulike prissatte og ikke-prissatte virkningene som inngår. Det som er sagt over gjelder følgelig også for kriteriet samfunnsøkonomisk effektivitet. Vanligvis handler databehovet i samfunnsøkonomiske analyser mest om de fysiske effektene, som er felles med det en søker under måloppnåelse og andre virkninger, mens prisene kan antas å være uendret i tiden som er gått etter ex ante analysen²⁷.

Scoresetting

Etter at data er samlet inn og analysert skal evaluator gi en samlet vurdering av vellykkethet for hvert av de seks evalueringskriteriene. Konklusjonene vedrørende alle indikatorene under hvert kriterium skal aggregeres opp til en samlet score. Concepts opplegg innebærer at det skal benyttes en poengskala fra 1 til 6, hvor 1 er laveste score (helt mislykket) og 6 er høyeste (svært vellykket). I de tre første evalueringene opererte vi ikke med denne typen aggregert vurdering og scoresetting, det ble innført senere. Vi utarbeidet etter hvert et foreløpig utkast til retningslinjer for scoresettingen (Concept, 2015) som har vært førende for omtrent halvparten av evalueringene. Her var det for eksempel antydnet at score 4 tilsvarer et resultat som er «helt greit», dvs. akseptabelt, men ikke noe mer. Eksempelvis at prosjektet akkurat holder kostnadsrammen eller har lønnsomhet rundt null. Retningslinjene vil bli

²⁷ Verdsettingsstudier er ressurskrevende og gjennomføres sjelden som del av en enkelt analyse. I sektorer med mange samfunnsøkonomiske analyser har etatene gjerne et fast sett av priser på de vanligste virkningene. Disse kan være oppdatert siden ex ante-analysen, men hvis en skal bruke nye priser ex post bør en også oppdatere ex ante-analysen med disse. Unntaket er hvis det er fundamentale endringer i befolkningens preferanser som har medført prisendringene.

oppdatert som resultat av denne gjennomgangen. Skalaen må betraktes som ordinal og en grov måte å kategorisere resultater på. Hensikten er å kunne sammenligne ytelse mellom ulike kriterier og opp mot andre prosjekter. Resultatene er derimot ikke ment å måle relativ styrke på effektene, og vår bruk av resultatene til beregning av gjennomsnitt må derfor leses med forbehold om dette.

Ved å oppsummere vurderingen i form av score, vil evalueringsteamet måtte definere akseptable nivåer for resultatene og deretter vurdere resultatene opp mot denne standarden. Dette er en øvelse med store frihetsgrader og betydelig innslag av skjønn. Derfor er det viktig at det er transparens om hvilke hensyn og perspektiver som vektlegges. Som påpekt av Tornes (2013) er det i evalueringer ofte påfallende lite diskusjon om valg av verdier og kriterier, for eksempel hvorvidt evaluator velger å utelukkende ta utgangspunkt i andres mål og vurderinger eller om man også gjør egne vurderinger. For en samfunnsøkonom kan det synes åpenbart at alt må vurderes ut fra effektens samfunnsøkonomiske betydning (dvs. implisitt vektet med befolkningens betalingsvillighet)²⁸ men for andre er dette ikke selvsagt. Relevans er som nevnt særlig krevende idet det handler om ytelse i forhold til behov og prioriteringer. Dette kan vurderes på ulike måter: ekspertvurderinger, brukernes egne formuleringer av behov, observerte etterspørsel, betalingsvillighet, evaluators eller andres tanker om samfunnsnytte, tanker om at noen har «fortjent» et gode, politiske mål, etc. I mange tilfeller er det konflikter mellom ulike perspektiver. Den samme utfordringen gjelder ved vurdering av andre virkninger, hvor helt ulike virkninger skal vurderes mot hverandre. For produktivitet og måloppnåelse vil derimot de avtalte målene (evt. den justerte målstrukturen) være naturlige holdepunkter. Tilsvarende har man for lønnsomhet en naturlig målestokk hvor høyere lønnsomhet er entydig positivt. I praksis vil scoresetting for alle kriteriene innebære utøvelse av skjønn. Som påpekt i Scriven (2013) er den beste måten å sikre en balansert vurdering på, et bredt sammensatt team med ulike faglige perspektiver.

Chianza (2008) argumenterer for at en i tillegg burde definere minstekrav for enkelte kriterier, slik at dersom resultatet ikke oppfyller dette, så skal prosjektet de facto defineres som helt mislykket (her score 1). Chianza foreslår at dette vurderes for produktivitet, virkninger og levedyktighet. I det første tilfellet

²⁸ Imidlertid fant Bull-Berg m.fl. (2014) at mens prissatte virkninger i samfunnsøkonomisk analyse vurderes etter klare prinsipper (verdsetting i henhold til befolkningens betalingsvillighet) så er oppsummeringen av ikke-prissatte virkninger (på en skala fra ---- til +++) i praksis basert på ulike og ofte ikke-oppgitte verddivurderingsprinsipper.

handler det om å sette en grense for mye pengesløsning (budsjettoverskridelse) som kan aksepteres, muligens også en grense for mangel på etisk og miljømessig ansvarlighet i gjennomføringsfasen. I det andre tilfellet om hvor store negative sidevirkninger som er akseptabelt (for eksempel på miljø eller svake grupper). Og i det siste tilfellet om hvor raskt en kan godta at effekten forsvinner. Dette er kanskje særlig relevant i ex ante evaluering, hvor det vil innebære ikke-aksept for oppstart av prosjekter som ikke oppfyller kravene (omtrent tilsvarende som absolutte krav i en KVV).

Teamsammensetning

Alle Concept-evalueringer må følge internasjonalt anerkjente kvalitetsstandarder, som for eksempel oppsummert i OECD (1991 og 2010). Det handler om kvalitet og nytteverdi, i tillegg til evaluators troverdighet, uavhengighet og transparente (etterprøvbare) resultater. Dette gir føringer for hvilken type kompetanse vi må etterspørre og andre krav som stilles til et evalueringsteam. Det impliserer ikke minst at en trenger flere personer i temaet (normalt minst tre).

For det første krever vi *uavhengighet* til prosjektet som skal evalueres. Det betyr at et evalueringsmiljø som for eksempel har bistått med KVV, kvalitetssikring eller liknende ikke kan involveres. Det samme kan gjelde andre typer bindinger til departementet eller etaten som eier prosjektet. Vi har også vurdert å utelukke miljøer som har rammeavtale med Finansdepartementet om ekstern kvalitetssikring, siden det er å vente at KS-ordningens bidrag til prosjektsuksess kan bli et element i evalueringen. Men de evalueringsfaglige miljøene i Norge er få og små (dessuten måtte en da antakelig ha utelukket bruk av egne forskere også, siden Concept er finansiert av Finansdepartementet). Erfaringen til nå, er at også evaluatorene fra KS-miljøene (som har vært involvert i seks av 20 evalueringer) later til å ha den nødvendige profesjonelle avstand til det som analyseres. En bør likevel være oppmerksom på hvilken innvirkning bruk av KS-miljøer som evaluatorene kan ha – eller oppfattes å ha.

For det andre er det helt nødvendig med *evalueringsfaglig kompetanse*. Ifølge Scriven (2013) bør evaluatorene ha kjennskap til både kvantitative og kvalitative metoder fra samfunnsvitenskapen, herunder er casestudie-metoder særlig relevante for evaluering av enkeltprosjekter og -programmer. I vårt tilfelle vektlegges spesielt kjennskap til logisk prosjektanalyse eller programteori som underlag for evalueringen, og OECD-kriteriene og samfunnsøkonomisk analyse for å vurdere suksess. Videre trengs en viss prosjektfaglig kompetanse,

både det som gjelder kostnadsestimering av prosjekter, prosjektgjennomføring, og overordnet prosjektstyring fra et eierperspektiv.

For det tredje er det nyttig med en viss *sektorkunnskap* – nøyaktig hvilken vil variere med prosjektet. Det handler om å kjenne sektoren, prosjektet og konteksten (lokalkunnskap). Dette er nyttig blant annet for å vite hvilken informasjon en bør etterspørre, samt muligheten for å kommunisere med ulike interessenter på en god måte. På den annen side vil sektorkunnskap også kunne virke avgrensende, for eksempel på evnen til å oppdage utilsiktede virkninger ut over egen sektor. Denne typen kompetanse bør derfor ikke være dominerende.

7. Erfaringer med evalueringene

I dette kapitlet presenterer vi en oppsummering og vurdering av hvordan evalueringene er gjennomført og det metodiske rammeverket anvendt. Alle rapportene er grundig gjennomgått med utgangspunkt i en sjekkliste med kritiske spørsmål (vedlegg 3). Spørsmålene er fremkommet basert på relevant litteratur (kapittel 6) og ikke minst de tidlige erfaringene fra pilotevalueringene (Volden og Samset, 2013). Det handler om forhold en forventet kunne være problematiske, som håndtering av mangelfulle målstrukturer og uklar tolkning og avgrensning av kriteriene.

Mot slutten av arbeidet ble det holdt et fokusgruppemøte med 11 erfarne ressurspersoner fra ulike evalueringsmiljøer, som alle har deltatt i gjennomføringen av evalueringene [se vedlegg 4]. Disse ble konfrontert med foreløpige funn og vurderinger, og fikk for øvrig anledning til å dele sine erfaringer og vurderinger av metoden, opplegget og behovet for veiledning.

Forfatterne har også trukket på egne erfaringer fra prosessen med å bestille, følge opp og kvalitetssikre alle de 20 evalueringene samt fra deltakelse i tre av dem.

Lærdommene etter denne runden vil resultere i et revidert opplegg med tydeligere krav til hvordan evalueringene skal gjennomføres, evalueringsteamets kompetanse, og nærmere veiledning til hvordan modellen og kriteriene skal forstås.

Generelle erfaringer

Erfaringene med modellen er generelt gode, men det er også en del tilfeller hvor faktagrunnlaget som vurderingene er bygd på er for svakt. Prosessen med å disaggregere evalueringskriteriene i konkrete indikatorer og deretter aggregere funnene opp for å gi kvalifiserte svar på evalueringskriteriene, gir en god balanse mellom behovet for standardisering og fleksibilitet. Vi har registrert noen utfordringer knyttet til sammensetningen av team, tolkningen av de taktiske og strategiske kriteriene samt scoresettingen. Dette viser at det er behov for klarere retningslinjer fra Concepts side.

Evalueringene har i stor grad funnet en felles form. De fleste holder også akseptabel kvalitet. De gjøres med begrensede ressurser og vil uansett bare kunne tilby en overordnet vurdering av prosjektene. Denne gjennomgangen tyder på at resultatene stort sett er velinformerte, med noen unntak. Over tid

vil de gi grunnlag for robuste konklusjoner om hva som påvirker statlige investeringsprosjekters vellykkethet. De vil også kunne avdekke forhold som krever nærmere gransking, for oppfølging av departement eller etat.

Det later til å være bred enighet om at den generiske modellen er godt egnet for formålet, og til sammen beskriver de viktigste aspektene ved prosjektsuksess. Evaluatorene påpeker at særlig de strategiske kriteriene (andre virkninger, relevans og levedyktighet) er bevisstgjørende. Ved å bruke den samme generiske metodikken på mange prosjekter åpner det seg interessante analysemuligheter på tvers. Prosessen med systematisk disaggregering og aggregering av informasjon gir samtidig evaluator frihet til å tilpasse evaluering for hvert enkelt prosjekt.

Gjennomgangen viser også at det er noen utfordringer som går igjen. En av disse er at prosjektene har et mangelfullt referansegrunnlag²⁹. Blant annet er målstrukturen ofte uegnet for evaluering. Evaluatorene håndterer dette på ulikt vis, og ikke alle gjør den nødvendige vurderingen av målstrukturen. Man må finne en avklaring mellom det formelt vedtatte og hvilke rammer som gjør det mulig å evaluere prosjektet.

Vi ser også enkelte utfordringer knyttet til tolkningen av de strategiske kriteriene og forholdet mellom dem, herunder avgrensningen mellom OECD-modellen og samfunnsøkonomisk analyse. Mange av evalueringsteamene har hatt et flertall av samfunnsøkonomer, noe som kan ha medført at de legger samfunnsøkonomiske tolkninger inn i de øvrige kriteriene også. Volden og Samset (2013) beskriver hvordan enkelte av pilotevalueringene lot lønnsomhetskriteriet overskygge de andre. Dette er fremdeles en utfordring.

Datatilgangen er sjelden så god som man kunne ønske. Det er ressurskrevende å samle inn data, slik at man må avgrense og prioritere strengt. Videre møter evaluator ikke alltid vilje eller mulighet i sektoren til å gi tilgang til eksisterende materiale og generell statistikk, samt at innsamlingsrutiner, datasystemer osv. kan være endret over tid. Mange evaluatorene benytter triangulering for å verifisere data, men ikke alle. Enkelte later til å søke hovedsakelig etter såkalt objektive og faktabaserte kilder. Faren kan bli at de strategiske kriteriene får en knapp og overfladisk behandling.

Scoresettingen er en vesentlig del av evalueringen. Den representerer en sammenfattende aggregering av evaluators funn og vurderinger på øverste nivå. Erfaringen er at bruken av score er nyttig, ikke minst for å kunne

²⁹ Ingen av prosjektene hadde gjennomgått KVV/KS1

gjennomføre sammenlignende analyser. Men vi ser også at det er behov for bedre kalibrering, og flere av evaluatorene sier de ønsker bedre veiledning til scoresettingen.

Evalueringsobjektet

Avgrensning til prosjektet (ikke tiltaket) og evaluering +/- 5 år ut i drift, er et godt utgangspunkt som gjør evalueringen håndterbar. Prosjektet må ses i sin kontekst. Evalueringene må få frem mer informasjon om både tilblivelsen og tidligfasen.

Evalueringsobjektet avgrenses til selve prosjektet slik det er definert ved KS2. Det er hensiktsmessig fordi det gjør evalueringen håndterbar og de ansvarlige for gjennomføringen kan identifiseres. Men flere evaluatorene er inne på at det er krevende å måle effekter der resultatet åpenbart avhenger av mer enn det ene prosjektet. Det gjelder for eksempel bygg- og IKT-prosjekter som ofte forutsetter organisatoriske og ledelsesmessige grep internt i brukerorganisasjonen. Prosjektet må derfor alltid studeres i sin kontekst, og alle begrensninger og forbehold må beskrives. De mest erfarne evaluatorene har løst dette på en god måte.

I tråd med erfaringene fra pilotevalueringene samt anbefalingen i Andersen m.fl. (2007) har vi valgt prosjekter som har kommet +/- 5 år ut i driftsfasen. Evaluatorene melder at timingen stort sett har vært god men at man i noen tilfeller kunne ha ventet enda lengre. Det gjelder for eksempel Halden fengsel, hvor evaluator påpeker at *når effekten handler om å unngå tilbakefall for innsatte med lange straffer, sier det seg selv at tidsdimensjonen i måloppnåelsen må være svært lang*. Det var allikevel mulig å gi en samlet vurdering av effekten. Venter man for lenge blir utfordringen at tilgangen til nøkkelinformanter blir dårligere, noe som blir påpekt i flere evalueringsrapporter. Sentrale personer kan ha skiftet jobb eller glemt viktige aspekter ved prosjektet.

I noen tilfeller avsluttes ikke prosjektet med overlevering til drift. Økonomiske etterspill i form av for eksempel rettssaker kan resultere i at sluttrapport og sluttkostnad ikke foreligger når prosjektet skal evalueres. Det kan også finnes andre grunner til at prosjektkonti ikke lukkes. I slike tilfeller bør evaluator basere seg på etatens avsetninger eller estimerer på det som går ut over evalueringstidspunktet.

Mange av prosjektene har en lang tidligfase. Vår vurdering er at der evaluator gir en grundig beskrivelse av prosjektets tilblivelse, involverte aktører, maktforhold og beslutninger, så er dette nyttig. For eksempel, dersom konseptvalget var basert på analyse og faglig vurdering, gir dette i seg selv en

indikasjon om relevans, sammenlignet med der konseptvalget var resultat av lokalpolitisk maktkamp eller stivhengighet. Dersom en vet at beslutningstakerne forholdt seg til et realistisk kostnadsestimat da konseptvalget reelt sett ble tatt, er det mer sannsynlig at løsningen er både relevant og formålseffektiv enn dersom konseptvalget var basert på et langt lavere (og kanskje urealistisk) kostnadsestimat.

De færreste av evalueringsrapportene omtaler imidlertid prosjektenes tidligfase og få gir en god oversikt over utviklingen i kostnadsestimatet gjennom tidligfasen og forprosjekt. Konfrontert med dette svarer evaluatorene at de ikke gjorde dette fordi de ikke spesielt ble bedt om det. De ser åpenbart nytten av å gå lenger tilbake for å forstå prosjektenes tilblivelse, men har noen bekymringer knyttet til ressursbruken det vil kreve.

Tabell 7.1 Evalueringsobjektet og dets beskrivelse i evalueringen. Vår vurdering

Evalueringsobjektets kjennetegn	Ja	Nei	Delvis eller ukjent
Prosjektet er klart avgrenset	20	0	0
Grensesnitt mot andre prosjekter	13	2	5
Tidligfasen er godt beskrevet	4	9	7
Første kostnadsestimat er oppgitt	4	14	2

Evalueringsteamet

Mange av teamene har hatt et flertall av samfunnsøkonomer. Større bredde i teamet er nødvendig og en må sikre tilstrekkelig evalueringserfaring.

Som nevnt i forrige kapittel bør evalueringsteamet ha den nødvendige faglige bredde, i tillegg til uavhengighet og habilitet. Faglig bredde er viktig for å unngå en skjev tilnærming og mangelfull vurdering av prosjektet. En konklusjon fra pilotevalueringene var at teamene var for homogene. I ettertid har vi forsøkt å tilstrebe tverrfaglighet uten å helt lykkes. Alle unntatt ett team har tyngst erfaring med samfunnsøkonomisk analyse, mens evalueringserfaring, prosjektkompetanse og sektorkompetanse er mer varierende. I noen tilfeller har de samfunnsøkonomiske miljøene knyttet underleverandører til seg for å dekke kompetansemessige hull, men disse fagressursene har ofte i realiteten har hatt en perifer rolle.

Sett under ett er det vår vurdering at kun 12 av de 20 evalueringene er gjort av team med tilstrekkelig bred kompetanse og erfaring som dekker både evaluering, samfunnsøkonomi, prosjektstyring og det sektorfaglige.

Evaluatorene er til dels enige. Noen påpekte at evaluering ikke er en egen profesjon i Norge og at mange samfunnsøkonomer har slik dobbeltkompetanse. Men andre minnet om at samfunnsøkonomisk nytte bare er ett av flere aspekter ved behov, og at *det blir lett til at man fokuserer på det man kan mye om og mindre på det som er perifert i forhold til eget kompetanseområde – derfor må teamet være bredt sammensatt*. Det var videre litt uenighet om hvor viktig prosjektkompetanse er – noen mente at dette *må* være representert i teamet, mens andre mente at det holder å ha *tilgang* til slik kompetanse. Det var enighet om at sektorkompetanse ikke må dominere, samtidig som det kan være nødvendig for å forstå «stammespråket». Regionfelt Østlandet ble nevnt som eksempel – der var det avgjørende at en hadde militærfaglig kompetanse med i teamet. Andre ganger trengs det ikke.

