



SINTEF



# Rapport

## En modell for evaluering av arbeid med planlegging og bygging av vegstrekninger

### Forfatter(e):

Håkon Finne, Lone-Eirin Lervåg, Lone Sletbakk Ramstad og Morten Welde

### Rapportnummer:

2022:01207 - Åpen

**Oppdragsgiver:** Nye Veier AS

# Rapport

## En modell for evaluering av arbeid med planlegging og bygging av vegstrekninger

**EMNEORD**  
Evalueringsmodell  
Vegprosjekter**VERSJON**  
1.0**DATO**  
2022-11-17**FORFATTER(E)**

Håkon Finne, Lone-Eirin Lervåg, Lone Sletbakk Ramstad og Morten Welde

**OPPDRAGSGIVER(E)**

Nye Veier AS

**OPPDRAGSGIVERS REFERANSE**

Dag Yngvar Åsland

**PROSJEKTNUMMER**

102027259

**ANTALL SIDER OG VEDLEGG**

54

**SAMMENDRAG**

Rapporten presenterer et rammeverk for en modell for evaluering av arbeid med planlegging og bygging av vegstrekninger, og anvisninger for hvordan modellen kan brukes før, under og etter gjennomføring av slike vegprosjekter, anvendt på hele prosjektet eller deler av det. Modellens seks evalueringskriterier (effektivitet i prosjektgjennomføring, realisering av effekter, relevans, levedyktighet og samfunnsøkonomisk lønnsomhet) er hentet fra Concept-programmets modell for etterevaluering av store investeringsprosjekter under statens prosjektmodell. Et sentralt metodisk grep, i tillegg til vanlige effektstudier, er å teste prosjektets programteori (antakelser om hvorfor tiltaket skal nå sine resultat-, effekt- og samfunns mål) og de dertil hørende endringsmekanismene (som ligger bak disse antakelsene) for å identifisere læringsmuligheter for fremtidig prosjektarbeid. Modellen legger også til rette for at læringspunktene enklere skal kunne implementeres i klientorganisasjonen. En kortfattet veileder er tilgjengelig separat.

**UTARBEIDET AV**

Håkon Finne

SIGNATUR

**KONTROLLERT AV**

Hans Torvatn

SIGNATUR

**GODKJENT AV**

Espen Gressetvold

SIGNATUR



## Forord

Denne rapporten er utarbeidet på oppdrag av Nye Veier. Prosjektet Utvikling av metode for ex post evalueringer av Nye Veiers arbeid med strekninger i oppstartporteføljen har vært gjennomført i perioden 1. mars 2022 til 17. november 2022. Det overordnede målet for arbeidet har vært å utarbeide en metode for ex post evalueringer av Nye Veiers arbeid med strekninger i oppstartporteføljen. Sentrale spørsmål som metodikken skal besvare er:

- Har Nye Veier levert på sitt samfunnsoppdrag?
- Hvilke forbedringspunkter har ex post evalueringen avdekket?
- Hvilke forbedringspunkter er viktigst å ta fatt i (prioritering)?
- Hva er forutsetningene for en vellykket gjennomføring (implementering) av forbedringspunkter som ex post evalueringen avdekket?

Oppdragsgiver har lagt vekt på at kunnskap fra ex post evalueringer skal benyttes til forbedringsarbeid i Nye Veier i samsvar med selskapets verdier; forsikre, forbedre og fornye. Det har vært et viktig premiss for arbeidet med å utvikle en modell for evaluering av arbeid med planlegging og bygging av vegstrekninger.

Det er også laget en egen veileder som gir en kortfattet innføring i bruken av evalueringsmodellen, *Veileder for evalueringer i Nye Veier, case utbyggingsprosjekter*. Veilederen omfatter en trinnvis beskrivelse av de ulike oppgavene i evalueringsprosessen, i hovedsak rettet mot den som skal gjennomføre en evaluering.

Rapporten har fire kapitler. Kapittel 1 gir en redegjørelse av bakgrunnen for evalueringsmodellen. Deretter begrunnes og dokumenteres modellens rammeverk ut fra forskningsbasert kunnskap og teorier om evaluering av store investeringsprosjekter i kapittel 2. Etter dette presenteres evalueringsmodellens ulike trinn, fra forarbeid i Trinn 0 til gjennomføring av selve evalueringen i Trinn 1-5, og etterarbeid i Trinn 6. Etterarbeidet omfatter implementering, læring og forbedringsarbeid i Nye Veier (kapittel 3). Til slutt oppsummerer kapittel 4 styrker og svakheter ved modellen.

Arbeidet er gjennomført av et forskerteam fra SINTEF og NTNU; Håkon Finne, Lone-Eirin Lervåg og Lone Sletbakk Ramstad fra SINTEF, og Morten Welde fra NTNU (Concept-programmet). Vi har fått tilgang til mye data og dokumentasjon i Nye Veier, og gjennomført flere intervju med personer i ulike roller. Vi takker alle som har stilt opp i arbeidet fra Nye Veier. En spesiell takk til Dag Yngvar Åsland for tett oppfølging og god tilrettelegging gjennom hele prosjektperioden, og til prosjekteier Øystein Engedal.

En bredt sammensatt arbeidsgruppe i Nye Veier har fulgt arbeidet med utvikling av evalueringsmodellen, fra oppstart til ferdigstilling. De har gitt innspill og tilbakemelding i flere faser av utviklingsarbeidet gjennom workshops og møter. Dette har vært viktige bidrag i arbeidet. Vi vil rette en stor takk til dere alle.

Trondheim, 2022-11-17

Lone Sletbakk Ramstad  
Prosjektleder

## Innholdsfortegnelse

<b>EN MODELL FOR EVALUERING AV ARBEID MED PLANLEGGING OG BYGGING AV VEGSTREKNINGER ....</b>	<b>1</b>
<b>Forord .....</b>	<b>3</b>
<b>Innholdsfortegnelse .....</b>	<b>4</b>
<b>1 Bakgrunn for evalueringsmodellen .....</b>	<b>5</b>
<b>2 Evalueringsmodellens rammeverk .....</b>	<b>8</b>
2.1 Innledning .....	8
2.2 Terminologi og grunnleggende evalueringstenking .....	8
2.3 Seks evalueringskriterier .....	12
2.4 Mål og målstruktur i en programteoretisk forståelse .....	16
2.5 Ulike interessenters rolle i dette rammeverket .....	20
2.6 Forholdet mellom mål og evalueringskriterier .....	21
2.7 Design og metodetilfang .....	23
2.8 Læring .....	27
<b>3 Evalueringsmodellen i bruk .....</b>	<b>30</b>
3.1 Fremgangsmåte .....	30
3.2 Trinn 0 Forarbeid: Avgrense evalueringsobjekt, formål og omfang .....	32
3.3 Trinn 1: Kartlegge behov, målformuleringer og programteori .....	34
3.4 Trinn 2: Velge evalueringskriterier og evalueringsspørsmål .....	36
3.5 Trinn 3: Velge design, metoder og datakilder, og samle data .....	40
3.6 Trinn 4: Analysere dataene .....	41
3.7 Trinn 5: Oppsummere resultater og anbefalinger .....	44
3.8 Trinn 6 Etterarbeid: Implementere læring .....	45
<b>4 Evalueringsmodellens styrker og svakheter .....</b>	<b>48</b>
4.1 Viktige egenskaper ved evalueringsmodellen .....	48
4.2 Hva modellen er, og ikke er, egnet til .....	49
<b>5 Litteraturreferanser .....</b>	<b>51</b>
<b>Vedlegg .....</b>	<b>53</b>

## 1 Bakgrunn for evalueringsmodellen

Vi presenterer med dette en modell for å planlegge og gjennomføre en avgrenset type evalueringer i vegsektoren. Modellen er beskrevet i to dokumenter: en kortfattet veileder på ca. 20 sider, som beskriver hvordan evalueringer kan gjennomføres i tråd med modellen, og dette bakgrunnsdokumentet, som beskriver, utdyper, forankrer, kontekstualiserer og begrunner modellen og dens enkelte elementer. Veilederen er skrevet for de som skal utføre evalueringene, men den er skrevet slik at den med fordel også kan og bør leses av de som bestiller evalueringer og de som er ansvarlige for de forhold som skal evalueres. Dette vil gjøre alle parter kjent med både selve evalueringsprosessen og deres rolle i den, og med begrepsbruken. Bakgrunnsdokumentet (denne teksten) gir grunnlag for dypdykk i modellen og dens enkelte elementer for de som har behov for det.

Modellen og veilederen er utarbeidet av forskere i SINTEF, med bistand fra NTNU, på oppdrag fra Nye Veier AS. Oppdraget omfatter utarbeidelsen av en metode eller metodikk, eller **modell**<sup>1</sup>, for evaluering av Nye Veiers arbeid med sine tildelte vegstrekninger. Evalueringsmodellen skal ifølge oppdragsdokumentene benyttes til å besvare sentrale spørsmål som vist i Tabell 1:

**Tabell 1: Sentrale spørsmål som evalueringsmodellen skal brukes til å besvare**

Evalueringsoppgave
1) Har Nye Veier levert på sitt samfunnsoppdrag?
2) Hvilke forbedringspunkter har evalueringen avdekket?
3) Hvilke forbedringspunkter er viktigst å ta fatt i (prioritering)?
4) Hva er forutsetningene for en vellykket gjennomføring (implementering) av forbedringspunkter som evalueringen avdekket?

Den evalueringsmodellen vi presenterer, kan brukes til å besvare de fire spørsmålene, og den tilfredsstillende selv sagt faglige, praktiske og etiske krav til gode evalueringer.

Evalueringsmodellen er basert på **Concept-modellen** (Concept-programmet 2021), en modell utviklet av forskningsprogrammet Concept ved NTNU for **etterevaluering** av vegprosjekter og andre store statlige investeringsprosjekter som er underlagt **statens prosjektmodell for store investeringer** (FIN 2019, 2021). I tilpassingen til Nye Veiers behov har vi dratt nytte av intervjuer med et titalls ledere og eksperter i Nye Veiers organisasjon om behov for, og praksis og erfaringer med, å bruke evaluering til læring og kontroll. Vi har også dratt nytte av vår egen og andres erfaringer med andre evalueringstilnæringer med et spesielt fokus på læring (se særlig Finne mfl. 1995; Lervåg 2021).

Concept har basert sin modell på den modell som OECD har utviklet for evaluering av utviklings tiltak av alle slag (OECD 1991; OECD/DAC Network on Development Evaluation 2019), og tilpasset den til store statlige investeringsprosjekter, særlig i infrastruktur, som går gjennom prosjektutviklings- og gjennomføringsforløp spesifisert i den norske statens prosjektmodell. Per november 2022 har Concept selv, foruten en rekke forsknings- og rådgivermiljøer, med hell evaluert i alt 34 prosjekter, hvorav 12 i vegsektoren, etter Concept-modellen.

<sup>1</sup> Ordene metode og metodikk brukes både i dagligtalen og i ulike fagmiljøer om hverandre. Begge ordene kan vise både til konkrete (tekniske) metoder og til et overordnet metodisk rammeverk. Vi knytter heretter metodebegrepet til innsamling, behandling og analyse av data, tilpasset de problemstillinger som skal belyses eller spørsmål som skal besvares. For å unngå forvekslinger, kaller vi hele tilnærmingen en **evalueringsmodell**. Dette gir også bedre rom for å spesifisere ikke-metodiske forhold. Videre sier vi at evaluator setter sammen en metodikk for det enkelte evalueringsprosjekt, med utgangspunkt i evalueringsmodellens rammeverk.

Concept finansieres av Finansdepartementet og arbeider med kunnskapsgrunnlaget for statens prosjektmodell. Nye Veiers arbeid er ikke direkte underlagt statens prosjektmodell, men Nye Veier har et planleggings- og kvalitetssikringsregime som ivaretar noenlunde de samme funksjonene som ordningen med statens prosjektmodell anvendt på andre vegbyggingsprosjekter (Volden og Andersen 2019). Det gir derfor god mening å bygge denne evalueringsmodellen på designet fra Concept-modellen og erfaringene fra bruken av den i etterevaluering av vegprosjekter.