Tabell 7.2 Evalueringsteamet. Vår vurdering

Evalueringsteamet	Ja	Nei	Delvis eller ukjent
Samfunnsøkonomi	19	1	0
Prosjektfag	10	8	2
Sektorkompetanse	12	6	2
Evalueringserfaring	10	6	4
<u>Tilstrekkelig bred kompetanse og erfaring</u>	<u>12</u>	<u>1</u>	<u>7</u>
Uavhengighet og habilitet	20	0	0
Tilknytning til KS-ordningen	6	14	0
Følgereforskere fra Concept involvert	8	11	1

Spørsmålet om habilitet er også sentralt. Seks av evalueringene har vært ledet av miljøer som også har rammeavtale med Finansdepartementet om KS-oppdrag. Vi har ikke grunnlag for å si at disse skiller seg fra de øvrige evalueringene hva gjelder fokus, scoresetting, eller omtale av kvalitetssikrers rolle i prosjektene som evalueres. Men det er desto viktigere at alle vurderinger i evalueringene er transparente.

Vi har også notert (ikke vist i tabellen) det store antallet evaluatorene som har vært involvert: 17 miljøer (i tillegg til NTNU) og 47 personer. Dette har delvis vært tilsiktet i denne utprøvingsfasen, men har også medført at den kumulative læringseffekten blant evaluatorene har vært begrenset. Av de 35 eksterne ressurspersonene har kun syv (17 %) deltatt i mer enn en evaluering. Blant de 12 følgeforskerne fra Concept har en lagt noe mer vekt på erfaringsoverføring og læringseffekter – her har syv (58 %) deltatt i mer enn en evaluering. Kapasiteten har imidlertid satt begrensninger på omfanget av involvering av følgeforskere. Dette støtter opp om viktigheten av gode skriftlige retningslinjer som nye evaluatorene skal forholde seg til prosjektets resultatkjede og målhierarki.

Målhierarkiet har ofte mangler, og prosjektene bygger ikke nødvendigvis på en tydelig formulert programteori. Likevel er det få evaluatorene som gjør en grundig vurdering av dette. Mange gjør mindre justeringer i målformuleringene for å gjøre effektene mer målbare.

Prosjektene skal vurderes i forhold til formelt vedtatte mål. Problemet er at dersom disse er urealistisk ambisiøse vil vurderingen uansett bli negativ. Tilsvarende, dersom målet er trivielt vil prosjektet uansett måtte vurderes som vellykket.

Ikke uventet er prosjektenes resultatkjede og målstruktur ofte mangelfull, til dels helt uegnet som utgangspunkt for evaluering. Mange prosjekter mangler for eksempel samfunns mål. Kun én evaluering har vurdert den eksisterende målstrukturen og funnet den godt egnet (PERFORM). Evaluators vurderinger av målstrukturen er imidlertid sjelden særlig omfattende, med noen få unntak. Det gjelder blant annet to byggeprosjekter hvor evaluator i begge tilfeller pekte på at forventningene om synergieffekter mellom brukerne i bygget var tuftet på uklare og til dels urealistiske forutsetninger. I det ene tilfellet hadde politiske myndigheter nærmest lagt samlokalisering som en premiss for å utløse byggemidler.

Noen evaluatorene bruker de opprinnelige målene slik de står, også der de har store mangler. Andre gjør mindre justeringer (*vi ryddet litt opp i målstruktur og formuleringer*), mens atter andre reetablerer resultatkjede og målstruktur basert på ulike bakgrunnsdokumenter (herunder formålsbeskrivelser i de ansvarlige etaters styrende dokumenter). På workshopen kom det frem en grunnleggende uenighet om hvorvidt det er riktig av evaluator å justere målstrukturen. Flere ga uttrykk for at *det føles feil å skulle tukle med målstrukturen*. Det er både tidkrevende og vanskelig og har ikke nødvendigvis forankring hos prosjekteier. Andre mente at målstrukturen ofte er så mangelfull at en ikke har noe valg.

Det var enighet om at selv om evaluator ikke foreslår konkrete endringer, så bør en tydeliggjøre prosjektets resultatkjede, synliggjøre forutsetninger og påpeke urealistiske eller trivielle mål. En eventuell justert målstruktur må presenteres som evaluators egen tolkning, da det er urealistisk å forvente en formell godkjenning fra prosjekteier.

Tabell 7.3 Prosjektenes resultatkjede og mål, og evaluators vurderinger av dem. Vår vurdering

Resultatkjede og mål	Ja	Nei	Delvis eller ukjent
Effekt- og samfunnsmål finnes	11	6	3
Klar og logisk resultatkjede	1	7	12
Riktig ambisjonsnivå for målene	3	7	10
Evaluator justerer målene	7	8	5

Det må videre påpekes at en har avdekket relativt betydelige målkonflikter i mange av prosjektene. Den enkle lineære kjeden av mål på ulike nivåer, som bygger opp om ett enkelt samfunnsmål, gir dermed ikke alltid et realistisk bilde av de norske investeringsprosjektene. En tilstreber likevel å beskrive prosjektene på denne måten og håndterer i stedet andre viktige samfunnsmål som *krav* til løsningen eller sideeffekter. Prosjektets taktiske og strategiske vellykkethet vil da typisk hvile på at man har greid å finne en tilfredsstillende balanse mellom ulike hensyn.

Evalueringskriteriene

Rammeverket med de seks kriteriene skal gi en bred og overordnet vurdering av prosjektet. Det forutsetter at alle kriteriene vies tilstrekkelig oppmerksomhet og at de forstås og operasjonaliseres likt. I praksis ser vi en tendens til at enkelte evaluators fokuserer uforholdsmessig mye på de lettest målbare kriteriene. Målt ved fordelingen av sidetall i rapporten, så vies i snitt en fjerdedel av rapporten til produktivitetskriteriet alene. Det kan kanskje begrunnes med at dette ene kriteriet skal dekke operasjonell prosjektsuksess. Men samtidig ser vi at enkelte av rapportene gjengir unødvendig mange detaljer fra selve prosjektet, kanskje nettopp fordi de var lett tilgjengelige. Også samfunnsøkonomisk effektivitet vies ofte stor oppmerksomhet, selv om det normalt er en forenklet beregning som gjøres. Den samme iveren kan ikke alltid spores når det gjelder relevans og levedyktighet. Dette er de kriteriene det

brukes minst ressurser på, noe som er uheldig fordi det er helt essensielle indikatorer på om en lykkes strategisk – det handler om hvorvidt behovet var/er tilstede og om nytteeffektene vedvarer over tid.

Tabell 7.4 Evalueringenes relative ressursbruk og tolkning av kriteriene. Vår vurdering

Kriteriene	Fordeling sidetall i rapporten	Evaluatorenes tolkning
Produktivitet	24 %	Lik tolkning. Tid, kostnad, kvalitet. Tids- og kostnadseffektivitet.
Måloppnåelse	19 %	Lik tolkning. Første ordens effekt for brukere. Henger mye på godheten av målstrukturen. Uformelle analyser.
Andre virkninger	17 %	Varierende tolkning. Ikke alle avgrenser seg til utilsiktede virkninger. Delvis sammenblanding med andre kriterier.
Relevans	13 %	Varierende tolkning og ofte knapp vurdering. Delvis sammenblanding med andre kriterier.
Levedyktighet	12 %	Varierende tolkning og ofte knapp vurdering. Delvis sammenblanding med andre kriterier.
Samfunnsøkonomisk effektivitet	14 %	Lik tolkning, men krevende i praksis, pga. manglende data og referansegrunnlag. Forenklete analyser er troverdige nok.

En annen utfordring er den store variasjonen i hvordan de strategiske kriteriene tolkes. Evaluatorene bekrefter at de er usikre og ønsker mer veiledning. Enkelte later til å legge et samfunnsøkonomisk verdivurderingsperspektiv til grunn for disse tre strategiske kriteriene. Resultatet av dette kan bli at de målorienterte kriteriene i virkeligheten bare blir elementer i en samfunnsøkonomisk analyse, og man får *ikke* de to separate tilnærmingene som var tiltenkt. Igjen vil vi påpeke viktigheten av et tverrfaglig evalueringsteam.

Nærmere om tolkningen av de enkelte kriteriene under:

Produktivitet

Produktivitet er definert likt på tvers av evalueringsteamene, men det er behov for felles retningslinjer når det gjelder hvilken kostnadsramme en sammenligner med og hvordan denne skal prisjusteres. Det er også behov for en sjekkliste med informasjon som bør tas med om gjennomføringsfasen.

Alle evalueringene fokuserer på de tre resultatmålene for tid, kostnad og kvalitet slik de er definert i styringsdokumentet og godkjent av Stortinget. I fem tilfeller ble imidlertid prosjektet fremmet for Stortinget på nytt etter det opprinnelige vedtaket og fikk en høyere ramme. Tre evalueringer brukte da den oppdaterte rammen som referanse mens to holdt på opprinnelig rammen. Dette er uheldig da det gir helt forskjellige resultater. I henhold til Concepts praksis er det rammen i det *opprinnelige* Stortingsvedtaket som skal være referansegrunnlaget, se Welde (2017). Hvis evaluator velger noe annet må det redegjøres godt for årsaken til dette.

Det varierer også hvilken prisindeks som benyttes for å fremskrive rammene. De fleste bruker den etatsspesifikke indeksen, men én bruker konsumprisindeksen, og ett prosjekt, Operaen, skiller seg ut ved at det fikk godkjent en egen prisindeks som innebar økninger på inntil 15 % per år. Dette har stor betydning for resultatet. Alle indekser som gir en økning ut over konsumprisindeksen innebærer i realiteten en omprioritering av midler fra andre sektorer og til det aktuelle prosjektet. Vår vurdering er at de sektorspesifikke indeksene må aksepteres, men konsumprisindeksen bør alltid brukes som en følsomhetsanalyse.

Evaluator skal videre svare på om prosjektet er gjennomført effektivt og uten å sløse med ressursene. Noen forsøker å identifisere et referansegrunnlag for denne vurderingen, men de fleste antar uten videre at rammene var satt optimalt. Det kan være gode grunner til å anta dette, ikke minst at prosjektet har gjennomgått KS2. Men på den annen side er det mange forhold (teknologiske, markedsmessige, etc.) som kan være endret etter KS2. Når det gjelder dobbeltspor Stavanger-Sandnes, som hadde en stor overskridelse, går evaluator langt i å antyde at dette var utslag av en ekstraordinær markedssituasjon som ikke kunne forutses i forkant. Motsatt antyder evaluator i et par andre prosjekter at rammen var romslig gitt markedsforskningene.

Rapportene bruker ofte mye plass på produktivitetskriteriet, men skriver lite om organisering og styring, som for eksempel hvilken kontraktstrategi som ble valgt og hvorvidt det ble brukt styringsgruppe. Betydningen av kvalitetssikring, herunder om rådene fra KS2 er fulgt, er også påfallende lite omtalt.

Måloppnåelse

Måloppnåelse tolkes likt av alle, som første ordens effekter for brukerne. Dette belyses ved hjelp av en kombinasjon av kvalitative og kvantitative data. Flere evaluatorene kunne hatt nytte av å kombinere med en teoribasert tilnærming.

Alle evalueringene tolker måloppnåelse slik de skal, som første ordens effekt for brukerne. For veiprosjekter er det gjerne reisetid og ulykkesfrekvenser, og for et byggeprosjekt kan det være trivsel og produktivitet blant brukerne av bygget. Spørsmålet om kontroll for eksterne faktorer er relevant, men ingen av evaluatorene har håndtert dette på formelt vis eller med høyt presisjonsnivå. De fleste velger enkle evalueringsdesign hvor effekten sannsynliggjøres ved bruk av uformelle metoder. En søker gjerne tidsseriedata i den grad det foreligger på tilstrekkelig disaggregert nivå, og supplerer med intervjuer og andre kvalitative data.

Noen få har forsøkt seg på en mer omfattende komparativ analyse. Evalueringen av Halden fengsel er et eksempel, hvor en studerte tilbakefallsprosent både for innsatte fra det aktuelle fengselet og en gruppe andre, sammenlignbare høysikkerhetsfengsler. På grunn av kort tidsserie og mangelen på gode før-data var det likevel ikke mulig å konkludere basert på disse tallene alene. Det viser at en er nødt til å være edruelig når det gjelder å dokumentere effekter kvantitativt, samtidig som evaluator kan gi et nyttig bidrag ved å skissere et mulig opplegg for slik evaluering, for videre oppfølging i etaten og som innspill til hva en bør utrede og samle data om i før-evalueringer i fremtiden.

Vi observerer at det er ingen som har gått særlig detaljert inn på *hvordan* effekter skapes i møte med brukerne og deres respons på tiltaket. Dette hadde trolig vært en nyttig tilnærming i enkelte prosjekter, i det minste som en form for triangulering.

Andre virkninger

Kriteriet tolkes ulikt og blandes tidvis sammen med andre kriterier. Det er kun ikke-tilsiktete virkninger som skal dokumenteres og vurderes. Det gjelder alt som kan tilbakeføres som resultat av prosjektet, ut over første ordens effekter (Effektmålet). Og det skal ikke være en økonomisk analyse.

Evaluatorene har til dels ulike tilnærminger til hva de tar med under virkningskriteriet. Noen definerer kriteriet til å handle om samfunnsmålet eller det de anser som den «egentlige hensikten» bak prosjektet. Men dette er avgrenset til positive virkninger, og dekker ikke nødvendigvis utilsiktete negative virkninger som kan tilbakeføres som resultat av prosjektet. Andre har løst oppgaven ved å gjennomføre en fordelingsanalyse, som viser systematisk hvordan ulike grupper (herunder målgruppen) kommer ut. Dette kan lett medføre dobbelttelling da det også dekker effekter som allerede er vurdert under måloppnåelse.

Enkelte evaluatorene tolker oppgaven som en økonomisk analyse, hvor en først kartlegger virkningene og deretter balanserer dem (gjærne kvalitativt) mot hverandre i henhold til økonomiske verdivurderingsprinsipper. De vil da typisk omtale såkalt «mernytte» (produktivitetsgevinster) i positive ordelag mens rene omfordelingsvirkninger ikke tillegges vekt, og negative miljøeffekter nevnes kort men vektlegges ikke ved scoresetting på samme måte som de ville slått minimalt ut på netto nåverdi grunnet lav pris. Dette vil lett medføre dobbelttelling i forhold til lønnsomhet.

Faktisk er det knapt halvparten av evalueringene som tolker andre virkninger slik det var tenkt fra vår side, nemlig en oversikt over utilsiktede virkninger av prosjektet i et bredt perspektiv. Det kan handle om alt fra uheldige fordelingsvirkninger og miljøulempen, til incentiveffekter av selve investeringstiltaket, eller også om utilsiktede positive virkninger på miljø, økonomi og samfunn. Det finnes ingen felles måleskala eller standard for hvordan de skal vurderes samlet. I mange tilfeller kan det være behov for å ta forbehold samt påpeke målkonflikter og usikkerhet.

Relevans

Relevanskriteriet oppleves som krevende og får ofte en knapp (og kanskje ukritisk) behandling. Vurderingen her skal dreie seg om hvorvidt virkningene av prosjektet er i samsvar med samfunnets behov, uttrykt i form av politiske mål, interessentenes prioriteringer, og etterspørsel. En må også spørre om løsningen er den mest relevante i et tenkt mulighetsrom.

På samme måte som med andre virkninger har ulike evaluatorene vurdert relevans på ulike måter. Flere gir uttrykk for at relevans er det vanskeligste kriteriet, blant annet fordi det finnes lite kvantitativ dokumentasjon og fordi relevansen kan være omstridt. Mange evaluator ønsker ikke å fremstå som overdommer eller blande inn egne verdivurderinger. Utfordringen er at en neppe kommer utenom skjønnsutøvelse. Vurderingen blir på mange måter motsatsen til det å skulle vurdere om «behov» er godt begrunnet i en Konseptvalgutredning.

Noen ender opp med en kort utsjekk av om målgruppen er tilfreds. Andre tar utgangspunkt i ett av de tre anbefalte behovene: (1) overordnet politisk, (2) etterspørselsbasert eller (3) interessegruppebasert. Ofte velger man det som ligger nærmest effektmålet. Resultatet av dette blir at man i veiprosjekter vektlegger behovet for fremkommelighet (etterspørsel) mens man i jernbaneprosjekter vektlegger behovet for en omlegging til klima- og

miljøvennlige transportformer (politiske mål) – dermed får begge typer prosjekter god score på relevans, mens de grunnleggende målkonfliktene i transportsektoren blir skjøvet under teppet. I andre prosjekter (for eksempel bygg) vektlegges ofte lokale hensyn (det interessegruppebaserte perspektivet) uten å problematisere hvordan dette kan stride mot begge de to andre perspektivene.

Det er kun fem av de 20 evalueringene som vurderer relevansen systematisk opp mot alle de tre perspektivene og i tillegg identifiserer og drøfter konflikter mellom dem. Vår vurdering er at nasjonale politiske mål har en særlig status – det er tross alt staten som investerer og da bør en vektlegge selve begrunnelsen for at *staten* skal engasjere seg, ikke hva markedet eller lokalt forvaltningsnivå ville ha gjort. For øvrig må alle de tre perspektivene tas i betraktning, og et prosjekt som er relevant i alle de tre perspektivene bør ha høyere score enn et som er relevant i kun ett eller to perspektiver.

I tillegg kommer det kontrafaktiske spørsmålet om andre alternative løsninger som kunne tenkes i forhold til de aktuelle behovene. Dette drøftes i liten grad i evalueringene. Spesielt er det få som diskuterer relevans opp mot nullalternativet eller enklere/billigere måter å løse problemet på.

Levedyktighet

Også levedyktighet tolkes på forskjellig måte og vurderes ofte knapt. Evaluator beskriver trender, men perspektivet er ikke alltid bredt nok. Ikke alle vurderer robusthet overfor endrete behov eller forutsetninger.

Omtalen av levedyktighet er ofte kort og består gjerne av noen tanker om trender og prognoser for de tilsiktede virkningene. Anbefalingen i Lædre m.fl. (2012) tilsier imidlertid en bred vurdering hvor en sjekker at prosjektets utvikling er akseptabel både økonomisk, miljømessig og i et samfunnsmessig perspektiv. Vi ser eksempler på at levedyktighet blandes med lønnsomhet, ved at det argumenteres med at negative sideeffekter ikke har nevneverdig betydning fordi den neddiskonterte kroneverdien i dag er liten. Dette er misforstått – levedyktighet handler nettopp om de langsiktige virkningene, også for neste generasjon.

Når den gjelder den økonomiske/finansielle levedyktigheten, så er det vanligvis et spørsmål om hvorvidt drifts- og vedlikeholdskostnadene har en akseptabel utvikling. Få prosjekter har en inntektsside, så dette er midler som må bevilges år for år. I tilfellet Operaen gir billettsalget en viss egeninntekt, som har økt i årene etter åpningen. Men dette er marginalt og driftstilskuddet har økt enda

mer, også på grunn av økende pensjonsforpliktelser. Isolert sett er den finansielle levedyktigheten derfor svak, men evaluator velger i dette tilfellet allikevel ikke laveste score fordi man legger til grunn at regjeringen sannsynligvis vil redde institusjonen og bevilge det som behøves. For en sentral kulturinstitusjon som Operaen er det kanskje et realistisk scenario, men det endrer ikke på problemet med lav økonomisk/finansiell levedyktighet.

Ikke alle evalueringene drøfter hvordan levedyktigheten kan endre seg ved viktige milepæler – i veiprosjekter gjelder det for eksempel når bompengebetalingen faller bort, typisk etter 15 år. I andre tilfeller kan det handle om hvorvidt og når andre, relaterte prosjekter blir realisert. Kun seks evalueringer drøfter eksplisitt robustheten overfor endringer i sentrale forutsetninger, til tross for at flertallet av prosjektene har grensesnitt mot andre prosjekter og tiltak som vist tidligere.