Concept-modellen slik den er brukt i vegsektoren viser seg å være godt egnet til å besvare det første av Nye Veiers sentrale spørsmål ovenfor: om Nye Veier leverer på sitt samfunnsoppdrag gjennom de enkelte vegprosjektene. Dette er evalueringens **kontrollfunksjon**. Merk at dette ikke er formulert som at de har fulgt alle pålegg, men at de har oppnådd alle de mål som de har fått ressurser til å nå. De tre påfølgende sentrale spørsmålene handler om å legge til rette for læring fra evalueringsprosjektene, i vegsektoren generelt og hos Nye Veier spesielt. Selv om læring også er adressert i Concept-modellen, er det ikke like systematisk behandlet der som det forutsettes i oppdraget fra Nye Veier. Styrkingen av evalueringens **læringsfunksjon** er her ivare tatt på tre hovedmåter:

1. For det første er selve evalueringsmodellen styrket på å **forklare hva det er som forårsaker de positive eller negative prosjektutfall som avdekkes**. Dette gjør vi ved å videreutvikle Concept-modellens fokus på prosjektenes **målstruktur**, som vi bygger inn i en tydeligere årsaksforklaring<sup>2</sup> kalt en **programteori**. Her trekker vi på en evalueringsfaglig tradisjon kalt **teoribasert evaluering** (Chen 1990; van der Knaap 2004), og tilnærmingen rimer godt med årsakskart som brukes i sektoren generelt og i Nye Veier spesielt.
2. For det andre tilpasser vi modellen til **bruk i både tidlige og sene faser av et vegprosjekt**. Evaluering av et nytt element i for eksempel planlegging trenger da ikke vente til den inngår i etterevalueringen av hele vegprosjektet mange år senere. Slik kan evalueringsresultatene få bedre **aktualitet**.
3. For det tredje er arbeidet med **formidling av evalueringsresultater** utdypet, og vi har eksplisitt behandlet **hvordan evalueringsresultater kan omsettes til revidert praksis i sektoren**. Dette har vi først og fremst gjort ved å justere hvordan selve evalueringen planlegges, gjennomføres og rapporteres. Dernest har vi også sett på hvordan Nye Veier kan ta evalueringsresultater inn over seg og innarbeide dem i sine egne prosesser for innovasjon og kontinuerlig forbedring. Dette kan med andre ord innebære en gjensidig tilpasning mellom evalueringsopplegget og Nye Veiers måte å forholde seg til evalueringer på.

Kontrollfunksjonen og læringsfunksjonen kan noen ganger komme i konflikt med hverandre. Modeller som legger spesielt godt til rette for underveislæring gjennom sterk interaksjon mellom evaluator og evalueringsklient, som enkelte former for følgeforskning eller følgeevaluering, kan erfare denne konflikten (Olsen og Lindøe 2004). Vi løser dette med full transparens og med et skille mellom ekstern og intern bruk av modellen. Uavhengige evalueringer av hele eller deler av et prosjekt må gjøres av eksterne. Dette er en forutsetning for evalueringens kontrollfunksjon. Eksterne kan også med fordel gjennomføre evalueringer der læringsformålet er fremtredende, ikke minst fordi de også bringer inn andre perspektiver enn de som finnes i organisasjonen. Bruk av elementer i modellen til monitorering, styring og internevaluering av enkeltpraksiser, enkeltprosjekter og prosjektporteføljer kan gjøres av interne, men den uavhengige kontrollfunksjonen faller da bort.

<sup>2</sup> Vanligvis snakker vi om et årsaks-/virkningsforhold, men i det uttrykket omfatter virkning både det vi ellers i teksten har kalt effekt og det vi har kalt virkning, så vi kaller det en årsaksforklaring for å holde ordbruken konsistent.

Oppbyggingen av resten av dette dokumentet er som følger:

I kapittel 2 gjennomgås evalueringsmodellens rammeverk og dens faglige basis, med vekt på det som skiller modellen fra mange andre evalueringsmodeller.

I kapittel 3 beskrives hvordan evalueringsmodellen kan brukes til konkrete evalueringer. Utgangspunktet er en fullskala evaluering av Nye Veiers arbeid med en trafikkfattig vegstrekning som de har planlagt og vært byggherre for. Gjennomgangen i disse kapitlene viser også hvordan evalueringsmodellen kan anvendes til evaluering av kortere faser i slike prosjekter, på ulike tidspunkt, og med ulike balanser mellom læring og kontroll som formål.

I kapittel 4 synliggjøres noen sterke og svake sider ved modellen i forhold til formålet, og vi diskuterer grenseflater mellom evaluering etter denne modellen og andre, beslektede former for kunnskapsgenerering.

## 2 Evalueringsmodellens rammeverk

### 2.1 Innledning

Nye Veiers behov for en evalueringsmodell som ivaretar både kontroll og læring er grunnlaget for denne modellen. For å svare ut dette behovet spesielt, har rammeverket to hovedelementer som hver for seg og sammen understøtter begge funksjoner, og i tillegg et element som understøtter læring spesielt:

- **Seks evalueringskriterier** for en normativ bedømmelse av det enkelte prosjektets vellykkethet eller verdi<sup>3</sup>
- En **evalueringsteoretisk forståelse** av **målstrukturen** i et prosjekt som ryggraden i en **programteori** om hvordan aktiviteter skaper resultater, som igjen er en forutsetning for tilskittede effekter, som igjen bidrar til tilskittede virkninger, alt innenfor varierende kontekster
- En **tilnærming til læring** som gir føringer for hvordan evalueringen gjennomføres og evalueringresultatene formidles, og som gir Nye Veier noen idéer om hvordan de selv kan forankre evalueringsaktiviteter i egen organisasjon og legge til rette for læring fra evalueringer.

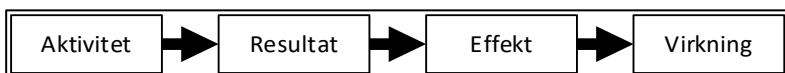
Dessuten omfatter rammeverket også elementer som overlapper med de mange andre evalueringsmodeller, nemlig

- Et **tilfang av design og metoder** som tilpasses det enkelte evalueringsprosjektets formål og evalueringsobjekt, og som holder høy evalueringfaglig standard.

I det følgende går vi gjennom disse fire elementene før vi går over til å se på hvordan de kan og bør brukes i evalueringsprosjekter under denne modellen. Vi trenger imidlertid først å fastlegge en enkel begrepsbruk for å unngå sammenblandinger.

### 2.2 Terminologi og grunnleggende evalueringstenking

Den terminologien vi bruker for å evaluere **utfallene** (eller konsekvensene) av en aktivitet (som et prosjekt), er vist i Figur 1, som er en mye brukt måte å framstille en slik **utfallskjede** på.<sup>4</sup>



**Figur 1: Utfallskjede for evaluering av en aktivitet**

Hver boks inneholder normalt mange elementer; i denne figuren er disse bare samlet sammen under en enkelt betegnelse. **Aktiviteter** fører til direkte utfall som vi kaller **resultater**, i første rekke leveranser med gitte egenskaper, for eksempel en veg med en viss standard mellom to punkter, og med tilslutningspunkter underveis. Disse resultatene medfører i neste omgang utfall som vi kaller **effekter**, etter hvert som leveransen settes i drift og/eller tas i bruk. Dette er ofte effekter på bruker-

<sup>3</sup> "Verdi" kan omfatte både økonomiske og andre verdier. "Vellykkethet" brukes her, som etter Concepts konvensjon, som et begrep som kan omfatte flere interessentperspektiver, i motsetning til "suksess", som vanligvis er knyttet til ett interessentperspektiv, om enn ofte som et kompromiss mellom flere parter.

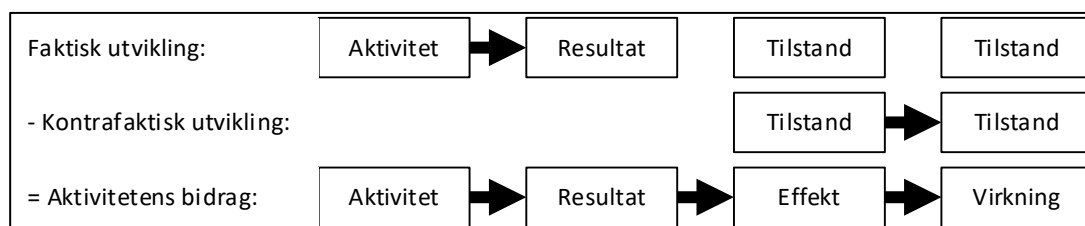
<sup>4</sup> Utfallskjedetenkingen er ikke spesiell for denne evalueringsmodellen, men den finnes i mange varianter. Begrepsbruken varierer mellom ulike evalueringsmodeller, og mange av begrepene brukes ofte om hverandre i en og samme modell. I sammenlikning mellom modeller må en derfor se på hvordan de enkelte ord og uttrykk begrepsfestes eller defineres. Her skjelder vi så konsekvent som mulig mellom resultater, effekter og virkninger som ulike utfall langs en utfallskjede, som i andre sammenhenger også kalles en utfallslinje.



nivå, for eksempel reisetid og sannsynlighet for ulykker, men det trenger ikke være bare det. På lengre sikt, som mer indirekte utfall, medfører disse effektene **virksomheter**, som kan være på brukere av leveransen (trafikkflyt), eller aggregerte brukereffekter (reisemønster), eller andre konsekvenser på samfunnsnivå (geografisk utstrekning av arbeidsmarked), eller for så vidt på planetens klima (klimagassutslipp) eller andre forhold. Med en slik utfallskjede får vi for det første frem en **årsakskjede** fra aktivitet til utfall eller konsekvenser, illustrert ved pilene i figuren, og for det andre en distinksjon mellom **direkte** og **indirekte utfall** eller konsekvenser i flere trinn. Ordvalget for første, andre og høyere ordens konsekvenser er en konvensjon; det finnes flere varianter av ordbruken i slike modeller, men i denne modellen bruker vi dem altså som vist i dette avsnittet.

Det er tre viktige modifikasjoner til en så vidt enkel lineær fremstilling. Begge rører ved årsaks- eller kausalitetsforutsetningene som ligger bak figuren. Det er uansett ikke snakk om noen mekanisk eller deterministisk kausalitet, men årsakssammenhenger som er så vidt sterk og så vidt gjentakende at vi kan tilskrive utfallene til aktivitetene.

Den første modifikasjonen er at de utfallene som skal kunne tilskrives aktiviteten, må være forskjellen mellom hvordan tilstanden på de aktuelle forholdene **faktisk** utvikler seg, og hvordan de **kontrafaktisk** ville ha utviklet seg uten aktiviteten.<sup>5</sup> Dette er illustrert i Figur 2.



**Figur 2: Utfallskjede som forskjell mellom utvikling uten og med aktiviteten**

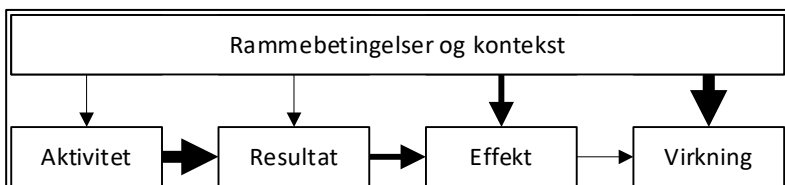
I denne figuren betyr **tilstand** tilstanden på de forhold som er av interesse for å forstå aktivitetens konsekvenser. Effekter og virkninger av aktivitetene og av deres resultater er altså forskjellen mellom det som faktisk skjer og det som ville ha skjedd uten aktivitetene. I bildet av den faktiske utviklingen (øverst i figuren) er det altså ikke gjort noen forutsetning om hvorvidt tilstandene på effekt- og virkningsnivå er forårsaket av aktivitetene og resultatene eller ikke. Hvordan man anslår denne forskjellen, når det kontrafaktiske forløpet per definisjon er ukjent, er et spørsmål om evalueringdesign, som vi kommer tilbake til i kapittel 2.7. Merk imidlertid at når evalueringen består i å sammenlikne utfall av aktiviteten med **mål** for dette, er målene noen ganger formulert som en tilstand og ikke som en endring i tilstand<sup>6</sup>. Da ligger strengt tatt den øverste linjen i figuren til grunn, altså uten de kausale pilene. Eller, som man ofte ser, pilene brukes, som i nederste linje, men da indikerer de bare tidsforløp og ikke egentlig kausalitet. Vi kommer tilbake til dobbeltheten i dette i kapittel 3.5.