Samfunnsøkonomisk effektivitet

Kriteriet er klart definert, men referansegrunnlaget er ofte mangelfullt og nytten vanskelig å kvantifisere. De fleste evaluatorene velger derfor en forenklet analyse. Resultatet er usikkert, men vurderes som troverdig nok.

Til forskjell fra de tre foregående kriteriene er samfunnsøkonomisk effektivitet klart definert og forstått. En kan i prinsippet velge to nivåer av effektivitetsvurdering, lønnsomhet og det vi kan kalle formålseffektivitet. Nesten alle evalueringene har lagt til grunn den mest ambisiøse tolkningen, nemlig lønnsomhet. Kanskje litt overraskende da mange har store utfordringer med å gjennomføre en slik analyse. Evalueringen av Svalbard forskningspark er et av unntakene – der gikk ikke evaluator inn på den økonomiske verdien av forskning og utdanning, men tok for gitt at Norge skal ha dette og diskuterte i stedet kostnadsnivået på Svalbard relativt til fastlandsuniversitetene.

Referansegrunnlaget er nesten alltid mangelfullt. Selv der hvor evaluator har hatt tilgang til en før-analyse er den lite transparent. Analyser av lønnsomheten i andre prosjekter er av samme grunn av begrenset verdi og trekkes sjelden frem som benchmark. I årene fremover, når en skal evaluere prosjekter som har gjennomgått KVV/KS1, vil tilgangen til samfunnsøkonomiske analyser forhåpentlig være langt bedre. Men Kvalheim (2015) fant at både KVV-er og KS1-rapporter har forbedringspotensial når det gjelder dokumentasjon av forutsetninger og begrensninger, så det er all grunn til å fortsette å kreve at analysene skal være transparente.

Et par av pilotevalueringene brukte mye av ressursene på den samfunnsøkonomiske analysen, uten at resultatet nødvendigvis ble mer troverdig. I tillegg har de fleste valgt enklere tilnærminger, enten en forenklet analyse hvor kun de viktigste virkningene er prissatt, og hvor en ikke bruker transportmodell (gjelder vei og jernbane), eller en kvalitativ analyse hvor evaluator har foretatt en samlet vurdering av lønnsomheten basert på faglig skjønn.

De fleste evaluatorene har oppgitt forutsetninger og forbehold, og forklart hva de har gjort, og analysene fremstår som troverdige. En ser likevel eksempler på enkelte tvilsomme antakelser, som der man uten videre forutsetter at en rekke andre prosjekter er gjennomført over andre budsjetter, eller der man forutsetter at et høyt nivå på årlige tidsgevinster opprettholdes også etter at bompengene bortfaller – til tross for at det andre steder i rapporten uttrykkes bekymring for opphopning av kø allerede i dag. I noen tilfeller kan en kanskje snakke om sviktende programteori.

Det kan nevnes at de fleste evaluatorene har en edruelig tilnærming til såkalt mernytte, som ofte løftes frem av prosjektenes pådrivere, spesielt i transportprosjekter. Flere av evalueringene nevner *mulige* indikasjoner på slike samfunnsøkonomiske tilleggsgevinster, men tar dem ikke med i beregningen. Tilsvarende, i analysen av Høgskolen på Remmen, nevnes generelle positive ringvirkninger av økt utdanning (på helse, levealder, kriminalitet og deltakelse i samfunnet) men en gir ikke noe påslag på nytten av det aktuelle byggeprosjektet. Dette er en fornuftig tilnærming.

Tabell 7.5 Evalueringenes relative ressursbruk og tolkning av kriteriene. Vår vurdering

Samfunnsøkonomisk effektivitet	Ja	Nei	Delvis eller ikke vurdert
Har evaluator tilgang til en brukbar før-analyse?	0	15	5
Er det mulig å (delvis) kvantifisere nytten?	10	10	0
Er beregninger og vurderinger transparente?	9	5	6
Er det samsvar mellom hva som tas med på nyttesiden og kostnadssiden?	15	2	3

Datagrunnlaget og triangulering

Datagrunnlaget anses som akseptabelt med noen unntak. Kvaliteten på evalueringene hviler sterk på at evaluator anvender en bred tilnærming og triangulerer mellom ulike kilder og metoder. De fleste gjør dette, men noen har et forbedringspotensial.

Alle evalueringene bruker i stor grad de samme typene datainnsamlings- og analysemetodene. Produktivitet vurderes med utgangspunkt i sluttrapport og prosjektreknskap, og suppleres med intervjuer. Måloppnåelse baseres gjerne på tidsseriedata der dette finnes, og suppleres med intervju med informanter samt befaringsnotat på stedet. Når det gjelder de strategiske kriteriene er det en overvekt av kvalitative metoder, men også bruk av statistikk til å belyse ulike trender. Lønnsomhet baseres på de øvrige dataene samt et sett av forutsetninger, enhetspriser og andre parameterverdier.

Ofte eksisterer det mye data ved oppstart av evalueringen, både i form av prosjektdata, interne evalueringer (herunder etatenes etterprøvinger av lønnsomheten i transportprosjekter) og medieoppslag både i lokal og nasjonal presse. Evaluator har en viktig rolle i å systematisere og vurdere det foreliggende og selv dekke hullene, typisk ved enkle og mindre formelle metoder. Unntaksvis har evaluator benyttet mer formaliserte datainnsamlingsmetoder selv. Et slikt unntak er evalueringen av Eiksundsambandet hvor evaluator gjorde en spørreundersøkelse blant husstandene i de seks kommunene om effekter på servicetilbudet i regionen.

I hovedsak later det til at evaluatorene har funnet tilstrekkelig gode datagrunnlag. Men det er eksempler på at en ikke får tilgang til dokumenter eller data som åpenbart burde vært tilgjengelig. Et særtilfelle er jernbanesektoren hvor evaluator nektes tilgang til passasjertall fra togoperatør. Videre har det foreliggende materialet ofte store svakheter (særlig samfunnsøkonomiske analyser). Som én evaluator sa: *Vi brukte veldig mye tid på å forstå de ulike analysene, og å korrigere feil og beregninger.*

Et par av evalueringene bærer preg av å primært være skrivebordsøvelser, hvor evaluator ikke har vært på befaring og hvor intervjuene begrenses seg til korte samtaler med noen få personer. Enkelte evaluatorene later til å anse kvantitativ informasjon som mer høyverdig enn kvalitativ informasjon og er kanskje ikke så komfortabel med å måtte gjøre egne skjønnsmessige vurderinger eller la interessentenes vurderinger få stor plass i evalueringen. Andre gjennomfører intervjuer, men benytter et skjevt utvalg av intervjuobjekter, for eksempel kun prosjektorganisasjonen. Det er derfor på sin plass å peke på viktigheten av

triangulering også i fortsettelsen. Tilstedeværelse/befaring av det som evalueres er selvsagt og skal være påkrevet.

Tabell 7.6 Evalueringenes datagrunnlag

Datagrunnlag	Ja	Nei	Delvis
Tilfredsstillende ifølge evaluator (slik vi leser rapporten)	14	1	5
God triangulering mellom kilder og metoder, vår vurdering	13	1	6

Ambisjonsnivå versus tilgjengelige ressurser

Gitt en fornuftig prioritering av indikatorer og triangulering av informasjon, vurderes rammene for evalueringene som tilstrekkelig til å gi troverdige svar på de seks kriteriene. Men det er viktig å sikre tilgang til nøkkelpersoner og sentrale data i forkant.

Alle evalueringene har hatt et begrenset budsjett, i gjennomsnitt rundt tre månedersverk. Tidsrammene har ligget på et halvt år eller noe lengre. Et betimelig spørsmål er om det er realistisk å gjennomføre en troverdig evaluering innenfor disse rammene. Da dette ble diskutert på workshopen, var det naturlig nok flere av evaluatorene som benyttet anledning til å ønske seg høyere budsjetter. Det overordnede inntrykket var likevel at evaluatorene er komfortable med kvaliteten på det de leverer. Det er også vår vurdering at de fleste evalueringsrapportene holder en akseptabel kvalitet.

Samtidig vet vi at det eksisterende underlaget (før-dataene og referansedokumenter) har vært variabelt og ikke alltid like godt. Nøkkelpersoner i departement og etater har vist varierende interesse for å bidra. En evaluator påpekte at dette innebærer en risiko, som kan begrenses ved at Concept som oppdragsgiver går lengre i å kartlegge og samle inn eksisterende data som del av forberedelsene til evalueringen. Dette er noe vi vil vurdere i fortsettelsen. Spesielt behovet for å forberede etatene på at de må sette av noe tid til å bistå evalueringsteamet. Samtidig vil det alltid være slik at en del datakilder vanskelig kan avklares før man er underveis i evalueringen.

Det må påpekes at problemene med referansegrunnlaget er midlertidig. I fremtiden vil en ha tilgang til KVV-/KS1-materialet, og dette vil avlaste evalueringsteamet. På den annen side har denne metaevalueringen pekt på enkelte forhold som bør beskrives *mer grundig* enn i dag, som for eksempel beskrivelsen av prosjektets begrunnelse og tidligfase. Dette kan gjøre evalueringen mer ressurskrevende. Vi har ikke noe entydig svar på hva som

eventuelt kan utelates, men det er klart at mange av rapportene er ordrike. Flere er på opp mot 100 sider og skiller ikke nødvendigvis mellom viktig og uviktig.

Det ble også reist spørsmål om det er riktig ressursbruk å evaluere alle prosjektene mot alle de seks kriteriene. Enkelte prosjekter er «åpenbart» vellykket for eksempel når det gjelder kostnadskontroll. Man kunne se for seg en mer fleksibel tilnærming hvor den enkelte evaluering kun går i dybden på de utfordrende aspektene. Selv om dette er et relevant poeng er det ikke gitt at man kan si på forhånd hvilke aspekter det gjelder. Videre er det i seg selv viktig at alle kriteriene blir besvart slik at en unngår huller i datasettet. Kvaliteten på senere metaevalueringer hviler på dette. Men vi er enige i at ressursbruken kan tilpasses. Dette har også vært praktisert i noen grad, for eksempel ved at vi i kun to evalueringer har bedt om mer detaljerte forklaringer på årsaken til kostnadsavvik, nemlig i de to prosjektene som hadde de største avvikene i henholdsvis negativ og positiv retning.

Erfaringer med scoresetting

Bruk av score 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.

Evaluatorene har oppsummert sin vurdering av prosjektet ved å gi en score mellom 1 og 6 for hvert kriterium. Grunnlaget for scoresettingen kan være varierende og konklusjonen tilsvarende usikker. Likevel var det ingen som ga uttrykk for at det å gi score er problematisk. Tvert imot påpekte flere evaluatorene at det er disiplinerende å måtte gi en tydelig konklusjon og begrunne den. Men enkelte var skeptiske til hvordan scorene (kan) leses og brukes. Usikkerheten knyttet til konklusjonen kan derfor med fordel synliggjøres. Én av evaluatorene har for eksempel angitt usikkerheten ved hver score (lav, middels eller høy) i evalueringsrapporten. Scoresettingen vil uansett bare være en forenkling, og leserne bør alltid oppfordres til å konsultere rapporten for nærmere begrunnelse og underlag for vurderingen. Det er tilrettelagt for akkurat dette i databasen med resultatene, som ligger åpent tilgjengelig på Concepts nettside.

Dersom komparative analyser skal ha troverdighet må en ha tillit til at alle evaluatorene legger til grunn de samme prinsippene for scoresettingen. Et viktig funn i denne metaevalueringen er at prosjektene har fått gjennomgående god score på alle kriteriene. Den laveste scoren er ikke brukt, og den nest laveste er

kun brukt 3 ganger. Til sammenligning er det tildelt 51 femmere og 12 seksere. Et spørsmål blir da om det er forhold ved evaluatorene, tolkningen av kriteriene, referansegrunnlaget eller dataene som gjør at scoren blir positivt skjev? Eller at prosjektene rett og slett er eksepsjonelt vellykkete. Vi har ikke noe klart svar på dette, men enkelte observasjoner kan helle i retning av positiv skjeving. Når det gjelder produktivitet, har vi allerede nevnt hvordan oppdaterte kostnadsrammer i noen prosjekter gjorde det lettere å holde rammen. Videre at noen kun benytter prosjektorganisasjonen som informanter. Vi har også et inntrykk av at enkelte evaluatorene har valgt en noe snever og kanskje ukritisk tilnærming til vurderingen av relevans og levedyktighet.

Det mest iøynefallende eksemplet er evalueringen av operabygget. Her er man kommet til at prosjektet scorer 5 på produktivitet til tross for en overskridelse på om lag 1 milliard eller i størrelsesorden 30 prosent om man sammenligner med Stortingets *opprinnelige* kostnadsramme. I dette tilfellet måtte prosjektet gis tilleggsbevilgninger hele syv ganger. Evaluator legger likevel til grunn siste vedtatte ramme for vurderingen og konkluderer at prosjektet holder seg innenfor kostnadsrammen.

I det samme prosjektet viser den samfunnsøkonomiske analysen en negativ nytte på hele 17 milliarder. En legger så til grunn at uten operabygget i Bjørvika ville bolig- og næringsutviklingen måtte skje i andre deler av Osloregionen, og at samlet nytte av prosjektet derfor er så stor at prosjektet allikevel scorer 4 på samfunnsøkonomisk lønnsomhet. Det er nokså opplagt at man skal unngå å legge hypotetiske argumenter til grunn. I dette tilfellet en antakelse som med stor sannsynlighet er feil.

Fra Concepts side har vi i de foreløpige retningslinjene signalisert at score 4 skal gis dersom resultatet er akseptabelt. Også dette kan i seg selv gi en positiv skjeving dersom det fører til at 4 (og ikke 3,5) blir ansett som midtpunkt på skalaen.

Evaluatorene selv har ikke entydige oppfatninger av dette. Mange sier at de står for vurderingene de har gitt og påpeker at prosjektene faktisk er vellykkete. Men som én påpekte, kan en ikke se bort fra at mangelen på et godt referansegrunnlag har kommet enkelte prosjekter til gode. Det kan være en tendens til å bruke score 4 der en er usikker eller har lite informasjon, spesielt om det som går i negativ retning. Vi ser også eksempler på at evaluatorene tar utgangspunkt i score 6 og deretter «trekker» for dokumenterte feil og svakheter. En bør vurdere tiltak for å unngå at resultatet av dette blir at KS1-

prosjekter som evalueres i fremtiden vil bli vurdert strengere enn de første 20 prosjektene, rett og slett fordi man vil ha et bedre referansegrunnlag og dermed et utgangspunkt for «trekk» for avvik fra dette.

Tabell 7.7 Balansert scoresetting eller mistanke om skjeving? Vår vurdering

Kriteriene	Balansert scoresetting?
Produktivitet	Mulig positiv skjeving, knyttet til valg av ramme, prisindeks og informanter.
Måloppnåelse	Later til å være balansert, men effektmålenes ambisjonsnivå er svært ulike (sektorforskjeller), det må tas forbehold om dette
Andre virkninger	Uklart da tolkningen av kriteriet varierer
Relevans	Mulig positiv skjeving, knyttet til selektivt behovsperspektiv. Mangel på referansegrunnlag kommer prosjektet til gode?
Levedyktighet	Mulig positiv skjeving, knyttet til selektivt behovsperspektiv. Mangel på referansegrunnlag kommer prosjektet til gode?
Samf.øk. effektivitet	Trolig balansert med noen unntak

Årsaksforklaringer og læringspunkter

Evalueringene skal forklare årsakene til resultatene og trekke ut lærdommer til nytte for nye prosjekter. Det gjelder først og fremst tidligfasen men også gjennomføringen. Så langt har de fleste hatt for lite fokus på dette.

Nytten av evalueringene avhenger av hva man kan lære av dem. For å lære må man forstå årsakene til et godt eller dårlig resultat, både hva aktørene gjorde riktig (eller galt) og hvordan kontekstuelle forhold bidro til resultatet. Dette er et svakt punkt i flertallet av evalueringene. Mange sier seg fornøyde med å svare på *hvorvidt* prosjektene lykkes – ettersom det var hva de ble spurt om.

Som tidligere nevnt gjør få evaluatorene grundige vurderinger av målstrukturen. Det kunne ha hjulpet dem med å forstå potensialet i tiltaket, som i etterkant kunne vært drøftes opp mot hva aktørene gjorde for å ta ut potensialet. Evalueringen av Høgskolen i Østfold er en av få som er inne på dette. Evaluatorene finner at de tilsiktede synergieffektene mellom fagmiljøene i bygget i

begrenset grad er tatt ut. Dette forklares *både* med at potensialet var begrenset og at bygget heller ikke var optimalt utformet for å realisere effektene.³⁰

I noen tilfeller kan det hevdes at prosjektet eller konteksten er så unik at det er lite å lære for andre. Evalueringen av Perform er en av få som eksplisitt diskuterer suksesskriterier og læringspunkter, men som samtidig antyder at mange av rådene har begrenset overførbarhet. Blant annet fordi det var et prestisjeprosjekt hvor kostnadsmålet var rangert lavest og fordi Statens pensjonskasse har så homogen kompetanse at mange ansatte fra linja kunne trekkes ut til å jobbe med IKT-utvikling. Likevel, lærdommene må nevnes, og de kan være mer overførbare enn man tror.

Flere evaluatorene ga uttrykk for at de er usikre på hvilken del av målgruppen de skal fokusere på når det gjelder læringspunkter (beslutningstakere, planleggere, gjennomføringsorganisasjonen, etc.). Vi tenker at all lærdom er relevant så lenge man presiserer hvem de ulike anbefalingene er nyttige for. Tidligfasen er særlig viktig – det er her de taktiske og strategiske forbedringene kan gjøres. Men evalueringen kan også avdekke nyttige erfaringer knyttet til detaljplanlegging og prosjektgjennomføring. Mange kan i tillegg identifisere nyttige læringspunkter til fremtidige etterevalueringer.

Involvering av målgruppen

Det må jobbes mer med å involvere departementer og etater slik at de ser nytten og læringspotensialet i evalueringene. God informasjon om hensikten med arbeidet og enkel tilgang til nøkkelresultater er viktigst.

Concept-programmet, som en tredjepart, bestiller og koordinerer evalueringene, men fagdepartementene og etatene er viktige målgrupper. På noe sikt vil nytten av evalueringsaktiviteten stå og falle med at disse tar resultatene i bruk.

Alle evalueringsprosesser starter med at Concept forankrer valget av prosjekt i fagdepartementet og får en kontaktperson i etaten. En etablerer en dialog med disse personene for å sikre at prosjektets mål, hensikt, kontekst og evt. spesielle forhold er oppfattet. De får også mulighet til å gi innspill til konkrete evalueringsspørsmål. Underveis i evalueringsprosessen vil evalueringsteamet forholde seg til de samme kontaktpersonene for tilgang til interne data og intervjuobjekter. Når utkast foreligger får de anledning til å kommentere.

³⁰ Altså både *theory failure* og *implementation failure*, ref. Edward Suchman.(1967)

Enkelte departementer har benyttet anledningen til å involvere seg, og erfaringene i disse tilfellene er gode. Enkelte av de prosjektansvarlige i etatene har også vært villige informanter. Men på dette området er det et potensiale for forbedring. Spesielt har det vært vanskelig å komme i inngrep med de som jobber med metoder og systemer for utredning og oppfølging av prosjekter, typisk på direktoratsnivå eller sentralt i etaten. Samlet sett er det vår vurdering at læringseffekten i fagdepartementer og etater har vært begrenset.