Den andre modifikasjonen er at aktivitetens påvirkning må forventes å svekkes jo lenger ut i utfallskjeden vi kommer. Dette har ikke så mye å gjøre med tidsspennet fra årsak til utfall, men med

<sup>5</sup> Den kontrafaktiske utviklingen er det sannsynlige utviklingsforløp uten prosjektet. I Statens prosjektmodell forutsettes det en utvikling i henhold til allerede vedtatte policyer (altså før prosjektet blir vedtatt), men dette kan være et fler-tydig kriterium. Den svarer omtrent til null-alternativet, eller fortrinnsvis null-pluss-alternativet, i en samfunnsøkonomisk analyse av aktivitetens konsekvenser.

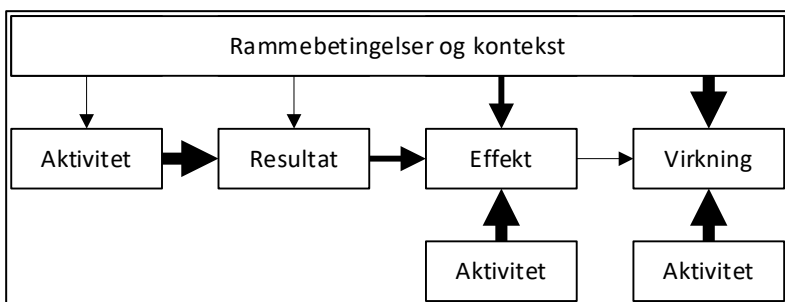
<sup>6</sup> En målformulering av typen "målet er en reisetid på under to timer".

at det er så mange andre forhold som spiller inn. Dette er illustrert i Figur 3, der bredden på påvirkningsspilene avspeiler styrken i påvirkningen fra utgangspunktet.



**Figur 3: Aktivitetens innflytelse svekkes utover i utfallskjeden**

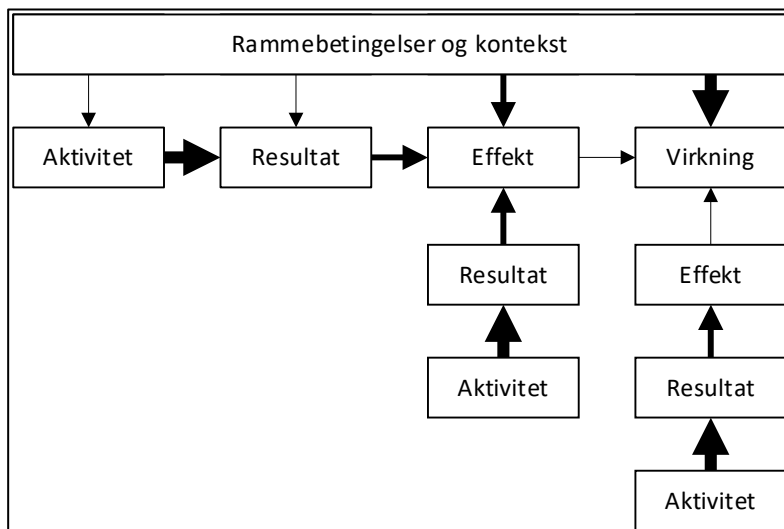
Selv om den som gir regi til eller gjennomfører aktiviteten har betydelig kontroll på resultatet, sviner påvirkningsmuligheten hen jo lenger ut i utfallskjeden man kommer, der andre forhold blir stadig sterkere. Vi har samlet alle disse forholdene under overskriften **rammebetingelser og kontekst**. Disse er heller ikke alltid statiske, de kan endre seg over tid, og prosjektets aktiviteter er blant de forhold som kan utløse slike endringer gjennom tilbakekoblingssløyfer (som ikke er tegnet inn i disse figurene). Men en svekket påvirkningskraft kan man gjøre noe med, se Figur 4.



**Figur 4: Utfallskjede med aktiviteter i flere ledd**

Et effektivt aktivitets- eller prosjektdesign kan altså innbefatte å sette inn støtteaktiviteter som virker direkte på senere ledd i utfallskjeden.<sup>7</sup> Slike støtteaktiviteter kan selvsagt også måtte virke gjennom sine egne utfallskjeder, som vist i Figur 5.

<sup>7</sup> Støtteaktiviteter er et nøkkelord her. Hadde de vært mer konsekvensrike enn den opprinnelige aktiviteten, kunne kanskje sistnevnte ha vært sløyfet.



**Figur 5: Utfallskjede med samvirkende aktiviteter**

Vi ser imidlertid at det som regnes som en effekt i én kjede, kan være en virkning i en annen kjede. Dette understreker at begrepsbruken er et analytisk hjelpemiddel snarere enn en absolutt kategorisering av hvert enkelt forhold som aktiviteten forventes å påvirke.

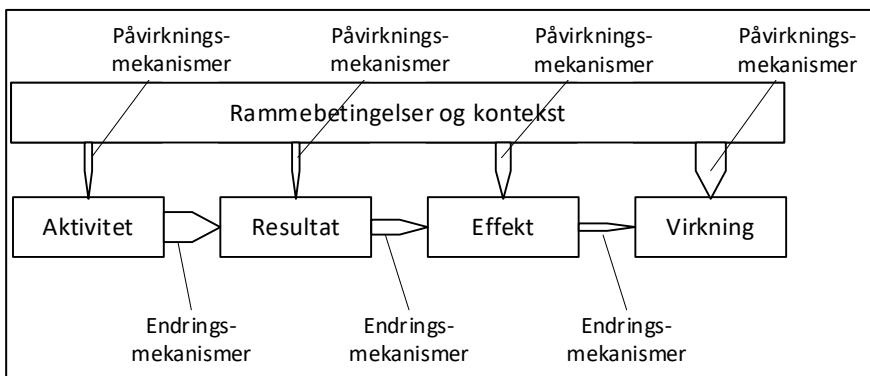
Om vi dessuten tar hensyn til at det kan være mange elementer innen hver av boksene, kan hele utfallskjeden bli ganske kompleks dersom man vil fremstille den i større detalj. I Figur 16 (på side 53 i Vedlegg) viser vi en ganske sammensatt utfallskjede fra et følgeforskningsprosjekt for Nye Veier der vi fulgte utprøvingen av nye arbeidsformer i planleggingsfasen av et vegprosjekt for å vurdere nytten av disse arbeidsformene (Ramstad mfl. 2020).

Det er derfor hensiktsmessig å konsentrere seg om de viktigste sammenhengene, i alle fall i innledende faser.

Men hvordan finner en de viktigste sammenhengene? Det er her den tredje modifikasjonen kommer inn, nemlig en synliggjøring av de mekanismene som er gjemt inne i pilene i de ovenstående figurene. For de kausale sammenhengene mellom utfallene i utfallskjedene ovenfor er ikke gitt bare fordi vi tegner dem som tynne eller tykke piler. Hvis vi ikke kan forklare hvordan et utfall bidrar til neste utfall i utfallskjeden, faller påstanden om at aktiviteten har noen betydning for utfallene, sammen. Denne forklaringen kan være godt eller dårlig begrunnet, og begrunnelsen kan være forankret i forskning eller erfaring eller overbevisning eller håp, men den må være begrunnet for at den skal kunne undersøkes. Dette forholdet underkommuniseres sterkt ved at utfallskjeder, uansett kompleksitet, nesten alltid fremstilles med tekst i boksene (utfallene) og ingen tekst i pilene (mekanismene). Mange slike **endringsmekanismer** er velkjente og kjennes enkelt igjen blant kompetente aktører i det feltet der aktivitetene settes inn for å oppnå endringer. Andre er som sagt dårlig begrunnet og vil være gjenstand for undersøkelser som kan gi betydelig læring gjennom gode evalueringer.

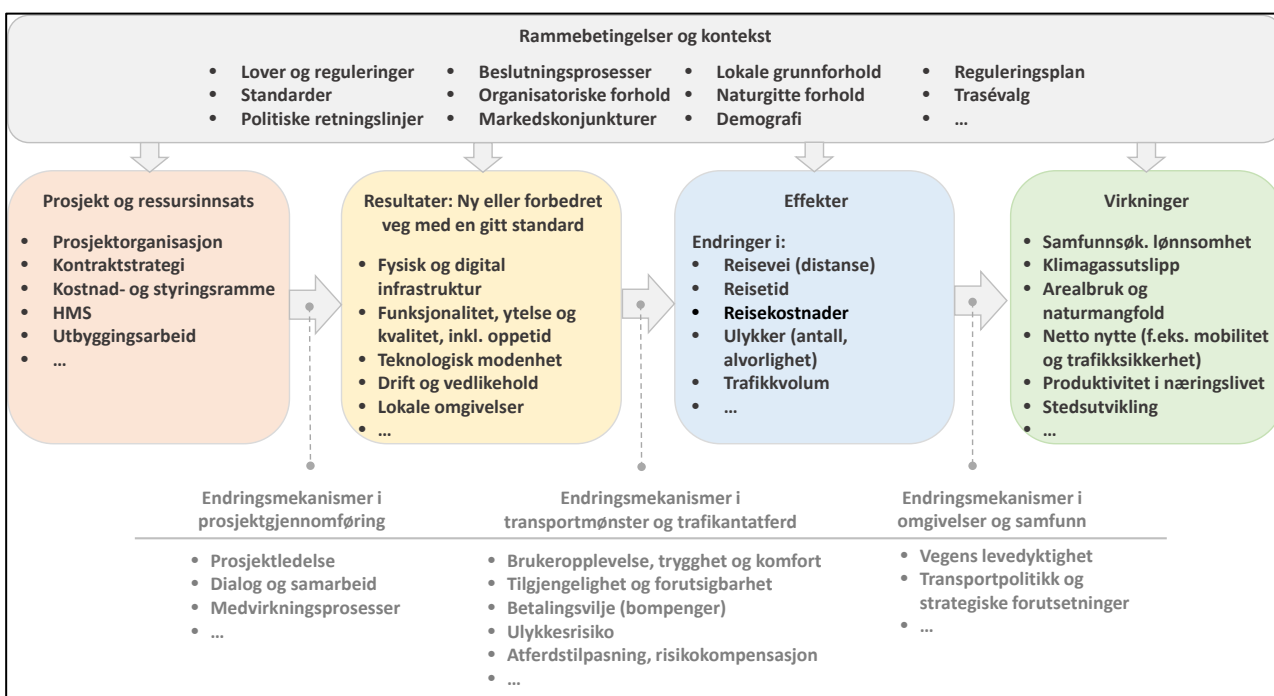
I den enkle varianten vil dermed illustrasjonen se ut som i Figur 6. Endringsmekanismer er her tegnet som enkle navngitte mekanismer; hver av disse kan i seg selv også omfatte kjedereaksjoner med flere ledd, og det kan være interaksjoner mellom dem. Påvirkninger fra omgivelsene har strengt tatt samme forklaringsstatus som påvirkning direkte gjennom utfallskjeden, men vi skjelner

mellom dem fordi vi i en evaluering er spesielt opptatt av om aktiviteten treffer de endringsmekanismene som kan stå seg mot allerede eksisterende påvirkningsmekanismer.



**Figur 6: Utfallskjede med endringsmekanismer og påvirkningsmekanismer**

En redegjørelse for hvordan en planlagt aktivitet forventes å forårsake bestemte resultater, effekter og virkninger under gitte eller varierende kontekstuelle betingelser, inklusive en redegjørelse av de endringsmekanismer som aktiveres, utgjør en **programteori** for aktiviteten. Et eksempel tilpasset vegsektoren er vist i Figur 7.



**Figur 7: Eksempel på programteori for vegbyggingsprosjekter**

Hvordan ulike typer mål knyttes til ulike punkter i en slik kjede, er noe vi vil komme tilbake til i kapittel 2.4 og 2.6. I første omgang er poenget å etablere en terminologi som er så entydig som mulig.

### 2.3 Seks evalueringskriterier

Det viktigste kjennetegnet ved denne evalueringsmodellen, sammenliknet med andre, er at den gjennomgående bruker seks faste evalueringskriterier for en normativ bedømmelse av det enkelte vegprosjektet og dets **vellykkethet**. Disse kriteriene blir sterkt førende for hvilke evalueringsspør-

mål som skal stilles, og dernest for hvilke data som trengs i evalueringsarbeidet. De seks evalueringsskriteriene er kort beskrevet i Tabell 2, sammen med tilsvarende betegnelser i sine forgjengermodeller hos Concept og OECD<sup>8</sup>. Rekkefølgen i dette avsnittet er omtrent fra det viktigste kriteriet og nedover, sett alle prosjekter under ett, men dette er ingen fast rangordning, og som vi skal se, er det fullt mulig å bruke bare ett eller noen få av evalueringsskriteriene for et spesifikt formål, men da blir det en annen evaluering enn om man bruker alle seks.<sup>9</sup>

**Tabell 2: Evalueringsmodellens seks faste evalueringsskriterier**

Tabellforklaring		
Denne modellen	Tilsvarende hos Concept	Tilsvarende hos OECD
Forklaring av hva som evalueres under hvert kriterium, og mot hvilken standard det primært eller typisk vurderes		
Evalueringsskriterier for strategisk vellykkethet		
Relevans	Relevans	Relevance
Tilskittede, oppnådde og fremtidige effekter, vurdert opp mot behovene på ett eller flere tidspunkt, også vurdert opp mot forventet relevans ved beslutningspunkt		
Samfunnsøkonomisk effektivitet	Samfunnsøkonomisk effektivitet	(Ikke i modellen)
Den samlede økonomiske nytten for samfunnet i forhold til de samlede kostnadene prosjektet har medført, vurdert opp mot mål eller forventninger		
Levedyktighet	Levedyktighet	Sustainability
Om, og i så fall hvor lenge, de planlagte effektene kan forventes å vare, vurdert opp mot mål eller forventninger		
Andre virkninger	Andre virkninger	Impact
Andre virkninger av prosjektet enn de planlagte effekter (se kriteriet for taktisk vellykkethet), vurdert opp mot hvorvidt de er positive eller negative for ulike grupper eller samfunnet som helhet		
Evalueringsskriterium for taktisk vellykkethet		
Realisering av effekter	Måloppnåelse	Effectiveness
Førsteordens (og eventuelle direkte høyere ordens) effekter av prosjektet, vurdert i forhold til avtalte effektmål		

<sup>8</sup> Fem av kriteriene er mer eller mindre identiske i alle tre modellene, men forskjellene i navngiving kan bidra til å forstå innholdet bedre. Det sjette kriteriet, samfunnsøkonomisk effektivitet (noen ganger kalt samfunnsøkonomisk lønnsomhet i Concept-modellen), har ingen motpart i OECD-modellen. Dette kriteriet avspeiler tilpasningen til Finansdepartementets rolle i statens prosjektmodell, og det er sentralt i opprettelsen av Nye Veier og i selskapets strategi, kultur og virkemåte. OECD-modellen er nylig (OECD/DAC Network on Development Evaluation 2019) utvidet med ytterligere ett evalueringsskriterium, *coherence*, som viser til prosjektets grad av samvirke (eller motvirke) med andre policyer som skal bidra til overordnede mål på samfunnsnivå. I norsk vegsektor kunne dette for eksempel dreie seg om hvorvidt vegbygging og klimaavtaler mellom staten og storbyene dro i samme retning. Den typen spørsmål om interessebalansering og fragmentering vs integrasjon i offentlig policy har vi heller behandlet som spørsmål om målkonflikter, og ved å åpne for flere interessentperspektiver.

<sup>9</sup> Det er stor fleksibilitet i hvordan evalueringsskriteriene brukes, og hva slags omfang det enkelte evalueringsskillet får. NORAD, som har lang erfaring med OECDs modell, legger modellen til grunn, men ikke nødvendigvis som en struktur for det enkelte evalueringsskillet. Deres evalueringsskillet er stort sett betydelig større enn Concepts evalueringsskillet, som bruker alle seks evalueringsskillet, typisk innenfor en ressursramme på noen få månedsverk (Concept-programmet 2021), og da må man ta enkelte ting for gitt og prioritere noen kriterier høyere enn andre.

## Tabellforklaring

Denne modellen	Tilsvarende hos Concept	Tilsvarende hos OECD
----------------	-------------------------	----------------------

### Evalueringskriterium for operasjonell vellykkethet

Effektivitet i prosjektgjennomføring	Produktivitet	Efficiency
Leveransens kostnad, leveringstid og egenskaper, vurdert i forhold til vedtatt budsjett, tidsplan og forventet kvalitet		

De seks kriteriene representerer i utgangspunktet seks ulike synsvinkler på et prosjekt, seks vinklinger som til sammen gir en robust måte å forstå et prosjekts vellykkethet på. De er neppe uttømmende, men de har vist seg å fungere godt sammen (OECD DAC Working Party on Aid Evaluation 1998; Volden og Samsø 2013). Innen vegsektoren kan de brukes noenlunde standardisert på tvers av prosjekter, men alltid med tanke på evalueringens formål, og de bør aldri anvendes mekanisk uten kontekstforståelse. Det kan selvsagt legges til ytterligere evalueringskriterier i enkelte evalueringer, alt etter hva man er opptatt av å evaluere, men man bør først undersøke om slike kriterier kan falle inn som en av flere dimensjoner under ett av de seks evalueringskriteriene i denne evalueringsmodellens rammeverk.

De fleste av disse evalueringskriteriene er forholdsvis selvforklarende, og detaljer om hvordan de kan anvendes i praksis vil bli behandlet i kapittel 3. Noen kommentarer er allikevel på sin plass her. Vi presenterer dem nå i noenlunde rekkefølge etter når i et prosjektløp det som skal vurderes, finner sted eller blir bestemt. Eksemplene forutsetter at evalueringsprosjektet er en etterevaluering av en ferdigbygd og trafikksett vegstrekning. Andre varianter vil gi andre eksempler. Merk også at sammenhengen mellom evalueringskriteriene og de ulike **mål** som er formulert for et prosjekt, i utgangspunktet ikke er entydig innenfor denne modellen.<sup>10</sup> Vi kommer tilbake til dette forholdet i kapittel 2.6.

- Vurdering av **relevans** er en sammenlikning mellom **behov** og **effekter**, eventuelt også mellom behov og utvalgte **virksomheter**. Dette er formodentlig det viktigste strategiske evalueringskriteriet, men i prosjekter som tar utgangspunkt i tydelig artikulerte behov, er disse ofte så godt tilfredsstillende at kriteriet er lett å overse. Planlagte effekter vurdert mot prosjektutløsende behov kan si noe om hvorvidt den valgte løsningen var relevant da den ble valgt. Faktiske og fremtidig forventede effekter kan relevansvurderes mot behovene slik de fremstår på evalueringstidspunktet eller i fremtiden. Endringer i både behov og tilgjengelige løsninger over tid kan også endre på utfallet av relevansvurderingen.
- Vurdering av **effektivitet i prosjektgjennomføring** handler i første rekke om ressurseffektiv (OECD: *efficient*; Concept: produktiv) produksjon av det som skal leveres, for eksempel en bestemt vegstrekning. Det er gjennomgående oppnåelsen av tre resultatmål med tett innbyrdes sammenheng som vurderes under dette kriteriet: gjennomføringskostnad, gjennomføringstid, og leveransens kvaliteter og omfang. Vurderingen av oppnåelsen av andre mål for, eller krav til, prosjektgjennomføringen, eksempelvis om tilfredsstillende HMS-tilstand eller konfliktnivå med omgivelsene i byggeperioden, inngår også i dette evalueringskriteriet.
- Vurdering av **realisering av effekter** (OECD: *effectiveness*; Concept: måloppnåelse) omfatter i alle tre modellene bare de planlagte effektene av prosjektet. De planlagte effekter,

<sup>10</sup> Begrepene resultatmål, effektmål og samfunns mål forklares i kapittel 2.4.

for eksempel redusert reisetid eller færre rasulykker, er hovedbegrunnelsen for å sette i gang prosjektet. Disse skal svare ut behovene, og det fortrinnsvis på en bedre måte enn andre løsninger. Graden av effektmåloppnåelse måles ved å sammenlikne faktisk realiserte effekter med avtalte effektmål. Relevans, derimot, vurderes ved å sammenlikne faktiske effekter med behov, og i prinsippet kan altså effektrealiseringen og relevansen vurderes ulikt.

- Vurdering av **andre virkninger** omfatter i alle tre modellene både tilsiktede virkninger (OECD: *planned impacts*) på samfunnsnivå og andre, både positive og negative, utilsiktede virkninger på flere observasjonsnivå (OECD: *unplanned impacts*).<sup>11</sup> Tilsiktede virkninger kan omfatte for eksempel økonomisk vekst, mens utilsiktede virkninger kan omfatte for eksempel økt støy (negativt) eller redusert avhengighet av værforhold (positivt). I Concept-modellen og denne modellen blir tilsiktede virkninger på samfunnsnivå, som spesifisert i prosjektets **samfunns mål**, først og fremst håndtert under dette kriteriet, men oppnåelsen av samfunnsmålene kan finnes under alle de strategiske evalueringskriteriene. Det kommer an på hvordan samfunnsmålene er formulert. Utilsiktede virkninger på lavere observasjonsnivå enn samfunnsnivå er også omfattet her, selv om altså kriteriet i første rekke er strategisk. Det er selvsagt et krav at virkningene med troverdighet skal kunne tilskrives prosjektet for å kunne vurderes under dette kriteriet.
- Vurdering av **levedyktighet** dreier seg strengt tatt om de tilsiktede effektene (forventede) varighet over tid, men i praksis blir dette lett en vurdering av **betingelsene** for at leveransen skal kunne fortsette å produsere slike effekter. Eksempler på slike betingelser er at vegens vedlikehold prioriteres og at personbiltrafikk ikke begrenses på strekningen. *Sustainability* i OECD-modellen har altså lite å gjøre med verdens økologiske, økonomiske og sosiale bærekraft. Slike bærekraftsspørsmål behandles i modellen på linje med andre substansielle forhold, og kan inngå i bedømmelsen på alle de seks kriteriene.
- Vurdering av **samfunnsøkonomisk effektivitet**, noen ganger uttrykt som samfunnsøkonomisk lønnsomhet, er spesielt viktig som evalueringskriterium for å kunne husholdere godt med de samlede finansielle ressurser i samfunnet. OECD-modellen omfatter ikke dette kriteriet, formodentlig fordi mange av de aktuelle prosjektene vanskelig kan analyseres noenlunde nøyaktig etter samfunnsøkonomiske metoder. Nye Veier bruker dette kriteriet både i forkant for å prioritere utbyggingsrekkefølge mellom delstrekninger (prosjekter), og i etterkant for å bedømme prosjektenes vellykkethet, og de legger betydelig vekt på dette kriteriet. Denne evalueringsmodellen tillater en prosjektspesifikk vektlegging av dette kriteriet i enkeltprosjekter, akkurat som for de øvrige kriteriene.

Noen vil argumentere med at samfunnsøkonomisk effektivitet inkorporerer alle de andre evalueringskriteriene, særlig hvis man følger praksisen med å inkludere ikke-prissatte virkninger i den samfunnsøkonomiske analysen. Vi benekter ikke at den økonomiske verdien av en virkning (en nytte eller en ulempe) i prinsippet ofte kan vurderes, selv der man ikke har klart å finne en god måte å fastsette en pris på den. Det berører selvsagt økonomien i et prosjekt dersom leveransen ikke lenger er en relevant løsning på behovene slik de har utviklet seg, enten fordi det etter hvert finnes bedre eller billigere alternativer (for eksempel at forstadsbane utkonkurrerer innfartsvei), eller fordi behovene endrer seg (for eksempel at hjemmekontor og digitale møter reduserer behovet for arbeidsreiser). Man kan i slike tilfeller ofte anslå om den økonomiske virkningen er positiv eller negativ, og om den er stor eller liten eller neglisjerbar. Men dette utelukker ikke at slike virkninger også kan ha en ikke-økonomisk egenverdi. Det er slike ikke-økonomiske verdier som de øvrige evalueringskriteriene er ment å ivareta bedre enn å innlemme dem i en (samfunns)økonomisk

<sup>11</sup> Alle **tilsiktede** virkninger på et lavere nivå enn samfunnsnivå skal som en følge av inndelingen i Concept-modellen og i denne modellen håndteres på effektnivå.

vurdering. Dessuten er det slik at **økonomiske fordelingsvirkninger** mellom regioner eller grupper ikke faller inn under samfunnsøkonomiens domene. Økonomiske og andre fordelingsvirkninger fanges opp under kriteriet andre virkninger, som kan være positive eller negative ut fra ulike interessenters perspektiv.