På workshopen ble mulige tiltak diskutert. God informasjon om målet og hensikten med evalueringsarbeidet er selvfølgelig viktig. Det ble kommentert at travle departementsansatte ikke kan forventes å lese en 100 siders rapport. Nøkkelresultatene må presenteres på en enkel og troverdig måte i rapportene. Det kan også være nyttig å tilby en kort presentasjon av et nesten ferdig utkast til rapport som et fast punkt i prosessen.

Et annet moment er muligheten for å ta med en representant fra den aktuelle etaten eller departementet i evalueringsteamet. En tenker da ikke på personer med direkte eierskap til prosjektet, men gjerne en person på direktoratsnivå som arbeider faglig med planlegging, gjennomføring eller overordnet styring av prosjekter. Det vil i så fall bli en litt annen type evaluering – ikke fullt ut uavhengig, men neppe et problem for troverdigheten så lenge det er snakk om én person blant flere i et team. Slik involvering vil kunne styrke mulighetene for læring og erfaringsoverføring i stor grad. Vi vil se nærmere på muligheten for å teste ut et slikt opplegg fremover.

8. Konklusjoner

De første 20 evalueringene har i stor grad vært en utprøvningsfase, hvor vi har involvert flere evalueringsmiljøer, og lært mye om hvordan evaluatorene tolker og bruker modellen og hva som er bestemmende for en god evaluering.

Vår vurdering av evalueringene er at det generelle inntrykket er positivt, tatt i betraktning de begrensede ressursene som er satt av til hver enkelt studie (tilsvarende ca. 3 månedsverk). Flertallet av evalueringene gir god valuta for pengene og leverer gode, overordnede analyser av prosjektene. Samtidig ser vi at enkelte evalueringer er for svake:

- Vi ser eksempler på at helt essensielle problemstillinger ikke er fanget opp.
- De fleste nøyer seg med å beskrive hva som har skjedd, og graver ikke dypt nok til å finne årsaken til det som skjer.
- Tolkningen og operasjonaliseringen av evalueringskriteriene er ikke konsistent
- Scoresettingen fungerer ikke helt, og vi kan ikke utelukke innslag av positiv skjeving.

Generelt mener vi allikevel at den valgte evalueringsmodellen fungerer godt. Den er standardisert slik at den kan tilrettelegge for læring på tvers av sektorer, samtidig fleksibel nok til å ivareta det forhold at alle prosjekter er forskjellige. De seks kriteriene gir en bred og balansert vurdering av prosjektets vellykkethet, som også er tilpasset de utredningene som gjøres i tidligfasen.

Det er viktig å presisere at dette skal være *overordnede* evalueringer, hvor man med relativt små ressurser vurderer prosjektet i flere perspektiver. Datainnsamling og analyse er ressurskrevende, og enkelte av evaluatorene har brent seg på å legge for mye av innsatsen i å svare grundig og presist på utvalgte spørsmål (for eksempel samfunnsøkonomisk effektivitet). Da blir det lite tid og ressurser igjen til å besvare de øvrige spørsmålene. Erfaringen er at en må triangulere mellom ulike kilder og metoder heller enn å satse på mengder av eksakt informasjon, og fordele innsatsen jevnt slik at man gir svar på alle spørsmålene, selv om dette kan gå på bekostning av presisjonen.

Vår erfaring er at evalueringsopplegget har gått seg til over tid, og at evalueringene har funnet en felles form. De fleste holder en akseptabel kvalitet. Samtidig er en del av utfordringene som ble påvist i de fire pilotevalueringene (Volden og Samset, 2013) fremdeles aktuelle. Noen av disse gjelder ikke bare evaluering av prosjekter, men evaluering mer generelt.

Én utfordring handler om manglende logikk mellom elementene i målstrukturen, og behovet for å justere eller omdefinere denne for at det skal gi mening å evaluere. Enkelte evaluatorene er for ukritiske til målstrukturen selv der den har store mangler. Logisk rammeverk og programteori er verktøy for å vurdere dette, som flere av evalueringene ville hatt nytte av å benytte mer eksplisitt.

En annen utfordring er at evaluatorene tolker de strategiske kriteriene ulikt og dels overlappende. Enkelte evaluatorene reduserer andre virkninger, relevans og levedyktighet til å være elementer i en samfunnsøkonomisk analyse. Vi ser også

at det er behov for å bedre kalibreringen ved scoresetting, ikke minst for å hindre mulig skjeving. Det er derfor behov for å etablere nærmere retningslinjer for tolkning og vurdering av kriteriene

Det største problemet, som samtidig er det enkleste å gjøre noe med, er den faglige sammensetningen av evalueringsteamene. Samfunnsøkonomi har vært overrepresentert, i noen tilfeller dominert. Det viktigste tiltaket fremover blir derfor å sikre riktig teamsammensetning. Samfunnsøkonomi er et viktig og relevant kompetanseområde, men blir for snevert alene. Vår konklusjon er at det bør være fem typer kompetanse representert:

1. Evalueringsfaglig. Denne ressursen bør normalt være teamlederen.
2. Prosjektfaglig
3. Estimering
4. Samfunnsøkonomisk
5. Samlet bør ressursene også ha god kjennskap til sektoren og lokale forhold.

En gitt ressurs vil ofte kunne dekke mer enn én type kompetanse, men det bør likevel være minst tre personer i teamet.

Erfaringene fra denne metaevalueringen vil resultere i et revidert opplegg med tydeligere krav til evaluatorene, tolkningen og vurderingen av kriteriene samt en sjekkliste over viktig informasjon som bør omfattes om prosjektet, konteksten og resultatene.

Referanser

- ALNAP (2006). *Evaluating humanitarian action using the OECD-DAC criteria. An ALNAP guide for humanitarian agencies*
- Andersen, B., Kvalheim, E.V. og Volden, G.H. (2016). *Prosjektmodeller og prosjekterstyring i statlige virksomheter*, Concept-rapport 50
- Andersen, B., Bråthen, S., Fagerhaug, T., Nafstad, O., Næss, P. og Olsson, N. (2007). *Effektvurdering av store statlige investeringsprosjekter*, Concept-rapport nr. 19
- Argyris C. og Schön, D. (1978). *Organizational Learning*. San Fransisco: Jossey-Bass.
- Askim J, Døving E. og Johnsen Å., (2013). Evalueringspraksis i den norske statsforvaltningen 2005-2011, i Halvorsen A, Lier Madsen E, Jentoft N (red.) *Evaluering. Tradisjoner, praksis, mangfold*, Fagbokforlaget.
- Baccarini, D. (1999). *The logical framework method for defining project success*. Project Management Institute
- Bakewell, O. og Garbutt, A. (2005). *The use and abuse of the logical framework approach, A review of international development NGOs' experiences*. A report for Sida. INTRAC, 2005
- Bamberger, M., Rugh, J., Church, M. og Fort, L. (2004). Shoestring Evaluation: Designing Impact Evaluations under Budget, Time and Data Constraints, *American Journal of Evaluation*, Vol. 25, No. 1, pp. 5-37
- Berg, P. m.fl. (1999). *Styring av statlige investeringer*. Sluttrapport fra styringsgruppen, Finansdepartementet, 10. februar 1999
- Blamey, A. og Mackenzie, M. (2007). Theories of change and realistic evaluation. Peas in a pod or apples and oranges, *Evaluation*, Vol 13 (4), pp 439-455
- Boardman, A., Greenberg, D., Vining, A. og Weimer, D. (2011). *Cost-benefit analysis. Concepts and practice*, 4. utgave, Pearson
- Brousselle, A. og Champagne, F. (2011). Program theory evaluation. Logic analysis. *Evaluation and program planning*, Volume 34, Issue 1, pp 69-78

Bull-Berg, H., Volden, G.H. og Grindvoll, I.L.T. (2014). *Ikke-prissatte virkninger i samfunnsøkonomisk analyse. Praksis og erfaringer i statlige investeringsprosjekter*. Concept-rapport nr. 38

Chen, H.T. (1990). *Theory-driven evaluations*. Thousand Oaks, Calif, Sage

Chianza, T. (2008). The OECD/DAC Criteria for International Development Interventions: An Assessment and Ideas for Improvement. *Journal of Multi-Disciplinary Evaluation*, Volume 5, Number 9

Concept (2015). Etterevaluering av statlige investeringsprosjekter i regi av Concept-programmet. Retningslinjer for evaluator, notat, versjon 2, 08.03.2015. http://concept-eval.ivt.ntnu.no/assets/Template_%20for_etterevaluering_av_KS_-_prosjekter_v2.pdf

Danida (2012). *Danida evaluation guidelines*, Ministry of Foreign Affairs, Danmark

DFØ (2007). *Evaluering av statlige tilskuddsordninger. Veileder*.

DFØ (2014a). *Veileder i samfunnsøkonomiske analyser*

DFØ (2014b). *Veileder. Gevinstrealisering – planlegging for å hente ut gevinster av offentlige prosjekter*

European Commission (2013). *The resource for the evaluation of socio-economic development*

European Commission (2004). *Aid Delivery Methods. Volume 1. Project Cycle Management Guidelines*.

European Commission (udatert). *Guidelines on evaluation and fitness checks*

Finansdepartementet (2003). *Reglement for økonomistyring i staten. Bestemmelser om økonomistyring i staten*. Fastsatt 12. desember 2003 med endringer, senest 5. november 2015

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

Finansdepartementet (2005). *Veileder til gjennomføring av evalueringer*

- Gasper, D. (2000). Evaluating the “logical framework approach” towards learning-oriented development evaluation, *Public administration and development*, 20, 17-28
- Glasgow R.E. (1999). Outcomes of and for diabetes education research. *Diabetes Educ* 1999; 25:74-88
- Hansen, W. (2011). *Mernytte. Næringsøkonomiske ringvirkninger av infrastrukturinvesteringer*, TØI-rapport 1180/2011
- Hatling, M., Damman, S. og Halvorsen, T. (2016). *Samlokaliseringseffekter – hva sier litteraturen? En oversikt over resultater fra relevante studier*, SINTEF-rapport nr. A27652
- Heinzerling, L. og Ackerman, F. (2002). *Pricing the priceless – Cost Benefit Analysis of Environmental Protection*. Georgetown Environmental Law and Policy Institute, Georgetown University Law Centre, Washington, USA
- Highways England (2015). *Post-Opening Project Evaluation (POPE) of main schemes*. Main report. Meta-analysis 2015.
- HM Treasury (2011a). *The Magenta Book. Guidance for Evaluation*
- HM Treasury (2011b). *The Green Book. Appraisal and Evaluation in Central Government*
- Jernbaneverket (2015). *Rutemodell 2027. Fase 3 Utrikling og anbefaling av rutemodeller. Oppsummeringsrapport*
- Kaplan, S. og Garrett, K. (2005). The use of logic models by community-based initiatives, *Evaluation and program planning*, 28, 167-172
- Karlsen, T. og Jentoft, N. (2013). Programteori versus samfunnsvitenskapelig teori, kapittel i Halvorsen, A., Madsen, E.L. og Jentoft, N. (red.). *Evaluering – Tradisjoner, praksis og mangfold*, Fagbokforlaget
- Kellogg Foundation (2004). *Logic model development guide: using logic models to bring together planning, evaluation, & action*.
- Kvalheim, E.V. (2015). *Forståelig og transparent? Formidlingen av resultater fra samfunnsøkonomiske analyser*. Concept arbeidsrapport

Kvalheim, E.V., Christensen, T., Samset, K. og Volden, G.H. (2015). *Har regjeringen fått et bedre beslutningsgrunnlag? Om effekten av å innføre konseptvalgutredning (KVU) og ekstern kvalitetssikring (KS1 og KS2) for store statlige investeringsprosjekter.*

Concept arbeidsrapport

Lædre, O., Volden, G.H., og Haavaldsen, T. (2012). *Levedyktighet og investeringstiltak. Erfaringer fra kvalitetssikring av statlige investeringsprosjekter.*

Concept-rapport nr. 38

NOU 2012:16 *Samfunnsøkonomiske analyser*. Utredning fra et utvalg oppnevnt ved kongelig resolusjon 18. februar 2011. Avgitt til Finansdepartementet 3. oktober 2012. Norges offentlige utredninger

NOU 1997:27 *Nytte-kostnadsanalyser. Prinsipper for lønnsombetsvurderinger i offentlig sektor*. Utredning fra et utvalg oppnevnt av Finans- og tolldepartementet 6. mai 1994. Avgitt 24. september 1997. Norges offentlige utredninger

Nyborg, K. (2012). *The ethics and politics of environmental cost-benefit analysis*, Routledge

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

OECD (1991). *Principles for evaluation of development assistance*, Development Assistance Committee, Paris, 1991

OECD (1999) *Guidance for Evaluating Humanitarian Assistance in Complex Emergencies*, Development Assistance Committee, Paris 1999

OECD (2002). *Glossary of key terms in evaluation and results based management*.

OECD (2010). *Quality Standards for Development Evaluation*, DAC Guidelines and Reference Series, Development Assistance Committee, 2010

Pawson, R. and Tilley, N. (1997). An introduction to scientific realist evaluation, chapter in Chelimsky, E. and Shadish, W. (eds.), *Evaluation for the 21st century. A handbook*

Practical Concepts Incorporated (1975). *Guidelines for teaching logical framework concepts*

Rogers, P., Petrosino, A., Huebner, T. og Hacsí, T. (2000). Program theory evaluation: practice, promise and problems, *New directions for evaluation*, Volume 2000, Issue 87, pp 5-13

Rosenberg, L., Posner, L.D. og Hanley E.J. (1970). *Project evaluation and the project appraisal reporting system*. Final report submitted to the Agency for International Development, by Fry consultants incorporated, July 24, 1970

Rossi, P.H., Freeman, H.E. og Lipsey, M.W. (2003). *Evaluation: A systematic approach*, 7. utgave, Sage Publications

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

Samset, K. (2012). *Lenge før og langt senere. Et tilbakeblikk på konseptvalget i historiske prosjekter*, Ex Ante Akademisk Forlag

Samset, K. (2008), *Prosjekt i tidligfasen. Valg av konsept*, Tapir akademisk forlag, Trondheim

Samset, K., Volden, G.H., Welde, M. og Bull-Berg, H. (2014). *Mot sin hensikt. Perverse incentiver – om offentlige investeringsprosjekter som ikke forplikter*. Concept-rapport nr. 40

Samset, K. og Volden, G.H. (2013). *Statens prosjektmodell. Bedre kostnadsstyring. Erfaringer med de første investerings tiltakene som har vært gjennom ekstern kvalitetssikring*, Concept-rapport nr. 35

Samset, K., Andersen, B. og Austeng, K. (2013). *Mulighetsrommet. En studie om konseptutredninger og konseptvalg*. Concept-rapport nr. 34

Samset, K., Strand A. og Hendricks, V.F. (2009). *Sykehus, fregatter og skipstunnel: Logisk minimalisme, rasjonalitet – og de avgjørende valg*, Concept-rapport nr. 21

Sartorius, R. (1991). The logical framework approach to project design and management, *Evaluation Practice* 12 (2), 1991

Scriven, M. (2013). *Key evaluation checklist (KEC)*, siste versjon funnet på nett er fra 24. juli 2013

Scriven, M. (1991). *Evaluation Thesaurus*, 5. utgave, Sage Publications, 1991

Sida (2007). *Looking back, moving forward. Sida evaluation manual*, 2nd revised version, Sverige

Concept rapport nr. 52

Statens vegvesen (2006). *Veileder for etterprøving av prissatte konsekvenser av store vegprosjekt*, Utbyggingsavdelingen UTB 2006/02

Suchman, E.A (1967). *Evaluative Research: Principles **and** Practice in Public Service **and** Social Action Programs*. New York: Russell Sage Foundation

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

Tornes, K. (2013). Om verdier og kriterier i evaluering, kapittel i Halvorsen, A., Madsen, E.L. og Jentoft, N. (red.). *Evaluering – Tradisjoner, praksis og mangfold*, Fagbokforlaget

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

Weiss, C. (1997). Theory-based evaluation: past, present, and future, *New Direction for evaluation*, Vol 1997, Issue 76, pp 41-55

Welde, M. (2017). *Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitetssikring*, Concept-rapport nr. 51

Welde, M. (2016). *Kostnadsutvikling i store, statlige investeringsprosjekter fra KS1 til KS2*. Concept arbeidsrapport.

Welde, M., Samset, K., Andersen, B. og Austeng, K. (2014). *Lav prising – store valg. En studie av underestimering av kostnader i prosjekters tidligfase*. Concept-rapport nr. 39

World Bank (2004). *Monitoring and evaluation. Some tools, methods and approaches*

Worsley, T. (2014). *Ex post Assessment of Transport Investments and Policy Interventions*, ITS Roundtable summary and conclusions. OECD/ITF Joint transport research centre discussion papers

Yin, R. (2013). *Case study research. Design and methods*, 5. utgave, Newbury Park, CA: Sage

Del 3 Kort sammenfatning av evalueringene

I denne delen presenteres resultatene av evalueringene i et kortfattet format, med ett prosjekt på hver side. For mer utfyllende informasjon, se Concept-programmets nettside hvor alle resultatene er samlet, og hvor en også kan laste ned evalueringsrapportene:

www.ntnu.no/web/concept/etterevaluering

1. Statlig kontrollområde Svinesund

Prosjektnavn: Statlig kontrollområde Svinesund

Sektor: Bygg

Sted: Svinesund

Årstall for vedtak (tilsvarer byggestart): 2002

Årstall for ferdigstilling: 2005

Årstall for gjennomført evaluering: 2012

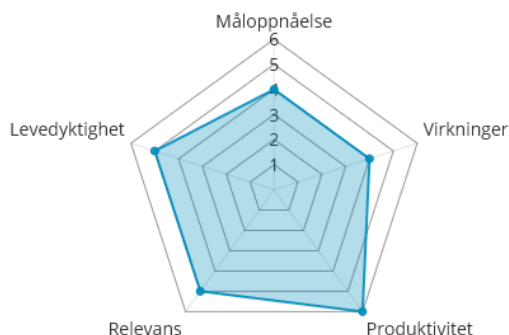
Styringsramme: 275 mill. kroner (2004)

Kostnadsramme: 290 mill. kroner (2004)

Sluttkostnad: 257 mill. kroner (2006)

Evalueringen er utført av SINTEF, i samarbeid med NTNU/Concept

Prosjektleder evaluering: Håkon Finne



Bakgrunn: Statlig kontrollområde Svinesund er et felles kontrollområde for Tollvesenet, Statens vegvesen, Politiet og private spedisjonsfirma ved landets mest trafikkerte grenseovergang til Sverige (E6 som firefelts motorvei og den nye Svinesundbrua).

- Prosjektet ble fullført etter planen, innenfor både styrings- og kostnadsramme. Kvalitetsmålet ble bare delvis innfridd.
- En mer effektiv grensekontroll er oppnådd, men i mindre grad grunnet samlokaliseringen av etatene.
- Kontrollområdet ligger ved landets viktigste grenseovergang, men beliggenheten er ikke den mest ideelle for Vegvesenet, og det kan ikke utelukkes at andre konsepter ville vært mer relevant.
- Man ser en del uheldige sidevirkninger, bl.a. smuglere kan velge grenseoverganger lenger nord, og periodevis store opphopninger av lastebiler på kontrollområdet.
- Prosjektet er vurdert som levedyktig, men avhengig av at fortollingsregimet mellom EU og EØS-land opprettholdes. Anlegget har begrenset alternativ verdi.
- En har funnet at det er vanskelig å prissette de sentrale virkningene. En anslår likevel skjønnsmessig at den samfunnsøkonomiske nytten er positiv.