Vurdering av måloppnåelse står selvsagt sentralt i evalueringsmodellen, men det er ingen entydig sammenheng mellom målformuleringer og de seks evalueringskriteriene. Før vi går nærmere inn på disse sammenhengene (i kapittel 2.6), skal vi se nærmere på hvordan målformuleringer og målstrukturer håndteres i modellen.

## 2.4 Mål og målstruktur i en programteoretisk forståelse

Vurdering av faktisk **måloppnåelse** (eller av sannsynligheten for fremtidig måloppnåelse) står sentralt i de fleste evalueringsmodeller. Noen modeller stopper også der. Teknisk sett kunne vi behandle måloppnåelse (på alle plan) som et ekstra evalueringskriterium i modellen vår, med en komponent for hvert av de formulerte målene. Vi har heller valgt å fordele vurdering av måloppnåelse på ulike plan som komponenter i de respektive evalueringskriteriene der de passer best. Vi kommer tilbake til denne fordelingen i kapittel 2.6, og i kapittel 2.7 vil vi se at vurderingen knyttet til det enkelte målet kan formuleres som et evaluerings spørsmål av formen "Er målet nådd?", med eller uten eksplisitt henvisning til prosjektet som årsak til måloppnåelsen.

De målene som skal etableres og styres mot i prosjekter under statens prosjektmodell, skal kategoriseres som samfunns mål, effektmål og resultatmål, og evalueringer skal selvsagt kontrollere oppfyllelsen av disse målene.<sup>12</sup> **Samfunns mål** er overordnet. De skal utformes slik at de reflekterer de behovene i samfunnet som utløser et prosjekt. De skal gi rom for ulike løsningskonsepter, samtidig som de skal kunne være etterprøvbare for det konkrete prosjektet som de utløser. **Effektmål** skal i prinsippet avledes av samfunns mål, men ettersom de først og fremst skal knyttes til tilfredsstillelse av behov på brukernivå, er en slik avledning langt fra entydig. Under utarbeiding av aktuelle effektmål kan en si at de "setter grenser for hvilke konsepter som kan vurderes" (FIN 2010 s. 3). Aktuelle løsninger med sine respektive egenskaper som kan utløse de ønskede effekter, skal deretter skisseres så langt at de kan vurderes opp mot hverandre på hvor godt de svarer til behovene og hvordan de skårer på en kost-/nyttevurdering. Egenskapene ved den valgte løsningen uttrykkes da som **resultatmålet** "kvalitet", i den grad disse egenskapene har betydning for effektrealisering. Og for å kunne målstyre leveranseprosjektet, legges kvalitetsmålet sammen med målene for kostnader og leveringstid som to andre resultatmål.

To viktige begreper vil gå igjen i det følgende: målstruktur og målhierarki. Disse to begrepene brukes ofte upresist, og av og til sammenfallende. I grunnlagsmaterialet for Concept-modellen og i veiledningene til statens prosjektmodell (Klakegg 2006; FIN 2010; Samset 2014) er det heller ikke helt konsistente definisjoner, men intensjonene kan uttrykkes slik:

- En **målstruktur** binder ulike mål sammen med kausale avhengighetsforhold, i motsetning til en målliste, der det ikke trenger være noen avhengighetsforhold mellom målene.
- Et **målhierarki** er en målstruktur der avhengighetsforholdene er hierarkiske, det vil si at mål på lavere nivå må oppfylles for at mål på neste nivå skal kunne oppfylles, helt opp til det overordnede målet på toppnivå. Hierarkiet angir dessuten en prioritering av målene i målstrukturen.

<sup>12</sup> Les mer om disse måltypene i kapittel 2.5.



Evalueringsmodellen tar eksplisitt utgangspunkt i gjentatte erfaringer med å evaluere måloppfyllelsen i prosjekter under statens prosjektmodell og lignende prosjekter. Vi ser at mange slike evalueringsprosjekter, ikke minst i vår egen portefølje, erfarer tre hovedproblemer når det gjelder å vurdere måloppnåelse:

- at mange mål er vagt formulert, slik at vurderingen av graden av oppnåelse av det enkelte målet blir vanskelig, og i alle fall lite presis
- at det blir for mange mål å forholde seg til
- at det er underuttalte målkonflikter mellom mange av målene, og fraværet av en tydelig rangering mellom dem overlater rangeringen til evaluator eller til den enkelte leser av evalueringsrapporten.

Dette er klart problemer for evaluator, men de er først og fremst prosjekteiers problemer, og de har mye større konsekvenser for prosjektene enn for evalueringene. De enkle rådene som ofte gis til prosjektutviklere er

- å forsyne prosjektene med SMARTE<sup>13</sup> formuleringer av de enkelte målene
- å fokusere på et lite antall mål, og
- å etablere et tydelig hierarki mellom målene.

Store, profesjonelle aktører med prosjektplanlegging som kjerneaktivitet har kommet langt i å strømlinjeforme prosjektporteføljene sine i tråd med disse rådene. Dette burde gjøre oppgaven med evaluering av store statlige infrastrukturprosjekter enklere. Allikevel er de langt fra helt i mål. Vår erfaring fra evalueringsprosjekter vi har gjennomført er at dette har fire hovedårsaker:

- Det er så mange legitime interesser og krav som skal ivaretas i et stort prosjekt at det er vanskelig å ikke gi dem målstatus i prosjekter som i utgangspunktet er målstyrte (for da risikerer de å ikke bli tilstrekkelig synlige)
- Målkonflikter blir lettere synlig jo mer SMARTE målformuleringene blir, noe som da går ut over muligheten til omforent aksept, altså Aen i SMARTE, og avveininger mellom ulike mål (og ulike aktørers interesser) gjøres kontinuerlig i alle faser av et prosjekt
- Mange mål er av en slik karakter at deres bidrag til et overordnet mål ikke blir tydelig, de blir hjemløse "delmål" i et hierarki eller en annen struktur, men de skal like fullt nås, noe som også kan bidra til at normen for alle mål blir at de skal oppfylles hver for seg
- Å formulere mål slik at de passer godt inn i en hierarkisk målstruktur er en vanskelig øvelse, og det er spesielt vanskelig å formulere gode samfunns mål.

Dette kan tyde på at det faktisk er gode grunner til å forvente ulike løsninger på målutforming – at ikke alt arbeid med målformuleringer og målstruktur kan passes inn i forutsetningene i ovenfra-og-ned-rammeverket for statens prosjektmodell. Ikke alle mål i målstrukturen vil falle godt inn i hierarkiet, og ikke alle målformuleringer vil være like evaluerbare. Forhold som har egenverdi vil finnes i målbildet, uten at de nødvendigvis henger godt sammen med øvrige mål. En ser også noen ganger at det som ser ut som et målhierarki, ikke har sterke nok avhengighetsforhold i de vertikale relasjonene til å fortjene navnet. Gode evalueringsmodeller vil derfor ta hensyn til dette, og ikke

---

<sup>13</sup> SMARTE: spesifikke, målbare, aksepterte, realistiske, tidfestede, og enkle (i motsetning til sammensatte) formuleringer av enkeltmål.

forutsette fullstendig samsvar med idealmodellen som er godt beskrevet i grunnlaget for statens prosjektmodell (FIN 2010).

Evalueringsmodellen tyr til innsikt, resultater og grep fra evalueringsforskningstradisjoner som kalles teoribasert og realistisk evaluering for å håndtere situasjonen med hybride målstrukturer på en slik måte at man kan kontrollere måloppnåelsen så langt mulig, og samtidig bidra til at organisasjonen blir i stand til å lære av sin egen praksis, både når det gjelder planlegging og bygging av veger, og når det gjelder å formulere mål og bygge målstrukturer som skal bidra både til bedre prosjektarbeid og til nyttigere evalueringer.

**Realistisk evaluering** er kort fortalt en tilnærming som bruker evaluering som en metode til å produsere (ny) kunnskap basert på enkeltstående case i kontekst, samtidig som man holder seg innenfor realistiske rammer (Pawson og Tilley 1997b). Dette innebærer blant annet at en evaluering sjelden bør følge opp alle mulige spor i et prosjekt, og prioritere undersøkelsene og metodevalgene etter behovene for kontroll og læring. Men det innebærer også at læringsaspektet tas på alvor. Og da er det spesielt viktig å forstå hvilke kontekstuelle betingelser som kan forklare evalueringfunnene, altså som forklaringer på hvorfor de samme aktivitetene og resultatene ikke alltid produserer de samme effektene og virkningene (Pawson og Tilley 1997a).

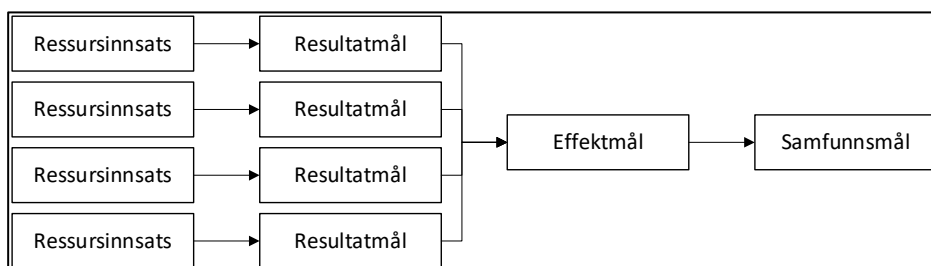
**Teoribasert evaluering** er kort fortalt evaluering som tar fatt i de eksplisitte og implisitte forestillinger, eller teorier, om hvordan aktivitetene i et prosjekt skal gi de resultater, effekter og virkninger som man tilstreber (Chen 1990; van der Knaap 2004). Det er altså ikke en evaluering som er basert på evalueringsteori, men en evaluering som studerer de teorier som måtte finnes blant prosjektaktører om hvordan det de gjør, skal lede til de målene de vil oppnå. I så måte er fokuset på teori et metodisk grep, som et alternativ til klassiske statistiske metoder for effektstudier, som krever data fra ganske mange og svært likeartede case – og kontrollgruppe – for i det hele tatt å kunne gi noe kunnskap. I evalueringssammenheng kalles disse teoriene **programteorier**, fordi arbeidet med dem for det meste har skjedd i kontekst av programevaluering (med mange likeartede caser eller prosjekter), men slike teorier omfatter naturligvis også forestillinger på prosjektnivå. Disse forestillingene trenger altså ikke å være vitenskapelig belagt på noen som helst måte. Tvert imot er relevant erfaring, taus eller tydelig uttrykt, ofte en bedre teori enn den som nyutdannede tar med seg fra forskningsbasert utdanning (Finne 2006; Funnell og Rogers 2011). Men jo sikrere forståelse planleggere og utøvere har av hvordan ulike aktiviteter påvirker resultater, effekter og virkninger i ulike kontekster, desto bedre kan man innrette aktiviteter slik at resultatmål, effektmål og samfunns mål kan både formuleres, knyttes kausalt sammen, og nås. Og det er nettopp dette en teoribasert evaluering gjør, nemlig å synliggjøre gode og dårlige kausale sammenhenger mellom aktiviteter, resultater, effekter og virkninger, i ulike kontekster.

Denne innretningen av teoribasert evaluering har mye til felles med det klassiske "*logical framework approach, LFA*" for planlegging, monitorering og styring av prosjekter, der alle aktiviteter legges inn og relateres til hverandre og til de resultater som skal oppnås og de mål som dermed skal nås (Rosenberg mfl. 1970). Varianter av dette er nå blitt standard tilnærminger til prosjektplanlegging, men utvidelsen til å omfatte effekter på brukernivå og virkninger på samfunnsnivå etter at selve prosjektet (den fokuserte ressursbruken) er avsluttet, krever innsikt fra mange flere fagdisipliner, og ligger som regel i liten grad i prosjektlederens hender, ettersom designet av leveransen (vegen) som regel betyr mer for nytteeffektene enn byggingen av leveransen gjør.<sup>14</sup> Teoridrevet evaluering har arbeidet spesielt mye med dette, og kan dermed bidra både til læring av og

<sup>14</sup> Dårlig arbeid kan selvsagt bety redusert nytte.

i enkeltprosjekter, og til en aggregert kunnskapsbase. Årsakskart mer generelt er også en velkjent tilnærming blant planleggere, i Nye Veier og i vegsektoren for øvrig, som enten vil forklare et avvik, løse et problem, eller prøve ut en ny tilnærming.

Programteorien starter med aktivitetene i prosjektet og viser hvilke utfall disse medfører, som et grunnlag for å evaluere hvilke resultater, effekter og virkninger som kan tilskrives aktivitetene i prosjektet. Evaluators **kartlegging** av programteorien, derimot, starter med prosjektets målhierarki, **fra samfunns mål og nedover**, og legger til ytterligere elementer etter behov. Dette kan synes unødvendig, all den stund det ser ut til å være et godt strukturelt samsvar mellom presentasjonen av et målhierarki (Figur 8) og en utfallskjede (Figur 1).



**Figur 8: Hierarkisk målstruktur**

Det er imidlertid flere grunner til å ikke anta at de er identiske, men starte med samfunns målet og gå bakover i kjeden.

For det første skal målhierarkiet i prosjekter under statens prosjektordning som sagt være utviklet ovenfra og ned, slik at man tar utgangspunkt i det samfunnsmessige behov, og avleder målene videre nedover derfra. Målhierarkiet skal dermed i utgangspunktet være i godt samsvar med en programteori som er bygget nedenfra og opp, men det er altså et empirisk spørsmål hvor godt samsvaret faktisk er. Det finnes som sagt mange eksempler på at pilene i slike diagrammer ikke viser noen kausal sammenheng (de avspeiler en fremstillingskonvensjon), og det kan være løsthengende mål i strukturen.

For det andre gir målhierarkiet en god indikasjon på hvilke resultater, effekter og virkninger som det er spesielt viktig å fokusere på i konstruksjonen av en programteori, i alle fall som et utgangspunkt.

Og for det tredje er det slik at interessen for at et prosjekt eller et tiltak bør evalueres, ofte tar utgangspunkt i utfall som ikke blir som planlagt, forventet eller ønsket, og at den læring som forventes i en slik evaluering, ikke har formen "hva var **effektene** av å gjøre a, b og c", men "hva var **årsakene** (d, e eller f) til at utfallet ble som det ble.

En vitenskapelig basert metode for evaluering tar utgangspunkt i å starte med en aktivitet og se hvilke konsekvenser den har (se kapittel 2.7). Når et evalueringsspørsmål starter med en konsekvens og vil undersøke hvilke årsaker den har, må man enten konvertere spørsmålene eller bruke andre metoder. Den konverteringsmekanisme som man gjerne bruker, er å omformulere spørsmål om årsaker til et sett med hypoteser om effekter og virkninger; men da må man først gjennom et detektivliknende arbeid med å finne gode kandidater til årsaksrollen. Dette detektivliknende arbeidet er godt hjulpet av en programteori som er formulert av tiltakshaverne selv, gjennom målhierarkiet. Samtidig er et slikt detektivarbeid et godt bidrag til nettopp å synliggjøre deler av

programteorien, særlig hvis det kan gjøres i samarbeid mellom evaluator og programaktører. Akkurat som en aktivitet kan påvirke mange forhold, kan et utfall ha mange årsaker. Evalueringsspørsmål som starter med å lete etter årsaker i et konkret tilfelle, kan dra nytte av ulike bakoverskuende metoder. To av disse metodene er *root cause analysis* (se for eksempel Andersen og Fagerhaug 2006) og *cause and effect analysis* med de velkjente fiskebeindiagrammene (Ishikawa 1990), som trenger bak symptomer og leter etter tiltak for å forhindre gjentak, snarere enn avbøtende tiltak i det enkelte tilfellet. Slike metoder kan brukes til å identifisere viktige potensielle årsaker, som sammen med det interessante utfallet kan bli kjernen i et evalueringsspørsmål som kan testes empirisk. Mange slike metoder brukes i kvalitetssikringsarbeid. De er også utbredt i vegsektoren, slik at de kan brukes i samarbeid med informanter for å videreutvikle deler av programteorien som så legges til grunn for nærmere undersøkelser. Dette kan også styrke en omforent forståelse av programteorien.

## 2.5 Ulike interessenters rolle i dette rammeverket.

Utgangspunktet for den enkelte evaluering er helt klart prosjekteiers perspektiv, på vegne av Stortinget som besluttsende organ for de store investeringene det er snakk om. Det er imidlertid flere legitime interessenter rundt de vegprosjektene som Nye Veier (og andre) arbeider med; og det er selvsagt ofte overlapp mellom dem.

- **Folkeviljen**, som representert gjennom demokratisk gjennomførte nasjonale politiske prioriteringsvedtak i Stortinget, og som implementeres av Samferdselsdepartementet gjennom **oppdrag** til Nye Veier, ofte kalt samfunnsoppdraget, er helt klart det førende perspektiv.
- Demokratiet finnes også uavhengig av dette på **lokalt** politisk plan, noe som tilsier at deler av samfunnsoppdraget i bredere forstand kommer fra de kommuner der vegen skal gå. Kommunestyrenes lokale perspektiv og prioriteringer er derfor en faktor i det enkelte vegprosjektet.
- Planleggingsmetodene forutsetter i noen grad at **brukere** er en homogen gruppe, med behov, perspektiver, interesser og prioriteringer som tilgodeses gjennom utforming av prosjektene. Det er imidlertid alltid stor variasjon i brukergrupper, noe som tilsier at også ulike brukerperspektiver bør ivaretas i evalueringene.
- Også **ikke-brukere** kan være berørt av et vegprosjekt, og kan derfor også tilføre et viktig perspektiv.
- **Interessentgrupper** som arbeider for andre saker i samfunnet enn de som tilgodeses av vegutbygging, kan ha legitime perspektiver på konsekvenser av vegbyggingen eller på alternativ bruk av offentlige ressurser som brukes til vegbygging.
- **Institusjoner** med fagmiljøer som forvalter andre offentlige interesser, eksempelvis hos statsforvalteren, bidrar i beslutningsprosessene; deres perspektiver kan også være relevante i evalueringer.
- **Nye Veier** som organisasjon vil naturligvis ha et perspektiv på sin egen utvikling som organisasjon, selv om organisasjonen i prinsippet "bare" utfører oppdragene fra samfunnet.
- **Fagpersoner** som arbeider med planlegging og gjennomføring av vegprosjekter, utgjør en formidabel ressursbase med tanke på hvordan prosjektene kan bli bedre. Deres perspektiver, erfaringer og idéer kan komme til sin rett gjennom kontinuerlig forbedring og innovasjon i deres respektive organisasjoner og prosjekter, og også synliggjøres gjennom evalueringer.

Ideelt sett bør alle disse og eventuelle andre aktører også kunne bidra med kandidater til evalueringskriterier eller evalueringsspørsmål, med data, og med fortolkning av evalueringresultater. Akkurat som en evaluering bringer inn et alternativt perspektiv på mål og aktiviteter i et prosjekt, gjør

andre interessenter også det samme. Aktuelle interessenter kan få sin stemme i mange former, blant annet gjennom deltakelse som intervjuobjekt, referansegruppedeltaker, og forfatter av skriftlige kommentarer og innsigelser i prosjektløpet. Dette avgjøres i det enkelte tilfellet.

## 2.6 Forholdet mellom mål og evalueringskriterier

Det er som antydning ikke noen en-til-en kobling mellom de seks evalueringskriteriene og målene på de tre nivåene i målhierarkiet. Sett fra målsiden er koblingene som følger:

- **Resultatmålene** er de som de tilsiktede faktiske<sup>15</sup> resultatene av prosjektgjennomføringen skal vurderes opp mot, for å avgjøre prosjektets operative vellykkethet. Merk at prosjekter også kan ha utilsiktede resultater. De relevante faktiske resultatene kan også vurderes opp mot faktiske resultater i sammenliknbare prosjekter, i en enkel form for referansemåling (*benchmarking*). Resultatmålene på sin side kan også noen ganger med fordel vurderes opp mot faktiske resultater i sammenliknbare prosjekter, for å vurdere om de er satt for høyt eller for lavt.
- **Effektmålene** er de som de tilsiktede faktiske effektene av å ta vegen med dens egenskaper i bruk skal vurderes opp mot, for å avgjøre prosjektets taktiske vellykkethet. Merk at prosjekter også kan ha utilsiktede effekter. De relevante faktiske effektene kan også vurderes opp mot faktiske effekter i sammenliknbare prosjekter, i en enkel form for referansemåling (*benchmarking*). Effektmålene på sin side kan også noen ganger med fordel vurderes opp mot faktiske effekter i sammenliknbare prosjekter, for å vurdere om de er satt for høyt eller for lavt.
- **Samfunnsmålene** er de som et utvalg av de faktiske virkningene på samfunnsnivå av å bygge og ta vegen i bruk skal vurderes opp mot, for å avgjøre en del av prosjektets strategiske vellykkethet. Merk at virkningene kan omfatte mye mer enn de som dekkes av samfunnsmålene. Det gir mindre mening å benchmarke prosjekter mot hverandre på tilsiktede virkninger på samfunnsnivå, all den stund samfunnsmålene ofte er vanskelig å formulere, enn si kvantifisere. Derimot kan faktiske virkninger, om de lar seg anslå eller måle, noen ganger være til hjelp med å formulere mer presise samfunns mål i senere, sammenliknbare vegprosjekter.

Sammenhengen mellom samfunnsmålene og de strategiske evalueringskriteriene er altså ikke helt entydig. Dette handler blant annet om at det er mindre konsensus om hvilke typer mål som kan regnes som samfunns mål. En kunne for eksempel tenke seg at det skulle formuleres mål for hvert av de fire strategiske evalueringskriteriene, men selv om det er gode erfaringer med å vurdere store prosjekter etter disse kriteriene, er det ikke sikkert at det ville være hensiktsmessig som et grunnlag for å formulere gode samfunns mål for et prosjekt eller for prosjekter innenfor en bestemt sektor. Dessuten er det viktig i seg selv at et prosjekt kan vurderes på andre dimensjoner enn de som er planlagt og innlemmet i målformuleringer på forhånd. Dette vil motvirke en sakte innlåsing mellom mål og evalueringskriterier, og det vil åpne for flere vinklinger når verden forandrer seg og interessene likeså. Derfor er det spesielt viktig at kriteriet Andre virkninger ikke blir for snevert avgrenset, selv om altså flere dimensjoner her etter hvert vil bli standardisert når de får økende og vedvarende betydning.

Som en begynnende forenkling kan man se om det finnes formuleringer av sektorspesifikke samfunns mål som gir retning for offentlig finansierte prosjekter innenfor sektoren. I vegsektoren er

---

<sup>15</sup> **Faktiske** resultater, effekter og virkninger forutsetter at evalueringen gjøres etter at de har funnet sted. For evalueringer på et tidligere tidspunkt er det eventuelt snakk om **forventede** resultater, effekter og virkninger.

målene for Nasjonal Transportplan (NTP) (Meld. St. 33 (2016-2017) ; Meld. St. 20 (2020-2021)) gode kandidater for å diskutere forholdet mellom samfunns mål og evalueringskriterier. Disse kan endres fra en planperiode til neste; de enkelte evalueringsprosjektene må ta høyde for det. NTP bruker ikke målterminologien fra prosjektverden, men det som kalles "det langsiktige målet for transportpolitikken" fram til 2050 og de "overordnede målene" (langs tre dimensjoner) for den enkelte planperioden er mål som de enkelte vegprosjekter må innrette sine samfunns mål mot om de skal kunne bli realisert med statlig finansiering.