2. Dobbeltspor Asker-Sandvika

Prosjektnavn: Dobbeltspor Sandvika-Asker

Sektor: Jernbane

Sted: mellom Sandvika og Asker

Årstall for vedtak (tilsvarende byggestart): 2001

Årstall for ferdigstilling: 2005

Årstall for gjennomført evaluering: 2012

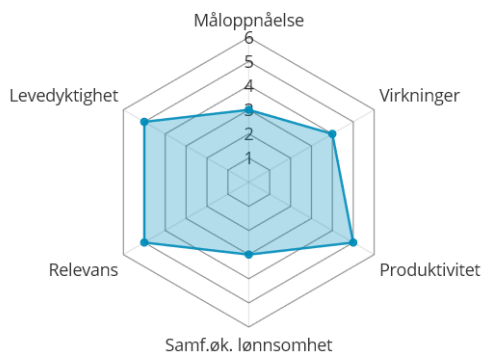
Styringsramme: 3710 mill. kroner (2006)

Kostnadsramme: 4165 mill. kroner (2006)

Sluttkostnad: 3714 mill. kroner (2008)

Evalueringen er utført av VTI (Statens väg- og transportforskningsinstitut)

Prosjektleder evaluering: Johan Nyström



Bakgrunn: prosjektet gjelder bygging av 9,5 km dobbeltspor på strekningen Asker-Sandvika samt arbeid på begge stasjoner og nytt signalanlegg. Prosjektet gjelder den ytterste av tre deletapper mot Oslo, på Norges mest trafikkerte jernbanestrekning.

- Prosjektet ble effektivt gjennomført og hadde god produktivitet. Det ble ferdigstilt innenfor tids- og kostnadsramme, men med noe reduksjon i kvalitetsnivået.
- Effektmålene for prosjektet anses bare i begrenset grad som oppnådd. Investeringen har, per 2012, hatt små effekter og gevinstene var trolig overvurdert i forkant, fordi togtilbudet og rutetabellene kun er marginalt forbedret. Intervjuobjekter oppgir at de fulle gevinstene ikke kan realiseres før alle etappene er ferdigstilt.
- Prosjektet syntes ikke å ha noen vesentlige indirekte gevinster.
- At tiltaket for å øke jernbanekapasitet på strekningen er relevant synes åpenbar. Evaluatoren stiller imidlertid spørsmål ved om rekkefølgen av utbyggingen var den mest hensiktsmessige.
- Evaluatoren anser at tiltakets levedyktighet samlet sett må sies å være høy. Prosjektet er prinsipielt sett viktig i et miljøperspektiv. Kollektivtransport bidrar til reduserte klimautslipp og et bedret lokalmiljø.
- Det har ikke vært mulig å etterprøve den samfunnsøkonomiske analysen i mangel av informasjon og data. Prosjektet vurderes likevel som lite samfunnsøkonomisk lønnsomt, basert på de kvantifiserbare virkningene.

3. E18 Momarken-Sekkelsten

Prosjektnavn: E18 Momarken - Sekkelsten

Sektor: Vei

Sted: Mellom Askim og Mysen

Årstall for vedtak (tilsvarer byggestart): 2005

Årstall for ferdigstilling: 2007

Årstall for gjennomført evaluering: 2012

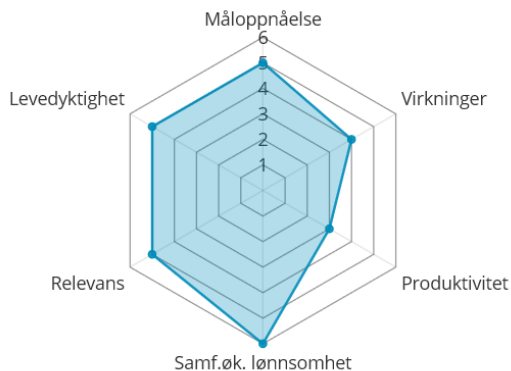
Styringsramme: 450 mill. kroner (2004)

Kostnadsramme: 490 mill. kroner (2004)

Sluttkostnad : 583 mill. kroner (2010)

Evalueringen er utført av forskere ved Concept-programmet, NTNU

Prosjektleder evaluering: Torbjørn Aass



Bakgrunn: Prosjektet omfatter bygging av en 6,2 km lang vegparsell på nye E18 gjennom Østfold. Dette er landets nest viktigste transportkorridor til utlandet på land. Hensikten var å etablere en sikker og effektiv vei mellom Mysen og Askim, bygget som en firefelts motorvei med 20 m total veibredde.

- Prosjekt hadde under middels høy produktivitet, og overskred kostnadsrammen med fem prosent. Det ble gjennomført innenfor planlagt tids- og kvalitetsramme.
- Alle tre effektmålene vurderes som oppnådd. Reisetiden er redusert og mer forutsigbar, det er observert en markant nedgang i antall ulykker, og trafikken langs den gamle E18 har blitt redusert med over 80 prosent.
- Lokalt ser man positiv virkning på nærmiljøet, men også økte reisekostnader og båndlegging av store naturområder. Regionalt sett har biltransport blitt mer attraktivt, og jernbanen opplever merkbar nedgang i passasjertallene.
- Prosjektet dekker behovene for raskere transport, færre ulykker og bedret miljø for lokalbeboerne. Trafikkutviklingen har vært høyere enn prognosen og vurderingen er at det var behov for veien.
- Trafikkprognosene i årene som kommer viser en betydelig økning i trafikkmengde, som i sin tur vil øke nytten og lønnsomheten av prosjektet uten å påvirke trafikksikkerheten og -avviklingen.
- Veiinvesteringen Momarken-Sekkelsten er et lønnsomt prosjekt. Netto nytte er betydelig høyere enn estimatet i beslutningsgrunnlaget for prosjektet.

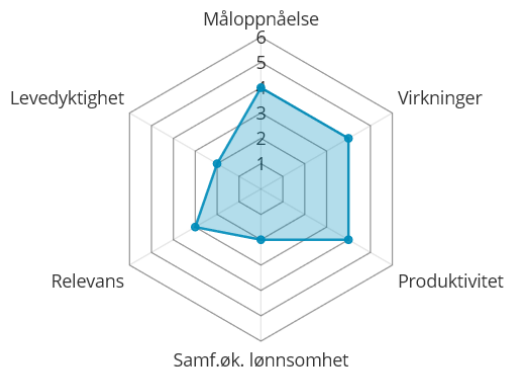
4. Skjold-klasse MTBer

Prosjektnavn: Skjold serie (MTB)
Sektor: Forsvar
Sted: Mandal (ble laget her, ved Umoe)

Årstall for vedtak (tilsvarer byggestart): 2003
Årstall for ferdigstilling: ikke ferdigstilt
Årstall for gjennomført evaluering: 2012

Styringsramme: ikke oppgitt
Kostnadsramme: 5412 mill. kroner (2010)
Antatt sluttkostnad: 5000 mill. kroner (2012)

Evalueringen er utført av Scanteam, i samarbeid med NTNU
Prosjektleder evaluering: Erik Whist



Bakgrunn: Forsvarsprosjektet omfatter utvikling av prototyp og bygging av fem hurtiggående missiltorpedobåter. Fartøyene erstatter den tidligere MTB Hauk klassen, og skal utføre overflateoperasjoner som et ledd i Sjøforsvarets utøvelse av sjøkontroll.

- Sluttkostnaden ligger godt innenfor kostnadsrammen og fartøyene leveres med tilfredsstillende kvalitet. Prosjektet er imidlertid betydelig forsinket i forhold til kontraktfestete leveransedatoer.
- Effektmålene vurderes som oppnådd, men i noe varierende grad. Det var også næringspolitiske mål utover selve gjennomføringen av leveransen, disse vil neppe bli oppnådd.
- Fartøyene har bidratt til å gi Norge et av Europas mest moderne sjøforsvar, men de er neppe optimale i en større forsvarmessig sammenheng. Prosjektet har gitt visse positive økonomiske og sysselsettingsmessige ringvirkninger lokalt (Mandal).
- Prosjektets relevans er omdiskutert både i forhold til forsvarsbehov og økonomiske rammer for Forsvaret. Det er betydelig overlapp med andre fartøyserier i forhold til egenskaper og oppgaver. Det gjelder spesielt de nye fregattene.
- Fartøyenes levedyktighet er satt til 40 år, og det årlige driftsbudsjettet er estimert til 200 millioner kroner (2012). MTB-ene er sårbare for å bli faset ut. Avhengighet av én leverandør som kun hadde dette ene oppdraget gjør også anskaffelsen sårbar.
- Evalueringen inneholder ingen egen samfunnsøkonomisk analyse, men ut fra en samlet vurdering anslår man at lønnsomheten vil være begrenset.

5. E6 Riksgrensen-Svingenskogen

Prosjektnavn: E6 Riksgrensen - Svingenskogen

Sektor: Vei

Sted: Østfold

Årstall for vedtak 2002

Årstall for ferdigstilling 2005

Årstall for evaluering: 2014

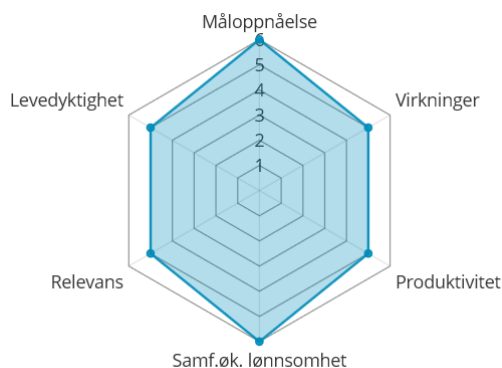
Styringsramme: 847 mill. kroner (i år 2006)

Kostnadsramme: 964 mill. kroner (i år 2006)

Sluttkostnad: 850 mill. kroner (i år 2006)

Evalueringen er utført av COWI

Prosjektleder evaluering: Hanne Samstad



Bakgrunn: Prosjektet omfatter bygging av ny grenseovergang på E6 med 4,3 km firefelts motorvei på norsk side samt ny Svinesundbru i samarbeid med det svenske Vägverket. Bakgrunnen for utbyggingen var køproblemer og trafikkuulykker på den gamle veien.

- Prosjektet ble gjennomført innenfor avtalte kostnads-, tids- og kvalitetsrammer, og evalueringsteamet konkluderer at produktiviteten i prosjektgjennomføringen var god.
- Effektmålene anses som oppnådd. Utbyggingen har ført til økt fremkommelighet, og det er observert en reduksjon i ulykker og alvorlighetsgraden av disse.
- Svinesundforbindelsen har hatt enkelte utilsiktede virkninger, både positive og negative. Dette inkluderer blant annet høyere lokale og globale utslipp til luft og mer støy enn forventet, det er gjort arkeologiske funn, og en ser noe økonomisk vekst regionalt.
- Evaluatør konkluderer med at tiltaket er relevant, da den har dekket behovet for økt fremkommelighet og trafiksikkerhet. Det er imidlertid målkonflikt med samfunnsmålet om mer miljøvennlige transportformer.
- Prosjektet anses som levedyktig. Veien har kapasitet til å møte fremtidig trafikkvolum, mulighet til å opprettholde tilstrekkelig drift og vedlikehold, og relativt lave miljøkostnader sammenlignet med nytten.
- Prosjektet er samfunnsøkonomisk lønnsomt, og trolig mer lønnsomt enn antatt på forhånd.

6. Svalbard forskningspark

Prosjektnavn: Svalbard forskningspark

Sektor: Bygg

Sted: Svalbard, Longyearbyen

Årstall for vedtak: 2003

Årstall for ferdigstilling: 2005

Årstall evaluering: 2014

Styringsramme: 332 mill. kroner (i år 2003)

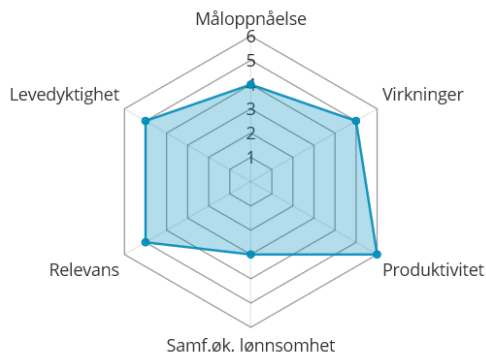
Kostnadsramme: 400 mill. kroner (i år 2003)

Sluttkostnad: 320 mill. kroner (i år 2008)

Evalueringen er utført av forskere ved

Concept-programmet, NTNU

Prosjektleder evaluering: Erik Whist



Bakgrunn: Prosjektet omfatter byggingen av et nytt bygg og integrering av det opprinnelige bygget til Universitetssenteret på Svalbard (UNIS). Prosjektet var planlagt å samlokalisere de akademiske miljøene i Longyearbyen, Svalbard Museum og et magasin for arkeologisk og kulturhistorisk materiale.

- Prosjektet ble gjennomført innenfor avtalt kostnads-, tids- og kvalitetsramme, og evalueringsteamet konkluderer at produktiviteten og gjennomføringen var god.
- Prosjektets tre effektmål anses i all hovedsak oppnådd: økt samarbeid mellom Polarinstituttet og UNIS, økt og bedret informasjonsvirksomhet om Svalbard, samt en viss grad av utvikling og utveksling av kompetanse mellom aktørene.
- Tiltaket har hatt størst positive virkninger for UNIS og Svalbard Museum, men har også medført samhandling og utvikling av nettverk mellom de øvrige institusjonene, og bidratt til en positiv utvikling for lokalmiljøet i Longyearbyen.
- Prosjektet blir vurdert som relevant, primært hva gjelder målet om norsk tilstedeværelse, mer sekundært i forhold til lokale behov (institusjonelt samarbeid, utviklingen av lokalsamfunnet og næringsvirksomheten).
- Prosjektets levedyktighet er rimelig god gitt dagens Svalbardpolitikk. Bygget er godt vedlikeholdt, leietakerne synes ikke å ha noen innvendinger mot husleienivået, og aktivitetene i bygget fremstår som relevante.
- På grunn av klimaet er kvadratmeterprisen på bygget høyere enn på fastlandet. Det er også langt dyrere å drive med forskning og undervisning på Svalbard. Flere mener tiltaket er verdt pengene ut fra politiske mål, men det kan likevel diskuteres om det finnes mer kostnadseffektive måter å oppnå tilstedeværelse på.

7. Lofoten fastlandsforbindelse LOFAST

Prosjektnavn: Lofotens fastlandsforbindelse (LOFAST)

Sektor: Samferdsel

Sted: Lofoten

Årstall for vedtak (tilsvarende byggestart): 2003

Årstall for ferdigstilling: 2007

Årstall for gjennomført evaluering: 2014

Styringsramme: 1343 mill. kroner (2009)

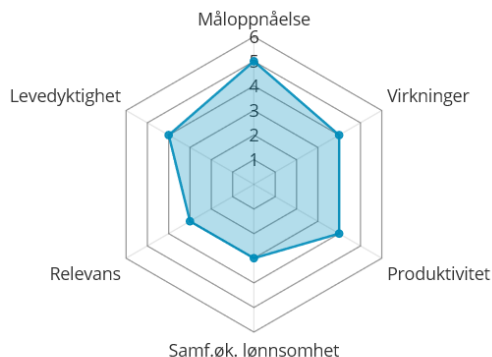
Kostnadsramme: 1500 mill. kroner (2009)

Sluttkostnad: 1380 mill. kroner (2009)

Evalueringen er utført av Universitetet i Nordland, i

samarbeid med Nordlandsforskning AS

Prosjektleder evaluering: Gisle Solvoll



Bakgrunn: Prosjektet gjelder bygging av tre mil vei mellom Raftsundet og Gullsfjordbotn, som gir Lofoten fergefri direkteforbindelse til fastlandsregionen Ofoten og E6. Prosjektet skapte lokal uenighet om trasévalget, der det andre alternativet var en trasé gjennom den mer folkerike delen av Vesterålen.

- Prosjektet ble gjennomført innenfor planlagt tidsramme og med avtalt kvalitet, og med en sluttkostnad om lag 8 prosent lavere enn kostnadsrammen. Evalueringsteamet konkluderer at produktiviteten var god.
- Veien har ført til reduserte transportkostnader og bedre forutsigbarhet for trafikantene. Trafikken til og fra Lofoten har økt betraktelig og det meste av den benytter LOFAST fremfor fergesambandene.
- Den økte tilgjengeligheten for turister fører til redusert oppholdstid, og endring i reisemønster. Prosjektet er i dag viktig for de sentrale næringene i regionen. Veien ble lagt gjennom inngrepsfrie naturområder og dette var og er kontroversielt.
- Veien brukes mer enn forventet, men det er uklart om den valgte traseen er den mest relevante i forhold til målet om å styrke Lofoten og Vesterålen som én region.
- Prosjektet har en teknisk kvalitet som tilfredsstillende dagens og fremtidens krav til veiinfrastruktur. Men dersom tunnel under Hadselsfjorden blir bygget vil behovet for LOFAST-traséen bli betydelig redusert.
- Prosjektet er ikke samfunnsøkonomisk lønnsomt, selv om trafikkvolumet er større enn antatt. Om man hadde valgt traseen via Hadselsfjorden og Vesterålen ville tiltaket blitt noe mindre ulønnsomt.

8. Eiksundsambandet

Prosjektnavn: Eiksundsambandet

Sektor: Samferdsel

Sted: Sunnmøre

Årstall for vedtak: 2003

Årstall for ferdigstilling: 2008

Årstall for evaluering: 2014

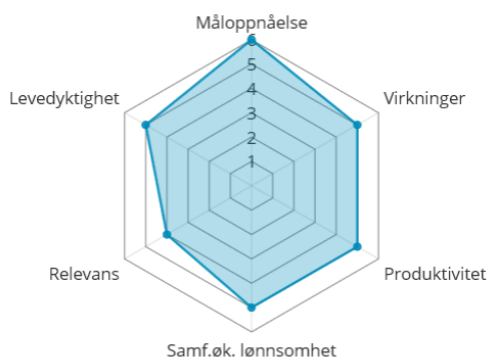
Styringsramme: 1078 mill. kroner (år 2009)

Kostnadsramme: 1145 mill. kroner (år 2009)

Sluttkostnad: 1062 mill. kroner (år 2009)

Evalueringen er utført av Menon Business
Economics

Prosjektleder evaluering: Heidi Ulstein



Bakgrunn: Eiksundsambandet gir fast veisamband mellom fire øykommuner og fastlandskommunene Ørsta og Volda på Nordvestlandet, med om lag 40 000 innbyggere og betydelig industrivirksomhet. Prosjektet omfatter en undersjøisk tunnel på om lag syv kilometer, to broer, og bygging og utbedring av tilslutningsveier.

- Prosjektet ble gjennomført under både styrings- og kostnadsrammen, kvaliteten på anlegget er god men det ble avsluttet tre måneder senere enn avtalt.
- Det er betydelig reduksjon i reisetid og tidskostnader. Bompengavgiften er høyere enn tidligere fergekostnad, men denne er tidsbegrenset. Den samlede vurderingen er at måloppnåelsen er god, med trafikkøkning mellom kommunene.
- En ser en positiv påvirkning på arbeidsmarked og produktivitet i næringslivet, men bare begrenset virkning på det offentlige og private servicetilbudet.
- Tidsinnsparinger og økt trafikk gir først og fremst nytte for arbeidsmarked, nærings- og servicetilbud og investeringen vurderes derfor som relevant. Evalueringen konstaterer at det har vært en betydelig folketilvekst til regionen i årene etter at sambandet stod ferdig.
- En finner at de langsiktige effektene av prosjektet trolig er større enn de korte- og mellomtsiktige effektene, som trolig vil forsterkes av en eventuell videre utbygging av fergefri E39 mot Ålesund.
- Prosjektet var beregnet å være ulønnsomt i forkant. En forenklet samfunnsøkonomisk analyse viser at prosjektet tvert imot har høy og positiv netto nytte, hovedsakelig fordi trafikkstrømmen har blitt større enn antatt.

9. NAV IKT Basis

Prosjektnavn: NAV IKT Basis

Sektor: IKT

Sted: Nasjonalt

Årstall for vedtak: 2006

Årstall for ferdigstilling: 2010

Årstall for gjennomført evaluering: 2014

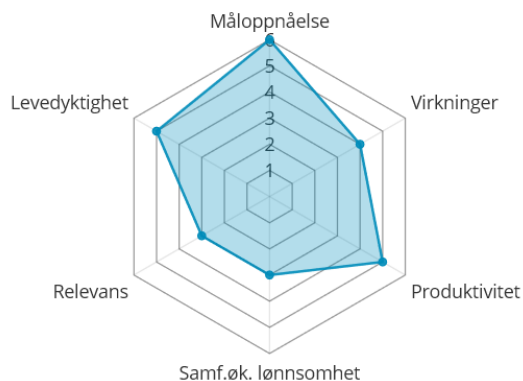
Styringsramme: 990 mill. kroner (2009)

Kostnadsramme: 1180 mill. kroner (2009)

Sluttkostnad: 867 mill. kroner (2009)

Evalueringen er utført av Norsk institutt for by- og regionforskning

Evaluator: Steinar Johansen og Sigrid Skålnes



Bakgrunn: IKT Basis-prosjektet i NAV gjelder investeringer i IKT systemer videreutvikling av eksisterende databaser og programmer, i tillegg til utrulling av nytt utstyr (telefoner og terminaler) til etatens vel 400 kontorer.

- Prosjektet ble gjennomført innenfor gitte kostnads- og tidsrammer. Det vurderes som godt planlagt og hadde prosjektstyring med oppfølging på alle nivåer.
- Samtlige effektmål vurderes å være oppnådd. Systemet med felles plattform fungerer og ansatte synes ikke å ha opplevd problemer med den nye ordningen.
- Det er ikke identifisert vesentlige utilsiktede virkninger av prosjektet, og IKT Drift mener at virkningene man har fått var forventet.
- IKT Basis var helt nødvendig for å kunne gjennomføre NAV-reformen, og var slik sett relevant. Evaluator mener at videre diskusjon av den samfunnsmessige relevansen av prosjektet må omfatte også den samfunnsmessige relevansen av selve NAV-reformen.
- IKT Basis var aldri ment som en varig løsning, og levedyktigheten vurderes ut fra hvor godt egnet det var å bygge videre på. Erfaringene er at IKT Basis var en fornuftig plattform, som også ga NAV viktig erfaring med store IKT-prosjekter.
- Det har vært vanskelig for evaluator å kvantifisere lønnsomheten av prosjektet, men en vurderer at det trolig er lønnsomt. Det er imidlertid for begrenset å se på IKT-Basis alene, da prosjektet bare er en avgrenset del av NAV-reformen.

10. Høgskolen i Østfold

Prosjektnavn: Høgskolesenter på Remmen

Sektor: IKT

Sted: Østfold, Halden

Årstall for vedtak: 2003

Årstall for ferdigstilling: 2006

Årstall for gjennomført evaluering: 2015

Styringsramme: 605 mill. kroner (2006)

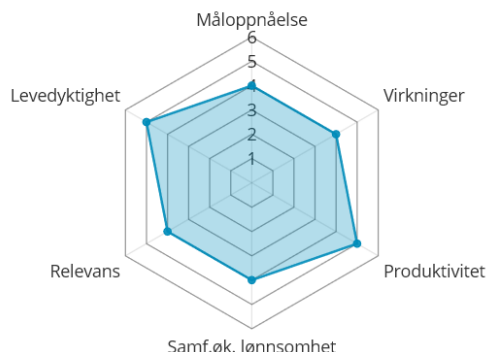
Kostnadsramme: 663 mill. kroner (2006)

Sluttkostnad: 560 mill. kroner (2006)

Evalueringen er utført av SINTEF, i samarbeid med

NTNU og Concepts forskere

Prosjektleder evaluering: Håkon Finne



Bakgrunn: Byggeprosjektet høgskolesenteret på Remmen i Halden omfatter rehabilitering av den tidligere lærerhøgskolen på Remmen og bygging av nybygg for å samlokalisere høgskolens avdelinger i Halden.

- Prosjektet ble utført godt innenfor avtalt styrings- og kostnadsramme, og innenfor tidsrammen. Når det gjelder kvaliteten var det lagt stor vekt på estetiske hensyn, kanskje mindre på funksjonelle. En har i ettertid hatt kostbare tilpasninger, endringer og oppgraderinger, blant annet etterisolering.
- Bygget legger bare i begrenset grad til rette for samlokaliseringseffekter slik det skulle, og det har vært begrenset samarbeid mellom fagavdelingene. Imidlertid har prosjektet ført til en kapasitetsøkning når det gjelder faglig og administrativt ansatte og studenter, som var et mål i seg selv.
- Prosjektet har enkelte negative utilsiktede virkninger, primært knyttet til det avseses lokaliseringsvalget utenfor Halden bysentrum. Det er lite studentaktivitet på kveldstid og det er en bilbasert campus.
- Prosjektet anses som relevant. Anlegget gir gode fasiliteter for forskning og undervisning, og det har vært en positiv økning i søknadstall. Lokaliseringen trekker noe ned når det gjelder relevans.
- Det er sannsynlig at bygget er holdbart i den antatte levetiden, og at Høgskolen vil bli værende der. Det er også utvidelsesmuligheter på tomten. Imidlertid er det flere usikkerhetsfaktorer, selv om både studentmiljø og faglig samarbeid later til å bli styrket i tiden som kommer.
- Samfunnsøkonomisk er det vanskelig å påvise at en kommer bedre ut etter byggeprosjektet. Om en sammenligner med tilsvarende byggeprosjekter i høgskolesektoren, er vurderingen at en kommer ut over middels.

11. E16 Kløfta-Nybakk

Prosjektnavn: E16 Kløfta – Nybakk

Sektor: Vei

Sted: Kløfta - Kongsvinger

Årstall påbegynt: 2005

Årstall ferdigstilt: 2007

Årstall evaluering: 2015

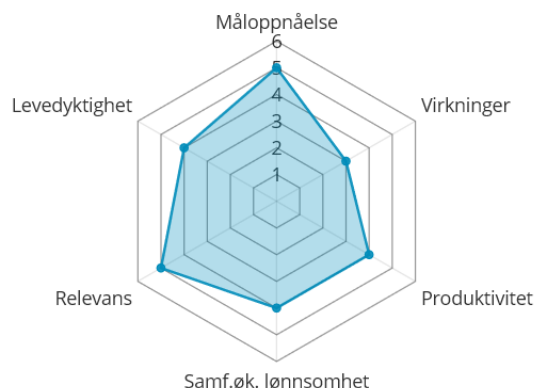
Styringsramme: 736 mill. kroner (2008)

Kostnadsramme: 759 mill. kroner (2008)

Sluttkostnad: 747 mill. kroner (2008)

Evalueringen er utført av Urbanet Analyse

Prosjektleder evaluering: Hilde Solli



Bakgrunn: Veiprojektet E16 Kløfta-Nybakk er en delstrekning i prosjektet E16 Kløfta-Kongsvinger. Prosjektet omfatter bygging av 10,5 km smal firefeltsvei. Vegen ble bygget for å gi økt trafikksikkerhet og bedre trafikkavvikling mellom Kløfta og Kongsvinger, og videre til Sverige og Oslo.

- Prosjektet oversteg styringsrammen, men var innenfor vedtatt kostnadsramme. Kvalitetsmessig har den nye veien løst de tidligere utfordringene med veistandarden, men har møtt enkelte andre utfordringer.
- Prosjektet hadde tre effektmål, lavere reisetid og –kostnader, bedre trafikksikkerhet, og bedring av miljøforholdene for veiens naboer. Evaluators anser disse som oppnådd og at prosjektet var et godt virkemiddel for å nå målene.
- Effekten for lokalt næringsliv kan ikke vurderes før hele strekningen mellom Kløfta og Kongsvinger er ferdigstilt. Prosjektet har hatt negative konsekvenser for dyrka mark. Veien er i stor grad lagt på dyrka mark og har dermed hatt en negativ virkning på miljøet.
- Prosjektet anses som relevant. Veien var et svar på opplevde utfordringer og behov både for transportkorridoren og lokalt.
- Prosjektet anses som levedyktig, og de positive virkningene av prosjektet vil vedvare over tid. Imidlertid er ikke veien et godt svar på klimamål, og har redusert kjørekomfort og standard grunnet uheldige konsekvenser av kvalitetsvalg i prosjektet.
- Kløfta-Nybakk har trolig negativ lønnsomhet. Imidlertid har mange veiprojekter negativ nytte, sammenlignet med disse har veien verken spesielt bra eller spesielt dårlig lønnsomhet.

12. Rv 519 Finnfast

Prosjektnavn: Rv519 Fastlandsforbindelsen til Finnøy (Finnfast)

Sektor: Vei

Sted: Finnøy utenfor Stavanger

Årstall start: 2006

Årstall ferdigstilt: 2009

Årstall evaluering: 2015

Styringsramme: 668 mill. kroner (2015)

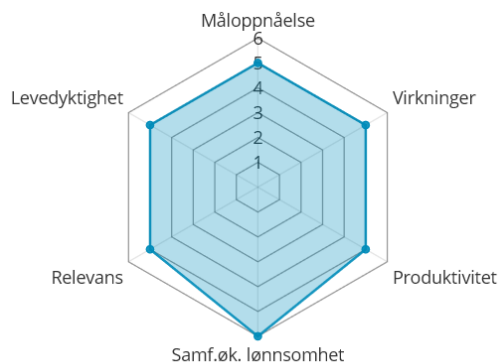
Kostnadsramme: 706 mill. kroner (2015)

Sluttkost: 614,5 mill. kroner (2015)

Evalueringen er utført av Menon Business

Economics

Prosjektleder Heidi Ulstein



Bakgrunn: Prosjektet omfatter en undersjøisk tunnel som kobler øyene Finnøy og Talgje med Rennesøy som allerede hadde fastlandsforbindelse. Prosjektet erstattet det tidligere fergesambandet, og gir øyboerne døgnåpen og raskere fastlandsforbindelse.

- Prosjektet ble gjennomført godt innenfor kostnads- og styringsrammen, og sambandet ble åpnet som planlagt i 2009. Evaluators mener at prosjektet har generelt god kvalitet, men er usikker på hvor ambisiøs rammen var.
- Måloppnåelsen i prosjektet har vært god. Reisetiden fra Finnøy til Stavanger er i gjennomsnitt halvert, fra 70 minutter til 35 minutter. Samlet sett har de fleste fått et bedre kommunikasjonsstilbud på grunn av Finnfast.
- Finnfast ser ut til å ha hatt positive effekter på samfunnsmålene om et større og mer fleksibelt arbeidsmarked, økt konkurransekraft for næringslivet og sikring av eksisterende bosetning. Det ser også ut til å bidra til økt overgang til elbil, dette er i hovedsak en utilsiktet virkning av de høye bomavgiftene.
- Betalingsvilligheten lokalt er høy og evaluator mener at alternative konsepter ikke ville ikke gitt de samme effektene. Prosjektet er finansiert nesten uten ordinære statlige midler.
- Evaluator antar at de positive effektene vil øke over tid. Utviklingen i petroleumsnæringen gir en viss usikkerhet. En avgjørende faktor er den fremtidige utviklingen i havbruk som forventes å bli en vekstnæring.
- Finnfast er et samfunnsøkonomisk lønnsomt prosjekt. En forenklet analyse viser en positiv netto-nytte på over to milliarder kroner. Dette er betydelig høyere enn anslått i ex ante analysen, og skyldes i hovedsak større trafikk enn planlagt.

13. Dobbeltspor Sandnes-Stavanger

Prosjektnavn: Dobbeltspor Sandnes – Stavanger

Sektor: Jernbane

Sted: Rogaland

Årstall for vedtak: 2005

Årstall for ferdigstilling: 2009

Årstall for gjennomført evaluering: 2015

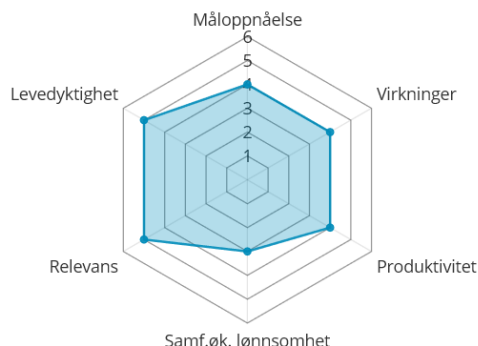
Styringsramme: 1344 mill. kroner (år 2005)

Kostnadsramme: 1517 mill. kroner (år 2005)

Sluttkostnad: 2421 mill. kroner (i år 2014)

Evalueringen er utført av Oslo Economics, i samarbeid med Atkins

Prosjektleder evaluering: Marit Svensgaard



Bakgrunn: Dette Jernbaneprosjektet gjelder bygging av 14,5 km jernbanespor mellom Sandnes og Stavanger, langs eksisterende spor med tilhørende tekniske anlegg, og en doubling av antall holdeplasser på strekningen. Bakgrunnen var den sterke befolkningsveksten i området og ønsket om å overføre trafikk fra vei til jernbane.

- Prosjektet har hatt en betydelig overskridelse av kostnadsrammen, men ble levert til planlagt tid og i hovedsak med planlagt kvalitet. Et vesentlig unntak er imidlertid signalanlegget som er midlertidig i påvente av innføring av ERTMS.
- Målene om økt driftsstabilitet, økning fra halvtimes til 15-minutters frekvens, samt fortetting langs bybåndet mellom Sandnes og Stavanger vurderes som oppnådd. Målet om økt markedsandel er imidlertid i liten grad oppnådd.
- Dobbeltsporet har hatt begrensede virkninger utover de direkte effektene. Samlet sett har utbyggingen hatt moderate negative virkninger på nærmiljøet, og moderate positive virkninger for regionen.
- Prosjektet anses som relevant, i forhold til behovene for mer miljøvennlig bytransport, redusert bilavhengighet og økt kollektivtrafikk, samt bedre fremkommelighet på veiene. Tiltakets utforming var allikevel ikke helt optimal.
- Prosjektet anses som levedyktig, med en forventet økning i trafikkgrunnlaget. Dobbeltsporet er imidlertid avhengig av utviklingen av det øvrige transportsystemet på Nord-Jæren, som vil påvirke levedyktigheten.
- Prosjektet vurderes som ikke samfunnsøkonomisk lønnsomt, med en beregnet netto nåverdi på minus 340 millioner kroner.

14. Regionfelt Østlandet

Prosjektnavn: Regionfelt Østlandet (RØ)

Sektor: Forsvar

Sted: Østlandet, ved Rena

Årstart påbegynt: 2002

Årstart ferdigstilt: 2012

Årstart evaluering: 2015

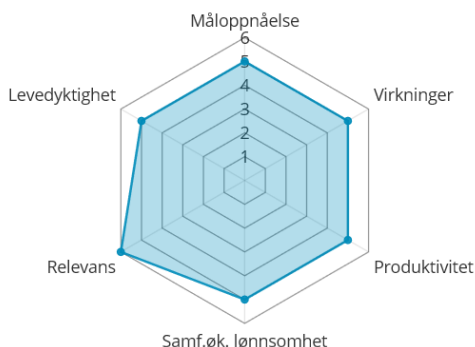
Styringsramme: 1832 mill. kroner (2016)

Kostnadsramme: 2113 mill. kroner (2016)

Sluttkostnad: 1806 mill. kroner (2016)

Evalueringen er utført av Prokonsult AS

Prosjektleder: Erik Whist



Bakgrunn: Prosjektet omfatter utbyggingen av et multi-funksjonelt treningsområde for Hærens styrker på Østlandet, der alle de viktigste våpengrenene kan samhandle. Det ble bygget som et ledd i omleggingen av Forsvaret til å bli mindre og mer slagkraftig.

- Prosjektets sluttkostnad er godt innenfor styrings- og kostnadsrammen. Utbyggingen ble om lag tre år forsinket i henhold til tidsrammen, og var ferdigstilt i 2012. Kvaliteten var som forventet.
- Prosjektets syv av åtte effektmål ble helt eller delvis nådd. Feltet ligger strategisk plassert nær Rena leir og gir mulighet for mange ulike øvelser. Områdets størrelse og utforming legger likevel noen begrensninger på bruken
- Regionfelt Østlandet har hatt hovedsakelig positive virkninger for blant annet Forsvaret, Forsvarsbygg, lokalbefolkningen, kommunen og grunneiere, som trolig er den gruppen lokalt som har kommet best ut av RØs opprettelse og drift.
- Prosjektet fremstår som klart relevant i forhold til de overordnede målene om å øke Forsvarets stridsevne. Øvingsfeltet har en lokalisering som tillater økt øvelsesfrekvens, og Forsvarsdepartementet uttrykker at de ville bygget anlegget også i dag.
- Prosjektet vurderes som levedyktig, hovedsakelig fordi det er av stor strategisk betydning for Norges forsvarsevne. Usikkerheten er knyttet til fordeling av driftsmidler mellom våpengrenene.
- Evaluatør vurderer skjønnsmessig prosjektet til å være samfunnsøkonomisk lønnsomt, men det er ikke mulig å beregne noen netto nåverdi.

15. Statens pensjonskasse PERFORM

Prosjektnavn: PERFORM

Sektor: IKT

Sted: Nasjonalt

Årstall påbegynt: 2008

Årstall ferdigstilt: 2012

Årstall evaluering: 2015

Styringsramme: 1006 mill. kroner (2010)

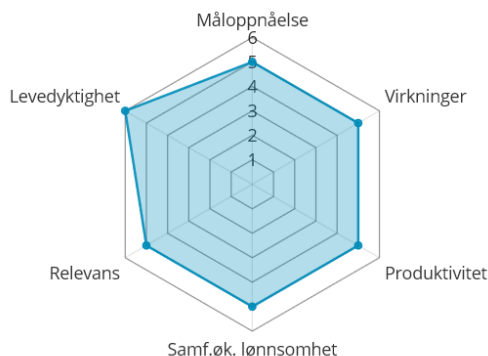
Kostnadsramme: 1287 mill. kroner (2010)

Sluttkostnad: 994 mill. kroner (2010)

Evalueringen er utført av Menon Business

Economics, i samarbeid med Vivento

Prosjektleder: Heidi Ulstein



Bakgrunn: PERFORM-prosjektet (Pensjon for fremtiden) var et stort IKT- og endringsprosjekt ved Statens Pensjonskasse som ble gjennomført i perioden 2008 til 2012. Foranledningen var Pensjonsreformen som trådte i kraft 1. januar 2011, og medførte store endringer i regelverket som krevde tilpasninger i Statens pensjonskasses produkter, arbeidsprosesser og IKT-systemer.