**Tabell 3: Målområder for Nasjonal Transportplan 2018-2029 og 2022-2033**

Målområder	2018-2029	2022-2033
<b>Langsiktig mål</b>		
Langsiktig mål (2050)	Et transportsystem som er sikkert, fremmer verdiskaping og bidrar til omstilling til lavutslippssamfunnet	Et effektivt, miljøvennlig og trygt transportsystem i 2050
<b>Overordnede mål (dimensjoner)</b>		
Sikkerhet	Redusere transportulykkene i tråd med nullvisjonen	Nullvisjon for drepte og hardt skadde
Økonomiske virkninger	Bedre framkommelighet for personer og gods i hele landet	Enklere reisehverdag og økt konkurransevne for næringslivet
Klima og miljø	Redusere klimagassutslippene i tråd med en omstilling mot et lavutslippssamfunn og redusere andre negative miljøkonsekvenser	Bidra til oppfyllelse av Norges klima- og miljømål
<b>Tilleggs mål for prioritering av ressursbruk</b>		
Teknologibruk		Effektiv bruk av ny teknologi
Kostnadseffektivitet		Mer for pengene

Transportpolitikken omfatter selvsagt mye mer enn bygging av veger og annen transportinfrastruktur. Vegbygging kan allikevel bidra til alle fem (tre) hovedmålområdene og tydeliggjøre dette gjennom sine respektive formuleringer av samfunns mål for det enkelte vegprosjektet. Formålet med å bygge en veg, eller det utløsende behovet på samfunnsnivå, kan i prinsippet forankres i ett eller flere av disse. Men gitt at en veg kan løse dette behovet, vil det være vanskelig å ikke forholde seg til forbedring på alle de andre hovedmålområdene, selv om disse i seg selv ikke ville være tilstrekkelig for å utløse vegprosjektet. Det skal for eksempel mye til for at et stort vegprosjekt blir satt i gang med hovedbegrunnelse å bidra til Norges klimamål, eller fordi det gjør effektiv bruk av ny teknologi. Men begge disse målene må ivaretas som føringer på vegprosjektet når det utløses for å styrke framkommelighet eller eliminere rasfare. I evalueringen av prosjektet kan det dermed være aktuelt å skjelve mellom mål som er utløsende for prosjektet, som dekkes av relevanskriteriet, og mål som ligger som føringer fra transportpolitikken når veger først skal bygges.<sup>16</sup>

Merk også at alle målene i Tabell 3 kan brytes ned til prosjektspesifikke samfunns mål og videre til effektmål og understøttes av resultatmål, avhengig av hvordan vegprosjektene utformes.<sup>17</sup> Det finnes også mange spor av en slik målnedbryting i den mer detaljerte teksten i NTP. Denne evalueringsmodellens operative og taktiske evalueringskriterier er med andre ord på sin plass for evaluering av prosjektene, sammen med det samfunnsøkonomiske kriteriet (som er sterkere fremhevet i

<sup>16</sup> Overordnede politiske føringer kan også formuleres som krav til løsningen, snarere enn mål som skal nås. De får uansett liknende status i evalueringsmodellen.

<sup>17</sup> Statens prosjektmodell forutsetter at målene skal utvikles på denne måten, ovenfra og ned.

planens siste utgave, men altså ikke formulert som et mål). Og, som forventet, fanger de overordnede målområdene mange av de tilsiktede virkninger som det er aktuelt å undersøke under evalueringskriteriet Andre virkninger. Ser man etter mål og prioriteringer i NTP som svarer til kriteriene Levedyktighet og Relevans, er disse ikke tematisert i like høy grad, men de kan sies å ligge under, særlig relevanskriteriet.

## 2.7 Design og metodetilfang

Evalueringskriterier er ofte uttrykt i en generisk form. Dette gjelder alle de seks kriteriene i evalueringsmodellen. Det er nødvendig å operasjonalisere dem. Hvert kriterium kan konkretiseres som ett eller flere evalueringsspørsmål<sup>18</sup>, som hver for seg eller samlet representerer kriteriet, og hvert av disse evalueringsspørsmålene kan i neste omgang operasjonaliseres som en eller flere indikatorer, som igjen kan belyses med data på en eller flere variabler. Dataene kan være objektive eller subjektive, de kan måles kvantitativt eller kvalitativt, og kvaliteten og tilgjengeligheten kan variere. Målformuleringer som gjelder for prosjektet, behandler vi evalueringsteknisk som evalueringsspørsmål av formen "Er målet nådd?" under de respektive evalueringskriteriene.

Triangulering mellom metoder og mellom datakilder på alle disse nivåene kan være aktuelt for å få et bedre samlet bilde. Tabell 7 (på side 37) viser en rekke eksempler på hvordan evalueringskriterier kan brytes ned til aktuelle evalueringsspørsmål (inklusive målformuleringer) med relevante indikatorer og datasett som enten eksisterer eller må genereres. Dette er langt fra noen komplett liste.

Evalueringsspørsmål av formen "er målet nådd?" besvares ofte med et veldig enkelt undersøkelsesdesign. Man velger indikator, finner data, og måler tilstanden etter at prosjektet (eller den prosessen som skal evalueres) er gjennomført. Tilsvarende, hvis tilstanden ikke kan måles ennå, oppgir man forventet tilstand. Faktisk eller forventet måloppnåelse måles da enkelt ved å sammenlikne tilstanden med målet slik det er formulert og operasjonalisert. Det er som regel mer arbeid dersom målet er formulert som en endring, for da må tilstanden måles både før og etter, noe som ofte strander på grunn av mangel på før-data. Men i begge tilfeller velger dette enkle designet bort muligheten til å tilskrive måloppnåelsen til det som er gjort i prosjektet. Strengt tatt svarer ikke dette enkle designet på spørsmålet om prosjektet har nådd målet, men om tilstanden etter prosjektet er i samsvar med målet.<sup>19</sup>

Dette er i utgangspunktet et tilfredsstillende design, ettersom det ikke evaluerer prosjektet, bare noen tilstander som prosjektet hadde som mål å påvirke. Et tilfredsstillende evalueringsdesign vil derfor måtte undersøke hvilke utfall (resultater, effekter og virkninger) aktivitetene i prosjektet har, sammenliknet med om prosjektet ikke hadde vært satt i gang, og først deretter vurdere disse opp mot prosjektets formulerte mål og/eller andre standarder. Slike design gir også grunnlag for å borre dypere ned i materialet for å se hvilke aspekter ved prosjektet og konteksten rundt som har bidratt til hvilke resultater, effekter og virkninger, og på hvilken måte. Programteorien og foreløpige vurderinger er godt egnet til å identifisere de sammenhenger som skal undersøkes. Undersøkelsesdesign vil variere mellom sammenhenger, ut fra nedenstående gjennomgang. Vi understreker at det

<sup>18</sup> Ikke alle evalueringsspørsmål trenger strengt tatt være en operasjonalisering av et overordnet evalueringskriterium. Det kan godt stå alene, men det må fortsatt operasjonaliseres.

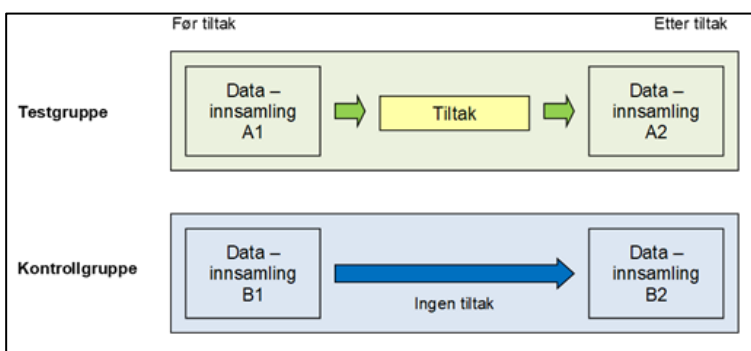
<sup>19</sup> Noen **resultater** kan som regel tilskrives prosjektet uten nærmere undersøkelser. Eksempelvis er selve vegen et resultat i seg selv. Dersom tilstanden var "gammel veg" eller "ingen veg" før prosjektet og "ny veg" etter prosjektet, og andre har bygd vegen mens prosjektet har brukt pengene sine på noe annet, er addisjonaliteten av prosjektmidlene liten, og man befinner seg sannsynligvis i et meget korrupt regime. Vi vil også vise i kapittel 2.6 at i noen tilfeller kan **effekter** tilskrives bestemte årsaker dersom det er en godt etablert sammenheng mellom årsak og effekt, og empirien støtter dette.

sjelden eller aldri er aktuelt å undersøke alle sammenhenger i et enkelt prosjekt; evalueringsressursene bør selvsagt settes inn der de gjør mest nytte for seg.

Bedre evalueringdesign beskriver evalueringens vitenskapelige fremgangsmåte og må defineres med utgangspunkt i de metodene som er best egnet til å besvare de relevante evalueringsspørsmålene. Evalueringen kan i utgangspunktet kombinere flere metoder og analyser (både kvalitative og kvantitative), og det forutsettes at metodeoppsettet utformes på bakgrunn av kunnskap om de ulike metodenes styrker og svakheter. I praksis betyr dette at de ulike variablene og sammenhengene i programteorien kan testes med ulike metodeoppsett, avhengig av prosjektets modenhet (tidspunktet for evaluering) og datatilgang.

**Kvasi-eksperimentelle metodeoppsett** innebærer at forsøkspersoner eller strekninger fordeles i en testgruppe og kontrollgruppe som sammenlignes før og etter gjennomføringen av prosjektet. Bruk av kontrollgruppe (forsøkspersoner eller strekninger som ikke har vært utsatt for tiltak) skal sikre at eventuelle observerte effekter kan knyttes til prosjektet og ikke er et resultat av en generell samfunnsutvikling. Dette metodiske oppsettet benyttes ofte for evaluering av veg- og trafikktiltak, f.eks. for å avdekke effekter på reisetid, ulykker, infrastrukturens oppetid, antall nedstenginger eller driftskostnader.

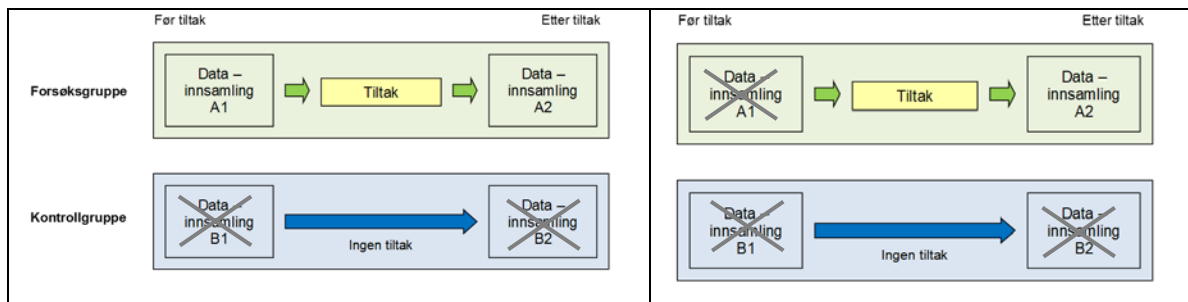
Dersom fordelingen av forsøkspersoner eller strekninger i testgruppe og kontrollgruppe skjer tilfeldig (randomisert) kalles dette for et **eksperimentelt evalueringdesign**. Dette anses å være selve "gullstandarden" for metodeoppsett, men er sjelden mulig å oppnå i praksis. I transportsektoren benyttes slike eksperimenter først og fremst i kjøresimulator eller intervjustudier, da det er vanskelig å oppfylle disse betingelsene i evalueringer som baserer seg på observasjoner i den virkelige verden.



**Figur 9: Det (kvasi-) eksperimentelle metodeoppsettet med datainnsamling før og etter gjennomføring av prosjektet, og sammenligning av resultater mellom testgruppe og kontrollgruppe**

**Ikke-eksperimentelle metodeoppsett** innebærer at prosjektet evalueres i en testgruppe, uten sammenligning mot en kontrollgruppe. Dette metodeoppsettet benyttes blant annet når det ikke er mulig å etablere en passende kontrollgruppe, enten på grunn av begrensede ressurser til datainnsamling eller fordi prosjektet berører hele befolkningen. Et ikke-eksperimentelt evalueringdesign forutsetter at det er etablert en velfundert programteori for å sikre evalueringens validitet. Når evalueringen struktureres rundt prosjektets programteori, omtales dette også som et **teoribasert evalueringdesign**.





**Figur 10: Eksempler på ikke-eksperimentelle metodeoppsett (observasjonsstudier med data-innsamling kun i forsøksgruppen).**

Prosjektets effekter og årsakssammenhenger vil kunne testes empirisk med ulike metoder og på ulike tidspunkt i gjennomføringsprosessen. Eksempelvis vil måling av faktiske effekter og virkninger kreve at vegen har fungert stabilt over en lengre driftsperiode. Tabell 4 gir en oversikt over relevante betingelser ved valg av evalueringsdesign og metoder.