- Prosjektet ble gjennomført innenfor den endelige styringsrammen, og klarte å legge til rette for utbetaling av pensjon i henhold til tidsrammen. Imidlertid ble rammene økt betydelig fra KS2 2008.
- Fem av prosjektets seks effektmål er oppnådd. Etter innføringen har man stort sett klart å utbetale rett pensjon til rett tid, og gitt god informasjon underveis i prosjektet. Imidlertid er ikke målet om høy kundetilfredshet oppnådd.
- Evaluatør finner at virkningene av prosjektet var negative under prosjektperioden, men at de positive effektene i ettertid mer enn oppveier dette.
- Prosjektet har høy relevans og er i samsvar med sentrale og høyt prioriterte mål, både innad i Statens pensjonskasse, for medlemmer og kunder, og for eierdepartement.
- Levedyktigheten vurderes som god, gitt kontinuerlig drift, vedlikehold og forbedring. Systemet har også høy grad av fleksibilitet.
- Netto nåverdien til tiltaket vurderes skjønnsmessig som større enn null, men med stor grad av usikkerhet. Evaluatør vurderer den samlede effekten til å være positiv.

16. Halden fengsel

Prosjektnavn: Halden Fengsel

Sektor: Bygg

Sted: Halden

Årstall start: 2007

Årstall ferdigstilt: 2010

Årstall evaluering: 2016

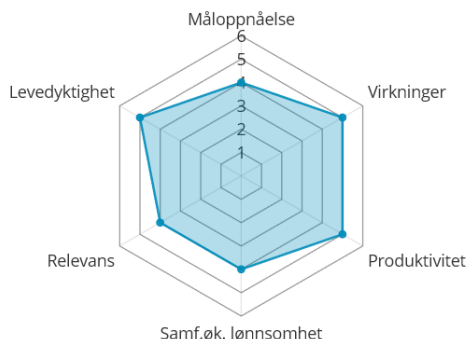
Kostnadsramme: 1477 mill. kroner (2010)

Styringsramme: 1566 mill. kroner (2010)

Sluttkostnad: 1352 mill. kroner (2010)

Evalueringen er utført av Oslo Economics i samarbeid med Tyrilistiftelsen og Sweco

Prosjektleder: Karl Rikard Løvhaug



Bakgrunn: Byggeprosjektet omfatter utbygging av et høysikkerhetsfengsel med 251 soningsplasser. Fengselet representerer en ny måte å utforme og drifte fengsel på, ved å tilby en soning som i større grad tilrettelegger for en kriminalitetsfri tilværelse etter løslatelse.

- Prosjektet ble gjennomført innenfor revidert kostnads- og tidsramme. Anlegget har god funksjonalitet og kvalitet på bygninger og uteareal.
- Halden fengsel har en kapasitet som er godt utnyttet. Tilbudet til de innsatte er godt. Det er imidlertid ikke påvist lavere sannsynlighet for tilbakefall sammenlignet med andre høysikkerhetsfengsler. Driftskostnadene er høye.
- Fengselet har blitt et utstillingsvindu der kriminalomsorgen får vist frem en praksis som etterlever prinsippene norsk straffegjennomføring skal bygge på. Aktivitetstilbud og sosial kontakt medfører positive virkninger for de innsatte. Institusjonen har bidratt til å skape arbeidsplasser i Halden-området.
- Prosjektet anses som relevant fordi det er stort behov for kapasiteten, og at det legger til rette for rehabilitering. De høye investerings- og driftskostnadene trekker imidlertid ned, og en ser at nye fengsler ikke har kopiert denne modellen.
- Prosjektet vurderes som levedyktig på grunn av fremtidig behov at bygget er fleksibelt med tanke på mulige utvidelser. Usikkerheten er hvorvidt fengselet vil få beholde ekstrabevilgningene for å dekke de relativt høye driftskostnadene.
- Den samfunnsøkonomiske lønnsomheten av prosjektet er svært usikker. Høye kostnader trekker vurderingen av lønnsomhet ned, og at det ikke er påvist lavere tilbakefall blant de innsatte enn i andre fengsler.

17. Nytt operahus i Oslo

Prosjektnavn: Nytt operahus i Oslo

Sektor: Bygg

Sted: Oslo

Årstart: 2003

Årstart ferdigstilt: 2008

Årstart evaluering: 2017

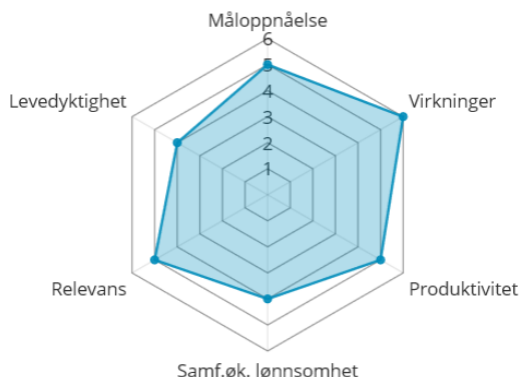
Kostnadsramme: 4356 mill. kroner (2009)

Styringsramme: 3836 mill. kroner (2009)

Sluttkost: 4278 mill. kroner (2009)

Evalueringen er utført av HR-prosjekt

Prosjektleder: Erik Whist



Bakgrunn: Byggeprosjektet omfatter etablering av nytt operabygg i Bjørvika i Oslo. Tidligere var Den Norske Opera lokalisert i Folketeateret, der man hadde plassproblemer. Det nye operahuset skulle tilrettelegge for økt produksjon og formidling av opera og ballett, med flere publikumsplasser samt produksjonslokaler.

- Prosjektet ble gjennomført til en sluttkostnad under siste oppjusterte kostnadsramme, men over styringsrammen, og åpningen fant sted seks måneder før opprinnelig plan. Prosjektet har i all hovedsak oppnådd kvalitetsmålene.
- Målsettingene for prosjektet Nytt Operahus er i all hovedsak oppnådd, bortsett fra på to punkter. Scene 2 har ikke innfridd forventningen til funksjonelle lokaler, og operaen har kun i begrenset grad oppnådd målsettingen knyttet til det å være et operahus for hele landet.
- En mener at tiltaket har hatt store positive ringvirkninger, ikke bare for operakunsten og dansekunsten, men generelt for kultursektoren. Videre har prosjektet hatt positive konsekvenser for byutviklingen i Bjørvika og Oslo.
- Prosjektet er forankret i de politiske målsettingene som gjelder på scenekunstrådet og generelt for kulturpolitikken. Evaluatoren vurderer at det også er relevant i forhold til utviklingstrekk i befolkningens kulturbruk.
- En sentral utfordring i forhold til levedyktighet er dagens lave egenfinansierungsgrad på 20 %, og fremtidige pensjonskostnader.
- Den samfunnsøkonomiske analysen av de prissatte effektene viser en negativ nåverdi av prosjektet på 17,08 mrd. kr. Det betyr at merkostnaden ved tiltaket er om lag 427 mill. kr per år i forhold til videreføring av virksomheten i Folketeateret.

18. E6 Svingenskogen-Åsgård

Prosjektnavn: E6 Svingenskogen-Åsgård

Sektor: Vei

Sted: Østfold

Årstart: 2005

Årstart ferdigstilt: 208

Årstart evaluering: 2017

Kostnadsramme: 3089 mill. kroner (2015)

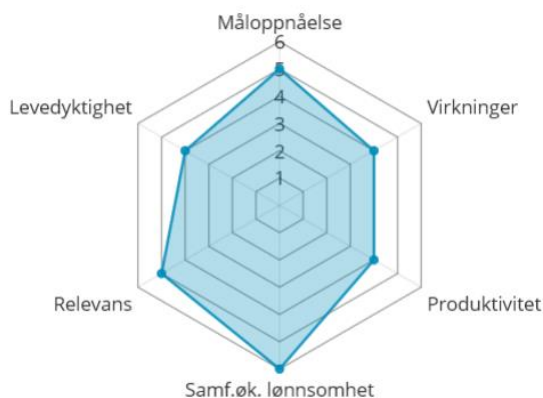
Styringsramme: 2944 mill. kroner (2015)

Sluttkost: 3034 mill. kroner (2015)

Evalueringen ble utført av Menon Business

Economics, i samarbeid med NTNU

Prosjektleder: Heidi Ulstein



Bakgrunn: Prosjektet er et delprosjekt i en større utbygging av E6 til firefelts motorvei med midtdeler gjennom Østfold fylke for å bedre trafiksikkerheten og fremkommeligheten. Veien er landets viktigste veiforbindelse til og fra utlandet

- Prosjektet ble gjennomført noe under kostnadsrammen, men over styringsrammen. Delstrekningen ble åpnet i henhold til tidsrammen, og har god kvalitet. Samlet sett vurderes produktiviteten i prosjektet som god.
- Brukerne har fått bedre fremkommelighet og vesentlig redusert reisetid. Det er økt forutsigbarhet på strekningen, til tross for noe kødannelse på enkelte punkter. Trafiksikkerheten er forbedret, både målt ved antall ulykker og alvorlighetsgrad.
- Utbyggingen har hatt positive virkninger for regionen i form av reduserte reisetider og økt konkurransekraft. Det er imidlertid også negative effekter knyttet til inngrep i dyrket mark, økt barriereeffekt for vilt og økt trafikkforurensning.
- Tiltaket er relevant i forhold til et prioritert behov for økt fremkommelighet og økt trafiksikkerhet på E6 i Østfold. Men prosjektet er i konflikt med mål om å redusere trafikkens påvirkning på miljøet.
- Veien anses som levedyktig da bedringen i trafiksikkerhet vil vedvare. Fremtidig trafikkvekst kan imidlertid true fremkommeligheten, og levedyktigheten er dermed noe usikker på lang sikt.
- En forenklet samfunnsøkonomisk analyse viser at delprosjektet er samfunnsøkonomisk lønnsomt. Netto nytte er høyere enn estimert i forkant, og også betydelig høyere enn de fleste veiprojektene som planlegges i dag.

19. E6 Åsgård-Halmstad

Prosjektnavn: E6 Åsgård-Halmstad

Sektor: Vei

Sted: Østfold

Årstart: 2004

Årstart ferdigstilt: 2005

Årstart evaluering: 2017

Kostnadsramme: 905 mill. kroner (2015)

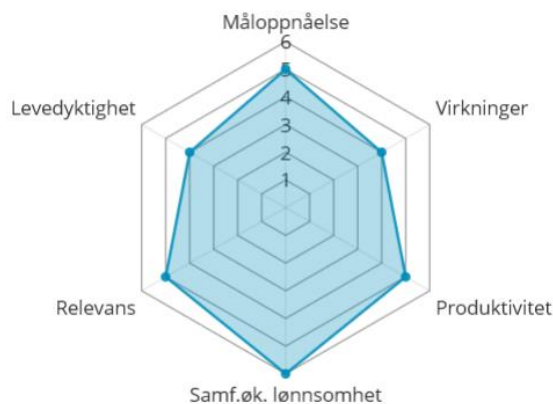
Styringsramme: 858 mill. kroner (2015)

Sluttkost: 526 mill. kroner (2015)

Evalueringen ble utført av Menon

Business Economics, i samarbeid med NTNU

Prosjektleder: Heidi Ulstein



Bakgrunn: Prosjektet er et delprosjekt i en større utbygging av E6 til firefelts motorvei med midtdeler gjennom Østfold fylke for å bedre trafikksikkerheten og fremkommeligheten. Veien er landets viktigste veiforbindelse til og fra utlandet.

- E6 Åsgård-Halmstad hadde en sluttkostnad som var betydelig under både kostnads- og styringsrammen, og delstrekningen ble åpnet i henhold til tidsrammen. Prosjektet har god kvalitet. Samlet sett vurderes produktiviteten i prosjektet som god.
- Prosjektet har medført bedre fremkommelighet, da reisetiden er vesentlig redusert og det er økt forutsigbarhet på strekningen. Trafikksikkerheten er forbedret, både målt ved antall ulykker og alvorlighetsgrad.
- Utbyggingen har hatt positive virkninger for regionen i form av reduserte reisetider og økt konkurransekraft. Det er negative effekter knyttet til inngrep i dyrket mark, økt barriereeffekt for vilt og økt forurensning som følge av økt trafikk.
- Det var et sterkt behov for økt fremkommelighet og økt trafikksikkerhet på E6 i Østfold. Prosjektet strider riktignok mot mål om å redusere trafikken påvirkning på miljøet. Det er følgelig ikke like relevant i alle perspektiver.
- Veien anses som levedyktig da bedringen i trafikksikkerhet vil vedvare. Fremtidig trafikkvekst kan imidlertid true fremkommeligheten, og levedyktigheten er dermed noe usikker på lang sikt.
- En forenklet samfunnsøkonomisk analyse viser at netto nytte er mye høyere enn estimert i forkant, og betydelig høyere enn de fleste veiprojektene som planlegges i dag.

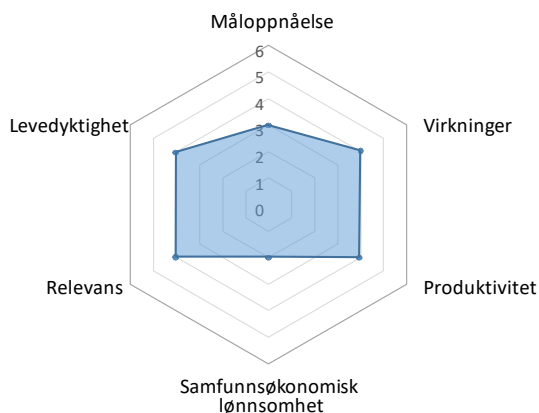
20. Gevingåsen jernbanetunnel

Prosjektnavn: Gevingåsen
jernbanetunnel
Sektor: Jernbane
Sted: mellom Trondheim og Stjørdal

Årstall start: 2009
Årstall ferdigstilt: 2011
Årstall evaluering: 2017

Kostnadsramme: 731 mill. kroner (2011)
Styringsramme: 635 mill. kroner (2011)
Sluttkost: 730 mill. kroner (2011)

Evalueringen ble utført av forskere ved
Concept-programmet, NTNU.
Prosjektleder: Morten Welde



Bakgrunn: Gevingåsen jernbanetunnel mellom Trondheim og Stjørdal var første trinn i en modernisering av Nordlandsbanens søndre del og består av 5,7 kilometer nytt spor, hvorav 4,4 kilometer i tunnel. Tunnelen har enkeltspor og åpnet for trafikk i august 2011. Tidligere trase gikk langs stranden, mens den nye traseen går gjennom Gevingåsen.

- Prosjektet ble gjennomført innenfor tidsfristen, med en sluttkostand litt under kostnadsrammen men 12 prosent over styringsrammen.
- Målet om redusert reisetid (5 minutter) er bare delvis nådd (2 minutter). Det er ingen påvist endring i antall togavganger eller punktlighet.
- Arealer i strandsonen i Hommelvik er blitt frigjort til friluftsmål og utbygging av boliger er i gang.
- Satsing på Trønderbanen var sterkt prioritert politisk. Strategisk sett er redusert lengde på den aktuelle strekning et egnet tiltak på en enkeltsporstrekning, men dobbeltspor er i tråd med eksisterende planer og redusert reisetid og forbedring av togtilbudet kunne vært oppnådd ved andre tiltak
- Evaluatør konkludere med at utbyggingen prosjektet fremstår som levedyktig, all den tid det vil fortsette å gå tog mellom Trondheim og Hell. Men dersom dobbeltspor skal bygges må det et ekstra tunneløp til.
- Investeringen har en netto nytte på minus 775 millioner kroner, og er så stor at det er lite sannsynlig at inkludering av flere mindre nyttekomponenter ville ha endret lønnsomheten vesentlig.

Vedlegg 1 Populasjon og utvalg av ferdigstilte KS-prosjekter

Listen omfatter alle prosjektene vi har sluttkostnad for per nå, ref. Welde (2017)

Prosjekt	Ferdig år	Antall hele år i drift	Evaluert (ja=kryss)
E6 Assurtjern Klemetsrud	2004	12	
E6 Skjerdingsstad - Jaktøyen	2005	11	
E6 Ny Svinesundsforbindelse	2005	11	x
FISBasis	2005	11	
Statlig kontrollområde Svinesund	2005	11	x
Golf leveranseprosjekt 1 (LP1)	2006	11	
E6 Åsgård Halmstad	2006	10	x
Dobbeltspor Sandvika Asker	2006	10	x
Svalbard forskningspark	2006	10	x
Høgskolen i Østfold Remmen (HiØ)	2007	10	x
Integrert strategisk ledelsesbygg (ISL-bygget)	2007	10	
E18 Høvik Frydenhaug	2007	9	
GSM-R	2007	9	
E18 Momarken Sekkelsten	2008	8	x
Rv 2 Kløfta Nybakk	2008	8	x
Lofoten fastlandsforbindelse 2 (Lofast 2)	2008	8	x
Eiksundsambandet	2008	8	x
Ringveg nord i Tønsbergpakken fase 1	2008	8	
E18 Kopstad Gulli	2008	8	
Nytt operahus	2008	8	x
Hærens taktiske treningssenter (HTTS)	2009	7	
E6 Svingenskogen Åsgård	2009	7	x
LOS Økonomiprojektet (ØP)	2009	7	
Pensjonsprogrammet fase 2	2009	7	
NAV IKT Basis	2009	7	x
E18 Frydenhaug Eik	2009	7	

E16 Wøyen Bjørum	2010	6	
E18 Langåker Bommestad	2010	6	
E6 Assurtjern Vinterbro	2010	6	
Lysaker stasjon	2010	6	
Rv 519 Fastlandsforb. Finnøy (Finnfast)	2010	6	x
Halden fengsel	2010	6	x
Rv 64 Atlanterhavstunnelen	2010	6	
Dobbeltspor Sandnes Stavanger	2010	6	x
E6 Nordre avlastningsveg	2011	5	
Ringveg vest byggetrinn 1	2011	5	
Høgskolen i Vestfold (HiV)	2011	5	
Nytt informatikkbygg UiO (IFI2)	2011	5	
E18 Krosby Knapstad	2011	5	
Pensjonsprogrammet fase 3	2011	5	
Skjold MTB	2012*	4	x
Fv 108 ny Kråkerøyforbindelse	2012	4	
Gevingåsen tunnel	2012	4	x
Dobbeltspor Lysaker Sandvika	2012	4	
Barkåker-Tønsberg	2012	4	
E6 Dal-Minnesund	2012	4	
Panserbekjempelse – middels rekkevidde	2012	4	
Perform	2012	4	x
Regionfelt Østlandet	2012	4	x
E18 Sky-Langangen	2013	3	
Odontologibygget Bergen	2013	3	
Fv 107 Jondalstunnelen	2013	3	
E39 Kvivsvegen	2013	3	
E6 Øyer-Tretten	2013	3	
Pansrede brukte spesialkjøretøy	2013	3	
Domus Medica (Gaustad UiO)	2013	3	
HiO Sykepleierutdanningen	2014	3	
Bremangersambandet 2	2014	2	
Patologibygget HiOA (Sykepleiehøyskolen)	2014	2	
Rv 7 Hardangerbrua	2014	2	

E6 Alta vest, Mølnes-Hjemmeluft	2014	2	
Fv. 47 T-forbindelsen	2014	2	
Rv80 Løding-Vikan	2014	2	
Fv609/Fv57 Dalsfjordsambandet	2014	2	
NASAMS II	2014	2	
EFFEKT-programmet	2014	2	
Rv7 Sokna-Ørgenvika	2015	2	
E18 Gulli-Langåker	2015	1	
Rv456 Kolsdalen-Lumberkrysset	2015	1	
E6 øst Trondheim-Stjørdal	2015	1	
Domus Media (Sentrum)	2015	1	
Rv70 Oppdølsstranda	2015	1	
Høyskolen i Bergen Samlokalisering	2015	1	
E16 Smedalsosen - Borlaug	2015	1	
E6 Gardemoen-Kolomoen	2015	1	
R6 Regjeringsbygg	2015	1	
Taktisk Data Link for fregatt og F16	2015	1	
LØFT Modulis	2015	1	
E18 Bjørvikaprojektet	2016	0	

Vedlegg 2 Sjekkliste for rapportgjennomgang. Prosjektenes suksess

Om prosjektet, diverse fakta

- Sektor
- Ansvarlig fagdepartement
- Ansvarlig etat
- Omfang (kostnadsramme)
- Grensesnitt mot andre prosjekter
- Prosjektmål (resultatmål, effektmål, samfunns mål)
- År ferdigstilt
- År evaluert

Om evaluator

- Navn på firma

Produktivitet

- Score
- Under kostnadsrammen?
- Under styringsrammen?
- Var rammene realistiske?
- Innenfor tidsmålet?
- Var tidsmålet realistisk?
- Kvalitet som forventet?
- Sentrale usikkerhetsfaktorer som slo ut
- KS2 sine anbefalinger og hvorvidt de ble fulgt

Måloppnåelse

- Score
- Vurdering av målene
- Vurdering av måloppnåelsen
- Prosjektets bidrag til måloppnåelsen

Andre virkninger

- Score
- Hvilke virkninger og retning (+/-)
- Prosjektets evne til å optimalisere andre virkninger enn de som er knyttet til målene

Relevans

- Score
- Vurdering av relevans
- Konflikter

Levedyktighet

- Score
- Vurdering av levedyktighet
- Har man fleksibilitet ved endrede behov eller rammebetingelser?

Samfunnsøkonomisk effektivitet

- Score
- Kvantifisering av lønnsomheten (NNV, N/K-brøk, e.l.)
- Vurdering av lønnsomheten (pos./neg.)
- Er prosjektet mer eller mindre lønnsomt enn forutsatt?

Vedlegg 3 Sjekkliste for rapportgjennomgang.

Gjennomføringen av evalueringene

Vurderingen av evalueringsobjektet

- Diskuteres evaluerbarhet (tidspunkt for evaluering, grensesnitt mot andre prosjekter)?
- Omtales prosjektets tilblivelse og tidligfasen?
 - Underliggende logikk
 - Utvikling av kostnadsestimater
 - Prosjektets interesser og pådrivere

Evalueringsteamet

- Fagdisipliner involvert
- Erfaring med evalueringsmodellen
- Bruk av underleverandør
- Følgforsker involvert
- Uavhengighet, habilitet

Generell vurdering av evalueringen

- Kvalitet
- Rapportstruktur

Vurderingen av prosjektets resultatkjede og målstruktur

- Vurderes kritisk av evaluator?
 - Logikken/årsak-virkningskjeden?
 - Realisme/sannsynlighet for realisering?
- Justeres/reetableres av evaluator?

Om modellen og kriteriene samlet

- Kommenterer evaluator modellens egnethet?
- Relativ vektlegging av kriteriene
 - Ressursbruk
 - Balanse operasjonelt, taktisk og strategisk perspektiv
 - Balanse målorientert perspektiv og lønnsomhet

Tolkning og vurdering av produktivitet

- Hvilken ramme vurderes sluttkostnad opp mot? Begrunnelse?
- Benchmarking av resultat?

Tolkning og vurdering av måloppnåelse

- Metode(r) for måling av effekter
- Benchmarking av resultat?

Tolkning og vurdering av andre virkninger

-
- Tilsiktede eller kun utilsiktede?
 - Direkte og/eller indirekte?
 - Sårbare grupper?

Tolkning og vurdering av relevans

- Dekkes de tre perspektivene (politisk, ettersp., interessegruppe?)
- Eksplisitt/ implisitt rangering?
- Konflikt mellom perspektivene?
- Problematisering av etterspørsel etter noe som er gratis?

Tolkning og vurdering av levedyktighet

- Hvor langt tidsperspektiv?
- Dekkes de tre perspektivene økonomi, miljø, sosiale hensyn?
- Vurderes fleksibilitet i forhold til endrede behov?

Tolkning og vurdering av samfunnsøkonomisk effektivitet

- Analysetype (nytte-kostnadsanalyse, kostnadseffektivitetsanalyse, kostnads-virkningsanalyse)
- Referansegrunnlag
- Direkte eller også indirekte virkninger
- Sammenheng mellom det som tas med på nytte- og kostnadssiden
- Kommer forbehold og forutsetninger frem?
- Transparens
- Håndtering av ikke-prissatte virkninger

Datagrunnlag

- Tilgang til data?

Valg av datainnsamlings- og analysemetoder

- Bredde og triangulering

Scoresetting

- Tendens til ukritisk (positiv) vurdering?
- Tendens til altfor kritisk (negativ) vurdering?

Læringspunkter

- Forklares resultatene?
- Anbefalinger til tidligfase i kommende prosjekter

Vedlegg 4 Fokusgruppemøte

Møte avholdt på Hotell Bristol 17. mars 2017 over en halv dag

Deltakere:

Erik Whist, Scanteam

Heidi Ulstein, Menon Business Economics

Kristina Wifstad, Menon Business Economics

Hanne Samstad, COWI

Rolf Sverre Asp, Oslo Economics

Marit Svensgaard, Oslo Economics

Kjell Ove Kalhagen, HR-prosjekt

Håkon Finne, SINTEF Teknologi og samfunn

Morten Welde, NTNU

Knut Samset, NTNU

Gro Holst Volden, NTNU

Concept rapportserie

Papirtrykk: ISSN 0803-9763

Elektronisk utgave på internett: ISSN 0804-5585

Lastes ned fra: www.ntnu.no/concept/publikasjoner/rapportserie

Rapport	Tittel	Forfatter
Nr. 1	Styring av prosjektporteføljer i staten. Usikkerhetsavsetning på porteføljnivå <i>Project Portfolio Management. Estimating Provisions for Uncertainty at Portfolio Level.</i>	Stein Berntsen og Thorleif Sunde
Nr. 2	Statlig styring av prosjektledelse. Empiri og økonomiske prinsipper. <i>Economic Incentives in Public Project Management</i>	Dag Morten Dalen, Ola Lædre og Christian Riis
Nr. 3	Beslutningsunderlag og beslutninger i store statlige investeringsprosjekt <i>Decisions and the Basis for Decisions in Major Public Investment Projects</i>	Stein V. Larsen, Eilif Holte og Sverre Haanæs
Nr. 4	Konseptutvikling og evaluering i store statlige investeringsprosjekt <i>Concept Development and Evaluation in Major Public Investment Projects</i>	Hege Gry Solheim, Erik Dammen, Håvard O. Skaldebø, Eystein Myking, Elisabeth K. Svendsen og Paul Torgersen
Nr. 5	Bedre behovsanalyser. Erfaringer og anbefalinger om behovsanalyser i store offentlige investeringsprosjekt <i>Needs Analysis in Major Public Investment Projects. Lessons and Recommendations</i>	Petter Næss
Nr. 6	Målformulering i store statlige investeringsprosjekt <i>Alignment of Objectives in Major Public Investment Projects</i>	Ole Jonny Klakegg
Nr. 7	Hvordan trur vi at det blir? Effektvurderinger av store offentlige prosjekt <i>Up-front Conjecture of Anticipated Effects of Major Public Investment Projects</i>	Nils Olsson
Nr. 8	Realopsjoner og fleksibilitet i store offentlige investeringsprosjekt <i>Real Options and Flexibility in Major Public Investment Projects</i>	Kjell Arne Brekke
Nr. 9	Bedre utforming av store offentlige investeringsprosjekter. Vurdering av behov, mål og effekt i tidligfasen <i>Improved Design of Public Investment Projects. Up-front Appraisal of Needs, Objectives and Effects</i>	Petter Næss med bidrag fra Kjell Arne Brekke, Nils Olsson og Ole Jonny Klakegg
Nr. 10	Usikkerhetsanalyse – Kontekst og grunnlag <i>Uncertainty Analysis – Context and Foundations</i>	Kjell Austeng, Olav Torp, Jon Terje Midtbø, Ingemund Jordanger, og Ole M Magnussen

Rapport	Tittel	Forfatter
Nr. 11	Usikkerhetsanalyse – Modellering, estimering og beregning <i>Uncertainty Analysis – Modeling, Estimation and Calculation</i>	Frode Drevland, Kjell Austeng og Olav Torp
Nr. 12	Metoder for usikkerhetsanalyse <i>Uncertainty Analysis – Methodology</i>	Kjell Austeng, Jon Terje Midtbø, Vidar Helland, Olav Torp og Ingemund Jordanger
Nr. 13	Usikkerhetsanalyse – Feilkilder i metode og beregning <i>Uncertainty Analysis – Methodological Errors in Data and Analysis</i>	Kjell Austeng, Vibeke Binz og Frode Drevland
Nr. 14	Positiv usikkerhet og økt verdiskaping <i>Positive Uncertainty and Increasing Return on Investments</i>	Ingemund Jordanger
Nr. 15	Kostnadsusikkerhet i store statlige investeringsprosjekter; Empiriske studier basert på KS2 <i>Cost Uncertainty in Large Public Investment Projects. Empirical Studies</i>	Olav Torp (red.), Ole M Magnussen, Nils Olsson og Ole Jonny Klakegg
Nr. 16	Kontrahering i prosjektets tidligfase. Forsvarets anskaffelser. <i>Procurement in a Project's Early Phases. Defense Aquisitions</i>	Erik N. Warberg
Nr. 17	Beslutninger på svakt informasjonsgrunnlag. Tilnærminger og utfordringer i prosjekters tidlige fase <i>Decisions Based on Scant Information. Challenges and Tools During the Front-end Phases of Projects</i>	Kjell Sunnevåg (red.)
Nr. 18	Flermålsanalyser i store statlige investeringsprosjekt <i>Multi-Criteria Decision Analysis In Major Public Investment Projects</i>	Ingemund Jordanger, Stein Malerud, Harald Minken, Arvid Strand
Nr. 19	Effektvurdering av store statlige investeringsprosjekter <i>Impact Assessment of Major Public Investment Projects</i>	Bjørn Andersen, Svein Bråthen, Tom Fagerhaug, Ola Nafstad, Petter Næss og Nils Olsson
Nr. 20	Investorers vurdering av prosjekters godhet <i>Investors' Appraisal of Project Feasibility</i>	Nils Olsson, Stein Frydenberg, Erik W. Jakobsen, Svein Arne Jessen, Roger Sørheim og Lillian Waagø
Nr. 21	Logisk minimalisme, rasjonalitet - og de avgjørende valg <i>Major Projects: Logical Minimalism, Rationality and Grand Choices</i>	Knut Samset, Arvid Strand og Vincent F. Hendricks
Nr. 22	Miljøøkonomi og samfunnsøkonomisk lønnsomhet <i>Environmental Economics and Economic Viability</i>	Kåre P. Hagen
Nr. 23	The Norwegian Front-End Governance Regime of Major Public Projects – A Theoretically Based Analysis and Evaluation	Tom Christensen
Nr. 24	Markedsorienterte styringsmetoder i miljøpolitikken <i>Market oriented approaches to environmental policy</i>	Kåre P. Hagen

Rapport	Tittel	Forfatter
Nr. 25	Regime for planlegging og beslutning i sykehusprosjekter <i>Planning and Decision Making in Hospital Projects. Lessons with the Norwegian Governance Scheme.</i>	Asmund Myrbostad, Tarald Rohde, Pål Martinussen og Marte Lauvsnes
Nr. 26	Politisk styring, lokal rasjonalitet og komplekse koalisjoner. Tidligfaseprosessen i store offentlige investeringsprosjekter <i>Political Control, Local Rationality and Complex Coalitions. Focus on the Front-End of Large Public Investment Projects</i>	Erik Whist, Tom Christensen
Nr. 27	Verdsetting av fremtiden. Tidshorisont og diskonteringsrenter <i>Valuing the future. Time Horizon and Discount Rates</i>	Kåre P. Hagen
Nr. 28	Fjorden, byen og operaen. En evaluering av Bjørvikautbyggingen i et beslutningsteoretisk perspektiv <i>The Fjord, the City and the Opera. An Evaluation of Bjørvika Urban Development</i>	Erik Whist, Tom Christensen
Nr. 29	Levedyktighet og investeringstiltak. Erfaringer fra kvalitetssikring av statlige investeringsprosjekter <i>Sustainability and Public Investments. Lessons from Major Public Investment Projects</i>	Ola Lædre, Gro Holst Volden, Tore Haavaldsen
Nr. 30	Etterevaluering av statlige investeringsprosjekter. Konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter <i>Evaluating Public Investment Projects. Lessons and Advice from a Meta-Evaluation of Four Projects</i>	Gro Holst Volden og Knut Samset
Nr. 31	Store statlige investeringers betydning for konkurranse- og markedsutviklingen. Håndtering av konkurransemessige problemstillinger i utredningsfasen <i>Major Public Investments' Impact on Competition. How to Deal with Competition Issues as Part of the Project Appraisal</i>	Asbjørn Englund, Harald Bergh, Aleksander Møll og Ove Skaug Halsos
Nr. 32	Analyse av systematisk usikkerhet i norsk økonomi. <i>Analysis of Systematic Uncertainty in the Norwegian Economy.</i>	Haakon Vennemo, Michael Hoel og Henning Wahlquist
Nr. 33	Planprosesser, beregningsverktøy og bruk av nytte-kostnadsanalyser i vegsektoren. En sammenlikning av praksis i Norge og Sverige. <i>Planning, Analytic Tools and the Use of Cost-Benefit Analysis in the Transport Sector in Norway and Sweden.</i>	Morten Welde, Jonas Eliasson, James Odeck, Maria Börjesson
Nr. 34	Mulighetsrommet. En studie om konseptutredninger og konseptvalg <i>The Opportunity Space. A Study of Conceptual Appraisals and the Choice of Conceptual Solutions.</i>	Knut Samset, Bjørn Andersen og Kjell Austeng

Rapport	Tittel	Forfatter
Nr. 35	Statens prosjektmodell. Bedre kostnadsstyring. Erfaringer med de første investeringstiltakene som har vært gjennom ekstern kvalitetssikring	Knut Samset og Gro Holst Volden
Nr. 36	Investing for Impact. Lessons with the Norwegian State Project Model and the First Investment Projects that Have Been Subjected to External Quality Assurance	Knut Samset og Gro Holst Volden
Nr. 37	Bruk av karbonpriser i praktiske samfunnsøkonomiske analyser. En oversikt over praksis fra analyser av statlige investeringsprosjekter under KVV-/KS1-ordningen. <i>Use of Carbon Prices in Cost-Benefit Analysis. Practices in Project Appraisals of Major Public Investment Projects under the Norwegian State Project Model</i>	Gro Holst Volden
Nr. 38	Ikke-prissatte virkninger i samfunnsøkonomisk analyse. Praksis og erfaringer i statlige investeringsprosjekter <i>Non-Monetized Impacts in Economic Analysis. Practice and Lessons from Public Investment Projects</i>	Heidi Bull-Berg, Gro Holst Volden og Inger Lise Tyholt Grindvoll
Nr. 39	Lav prising – store valg. En studie av underestimering av kostnader i prosjekters tidligfase <i>Low estimates – high stakes. A study of underestimation of costs in projects' earliest phase</i>	Morten Welde, Knut Samset, Bjørn Andersen, Kjell Austeng
Nr. 40	Mot sin hensikt. Perverse insentiver – om offentlige investeringsprosjekter som ikke forplikter <i>Perverse incentives and counterproductive investments. Public funding without liabilities for the recipients</i>	Knut Samset, Gro Holst Volden, Morten Welde og Heidi Bull-Berg
Nr. 41	Transportmodeller på randen. En utforskning av NTM5-modellens anvendelsesområde <i>Transport models and extreme scenarios. A test of the NTM5 model</i>	Christian Steinsland og Lasse Fridstrøm
Nr. 42	Brukeravgifter i veisektoren <i>User fees in the road sector</i>	Kåre Petter Hagen og Karl Rolf Pedersen
Nr. 43	Norsk vegplanlegging: Hvilke hensyn styrer anbefalingene <i>Road Planning in Norway: What governs the selection of projects?</i>	Arvid Strand, Silvia Olsen, Merethe Dotterud Leiren og Askill Harkjerr Halse
Nr. 44	Ressursbruk i transportsektoren – noen mulige forbedringer <i>Resource allocation in the transport sector – some potential improvements</i>	James Odeck (red.) og Morten Welde (red.)
Nr. 45	Kommunale investeringsprosjekter. Prosjektmodeller og krav til beslutningsunderlag. <i>Municipal investment practices in Norway</i>	Morten Welde, Jostein Aksdal og Inger Lise Tyholt Grindvoll
Nr. 46	Styringsregimer for store offentlige prosjekter. En sammenliknende studie av prinsipper og praksis i seks land.	Knut Samset, Gro Holst Volden, Nils Olsson og Eirik Vårdal Kvalheim

Rapport	Tittel	Forfatter
Nr. 47	Governance schemes for major public investment projects: A comparative study of principles and practices in six countries	Knut Samset, Gro Holst Volden, Nils Olsson og Eirik Vårdal Kvalheim
Nr. 48	Investeringsprosjekter og miljøkonsekvenser. En antologi med bidrag fra 16 forskere <i>Environmental impact of large investment projects. An anthology by 16 Norwegian experts</i>	Kåre P. Hagen (red.) og Gro Holst Volden (red.)
Nr. 49	Finansiering av vegprosjekter med bompenger. Behandling av og konsekvenser av bompenger i samfunnsøkonomiske analyser. <i>Financing road projects with tolls. The treatment of and consequences of tolls in cost benefit analyses.</i>	Morten Welde, Svein Bråthen, Jens Rekdahl og Wei Zhang
Nr. 50	Prosjektmodeller og prosjekteierstyring i statlige virksomheter <i>Project governance and the use of project models in public agencies and line ministries in Norway</i>	Bjørn Andersen, Eirik Vårdal Kvalheim og Gro Holst Volden
Nr. 51	Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitetssikring <i>Cost performance in government investment projects that have been subjected to external quality assurance.</i>	Morten Welde
Nr. 52	Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene <i>A Close-up on Public Investment Cases. Lessons from Ex-post Evaluations of 20 Major Norwegian Projects</i>	Gro Holst Volden og Knut Samset

Forskningsprogrammet Concept skal utvikle kunnskap som sikrer bedre ressursutnytting 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 Fakultet for ingeniørvitenskap og teknologi. Programmet samarbeider med ledende norske og internasjonale fagmiljøer og universiteter, og er finansiert av Finansdepartementet.

The Concept research program aims to develop know-how to help make more efficient use of resources and improve the effect of major public investments. The Program is designed to follow up on the largest public projects over a period of several years, and help improve design and quality assurance of future public projects before they are formally approved.

The program is based at The Norwegian University of Science and Technology (NTNU), Faculty of Engineering Science and Technology. It cooperates with key Norwegian and international professional institutions and universities, and is financed by the Norwegian Ministry of Finance.

Address:

The Concept Research Program
Høgskoleringen 7A
N-7491 NTNU
Trondheim
NORWAY

ISSN: 0803-9763 (paper version)

ISSN: 0804-5585 (web version)

ISBN: 978-82-93253-61-7 (paper version)

ISBN: 978-82-93253-62-4 (web version)