**Tabell 4: Valg av evalueringsdesign og metode**

Evalueringsdesign	(Kvasi-) eksperimentelt	Ikke-eksperimentelt (teoribasert)	
<b>Metode</b>	Kontrafaktisk analyse med statistiske metoder	Analyse av programteori (kvalitativ og kvantitativ metode)	Analyse av endringsmekanismer (realistisk metode)
<b>Formål</b>	Bevise effekter	Finne effekter	Forklare effekter
<b>Eksempel på evalueringsspørsmål</b>	<i>Har man oppnådd forventet effekt i løpet av prosjektperioden?</i>	<i>Før innføring: Hvordan kan prosjektet bidra til å utløse effekter? Etter innføring: Har man oppnådd effekter og var det prosjektet som førte til disse?</i>	<i>Før innføring: Hva virker, for hvem, i hvilken kontekst og på hvilken måte? Etter innføring: Hvilke endringer har skjedd og hvordan har de oppstått?</i>
<b>Tidspunkt for evaluering</b>	Etter implementering	Gjennom hele prosjektet: Før, underveis eller etter innføring	
<b>Egenskaper ved tiltaket og omgivelser</b>	Forventer å gi målbare resultater i løpet av prosjektperioden Stabilt og uforandret gjennom prosjektperioden Entydig årsak-virkning, og kan skilles fra andre tiltak Empiri etablert, robuste hypoteser og få andre forklaringer på endring	Langsiktig perspektiv på realisering av effekter Forventer små, kvalitative eller indirekte effekter Prosjektet/tiltaket er teknologisk umodent og/eller forventes å utvikle seg over tid Prosjektet innebærer komplekse årsakssammenhenger og/eller situasjonsavhengige effekter Dynamiske omgivelser og/eller variasjoner i implementeringen av tiltaket	
<b>Metodiske begrensninger og muligheter</b>	Representativt utvalg og kontrollgruppe Må være mulig å etablere baseline (nullalternativ) for kontrafaktiske analyser	Utforsker samspill mellom ramme-betingelser, eksterne faktorer og utfall Særlig egnet for svært komplekse, sammensatte tiltak (makronivå)	Utforsker underliggende mekanismer som utløser tiltakets effekter Særlig egnet for tiltak med begrenset fokus og influensområde (mikronivå)

Tilpasset etter Lervåg (2021) og Bredgaard mfl. (2016)

Tabell 4 er ikke uttømmende. En del sammenhenger, også slike som kan være av interesse, kan det være vanskelig å få data om som tilfredsstillende kravene under noen av de tre designene i tabellen. Den beste muligheten er da å bruke de tilgjengelige dataene, som for eksempel kan stamme fra ett eller to intervju, til å formulere en hypotese om en sammenheng, og presentere dette dersom det gir noen tilleggsverdi i evalueringen.

I prinsippet kan virkninger av en aktivitet undersøkes med det samme designrepertoar som effekter. Ettersom utfallskjeden frem til virkninger kan være lang, må designet ta høyde for flere kontekstuelle påvirkninger, og funnene må tolkes med større forbehold. Fremtredende evalueringsforskere anbefaler at man avgrenser slike effektstudier til det første utfallet av interesse i utfallskjeden (Mohr 1995).

I praksis er det en gjensidig avhengighet mellom tilgjengelig datatilfang og metodisk design. Realistisk evaluering strekker seg gjerne mot å bli vitenskapelige undersøkelser, men forholder seg til hva som er praktisk gjennomførbart. Ut fra et klassisk vitenskapssyn om hva som teller for en evidensorientert evaluering, kan man se for seg at aktuelle evalueringsmetoder rangeres som følger (Johansson mfl. 2015):

1. Randomiserte kontrollerte forsøk (RCT)
2. Kvasi-eksperimentelle studier
3. Sammenligning før- og etter tiltak i en gruppe
4. Tverrsnittstudier med randomiserte utvalg
5. Prosessevaluering, formative studier og aksjonsforskning
6. Kvalitative case-studier og etnografisk forskning
7. Deskriptive retningslinjer og god praksis
8. Ekspertvurderinger
9. Brukervurderinger

Slike rangeringer forutsetter at det bare finnes én kvalitetsnorm for evidens. Styrken med teoribasert evaluering er at det nå i tillegg er etablert en vitenskapsbasert kvalitetsnorm for evidens der tilgjengelige data ikke tilfredsstillende de klassiske kravene til en "blind" input-output-analyse, men der en forståelse av mekanismene i input-output-transformasjonen, som vi har omtalt som endringsmekanismen, står i sentrum.

Alle de nevnte designene (med unntak av det "enkle") forutsetter at man starter med en aktivitet og undersøker dens betydning for ett eller flere utfall. Som regel er det ikke bare ett resultat, en effekt og en virkning av en aktivitet; det kan være mange i parallell på samme nivå. Det er potensielt en stor vifte av mulige sammenhenger å undersøke. Som tidligere nevnt starter imidlertid ofte en prosjektevaluering med at man vil finne årsakene til at noe ikke gikk som det burde. Da peker viften av mulige årsaker å finne motsatt vei. Da starter man med et punkt i utfallsviften og går tilbake til mange deler av et tiltak (eller til mange tidligere utfall) i søket etter årsaker. Den viften starter altså langt ute i utfallskjeden og åpner seg innover. En detaljert programteori kan da være et hjelpemiddel til å finne hvilke årsakssammenhenger som det kan være verdt å bruke ressurser på å undersøke med et design etter valgkriteriene i Tabell 4, og hvilke årsakssammenhenger som kan fungere som støtteforklaringer basert på data som er for spinkle til å danne grunnlaget for en troverdig effektanalyse.

Data som Nye Veier samler i sitt strategihus ut fra behovene for ytelsesindikatorer i organisasjonen (*key performance indicators*, KPIer), kan gjerne inngå i datagrunnlaget for evalueringer, forutsatt at de er registrert slik at de kan knyttes til det aktuelle evalueringssubjektet. Noen data som er nyttige i evaluering, samles mest hensiktsmessig av Nye Veier selv. Dette gjelder særlig data som genereres i løpet av planlegging og gjennomføring av de enkelte prosjektene. Det er imidlertid et velkjent problem (for evaluator) at slike data ikke er tilgjengelig på evalueringstidspunktet, fordi de ikke er lagret på noen systematisk måte. En organisasjon som satser på å lære av sine evalueringer, må også sikre at prosessdata som genereres gjennom et prosjektforløp, også lagres for fremtidig bruk. Når det gjelder bruk av evalueringmodellen til kontrollert utprøving av ett eller annet tiltak, vil den evaluator som setter opp utprøvingen slik at de aktuelle effekter kan etterprøves, også kunne spesifisere data som må samles før, under og etter utprøvingen, og avgjøre hvilke av disse som mest hensiktsmessig bør samles av Nye Veier.

## 2.8 Læring

Nye Veier har lagt betydelig vekt på at evalueringmodellen eksplisitt skal brukes til evalueringer som de selv og aktører i vegsektoren generelt kan lære av, se Tabell 1 (på side 5). Læringspunktene fra evalueringene forventes å bidra til praksis som understøtter selskapets verdier:

- fornye (løsninger og arbeidsformer gjennom innovasjon)
- forbedre (eksisterende arbeidsprosesser og løsninger)
- forsikre (alle interessenter om at de ivaretar deres interesser)

Læring er et mangefasettert begrep. Det omfatter både individuell og organisatorisk læring, som er individers, respektive organisasjoners, tilegnelse av kunnskap og omsetting av denne til endret praksis. Når man lærer av det man har gjort som person eller organisasjon, snakker vi gjerne om enkeltsløyfelæring (eller førsteordens læringssløyfe) og dobbeltsløyfelæring (eller andreordens læringssløyfe) (Argyris og Schön 1978). En læringssløyfe som vurderer alternativ praksis innenfor de gitte rammer for hvordan man gjør ting i organisasjonen, en enkeltsløyfe, forekommer ofte i forbedringsarbeid. En læringssløyfe som inntar et annet perspektiv, en dobbeltsløyfe, kan gjerne påbegynnes, men den fullføres langt sjeldnere som en fornyelse eller innovasjon. Dette skyldes at den kan forutsette endringer ikke bare i den aktuelle praksisen, men også i andre forhold som henger sammen med vedkommende praksis, noe som gjør implementeringen betydelig tyngre.

Vi skal heller ikke glemme det vi kan kalle en nullteordens læringssløyfe. Vi lar dette begrepet betegne en situasjon der man bare tar vurderingen av det man har gjort til informasjon eller etterretning, uten at det har noen konsekvenser for fremtidig praksis. Vi vil også peke på en læringssløyfe som ligger mellom den nullteordens sløyfen og enkeltsløyfen. Denne beskriver en situasjon der en evaluering avdekker en feil, eller et avvik fra en norm, regel eller fastlagt rutine. Mange evalueringer med et kontrollformål resulterer i at man retter opp sin feil ved å gjøre slik det skal gjøres neste gang. Da har man ikke lært **hva** man burde gjøre, for det visste man fra før. Man har bare lært **at** man faktisk må gjøre det. Et læringsbegrep som forutsetter at det finnes en korrekt måte å gjøre ting på, åpner ikke for forsøk med å gjøre ting bedre, men det er vår erfaring at det ofte er dette begrepet som ligger under formuleringen "å lære av sine feil" i respons til et dårlig evalueringresultat.

Concept har undersøkt hvordan departementer og etater bruker evalueringer generelt og Concept-evalueringene spesielt. Dette har munnet ut i en rekke anbefalinger både til evaluator og til de

ansvarlige for de prosjekter som evalueres, anbefalinger som skal styrke nytten av evalueringer, ikke minst læringen fra dem.

Anbefalingene til evaluator er som følger (Bukkestein mfl. 2020):

- La målgruppen komme med innspill til hva som er spesielt interessant å få belyst i evalueringen
- Mer og bedre involvering, spesielt i utforming av læringspunktene
- Skille mellom målgruppene
- Enda mer fokus på formidling er nødvendig

Anbefalingene til etater og departementer er som følger:

- Ansvar for evaluering bør plasseres hos den som er ansvarlig for måloppnåelsen
- Tidlig tilrettelegging for evaluering, vil øke kvaliteten på evalueringene
- Evaluering bør inngå i departementenes virksomhetsstyring
- Krav om å bruke evalueringer der de allerede finnes
- Det bør være krav til utarbeidelse av standardiserte sluttrapporter innen bestemte frister
- Virksomhetene bør ha systemer for å spre relevant kunnskap internt i organisasjonen

I evalueringsmodellen har vi tatt ytterligere skritt for å sikre læringsutbyttet fra evalueringer. Dette har vi gjort med utgangspunkt i at klienten spesifikt har etterlyst en evalueringsmodell som sikrer både kontroll og læring, og med utgangspunkt i forskning om læring i en kjede av prosjekter, der ekstern og intern evaluering inngår blant flere kilder til læring av egen praksis.

Dette har konsekvenser for

- hvordan modellen er utformet, herunder både rollefordeling og faglig rammeverk
- hvordan evalueringene gjennomføres
- hvordan resultatene formidles
- hvordan Nye Veier tar til seg resultatene.

Det elementet i det faglige rammeverket som sterkest underbygger læringsmuligheter, er naturligvis vektleggingen av å synliggjøre endringsmekanismer og å teste dem som deler av en programteori. En bedre forståelse av hvilken innflytelse det man gjør, har på det man vil oppnå, er grunnlaget for både forbedring og fornyelse. Det er imidlertid ikke alltid sikkert at bedre løsninger eller praksiser springer ut av selve evalueringen. Et av våre intervjuobjekter i Nye Veier la vekt på at evaluator skulle påvise problemer og forbedringspotensialer, men ikke løse problemene.

Det kan ofte være krevende prosesser å lære av egen og andres erfaring. Ledelsen i organisasjonen må tilrettelegge for læring, både individuelt og – ikke minst – gjennom endring av rutiner og arbeidsprosesser, for at resultater av evalueringene skal bidra til fornying og forbedring i praksis. Figur 15 (på side 46) viser ulike faktorer som påvirker implementering av evalueringer i virksomheter, og som her kan forstås som “fremmere” av slike prosesser. Selv om ikke alle faktorene er like viktige, anbefales det at man har et bevisst forhold til disse, fordi de utgjør viktige elementer for å forstå en implementering som en prosess. Design og selve gjennomføringen av evalueringen er avgjørende for kvaliteten av evalueringen, men er også et viktig grunnlag for å lykkes med læring, forbedring og fornying av praksis. Derfor framheves også evaluators rolle som en

læringsagent. Det er viktig å understreke at faktorene i ulik grad påvirker hverandre og spiller sammen innen den aktuelle kontekst, og derfor bør vurderes i en helhetlig sammenheng.

### 3 Evalueringsmodellen i bruk

#### 3.1 Fremgangsmåte

Forløpet i evalueringsprosjekter etter denne modellen går gjennom prosessfaser som vist i Tabell 5, med angivelse av hvilken aktør som har hovedansvar for den enkelte fasen.

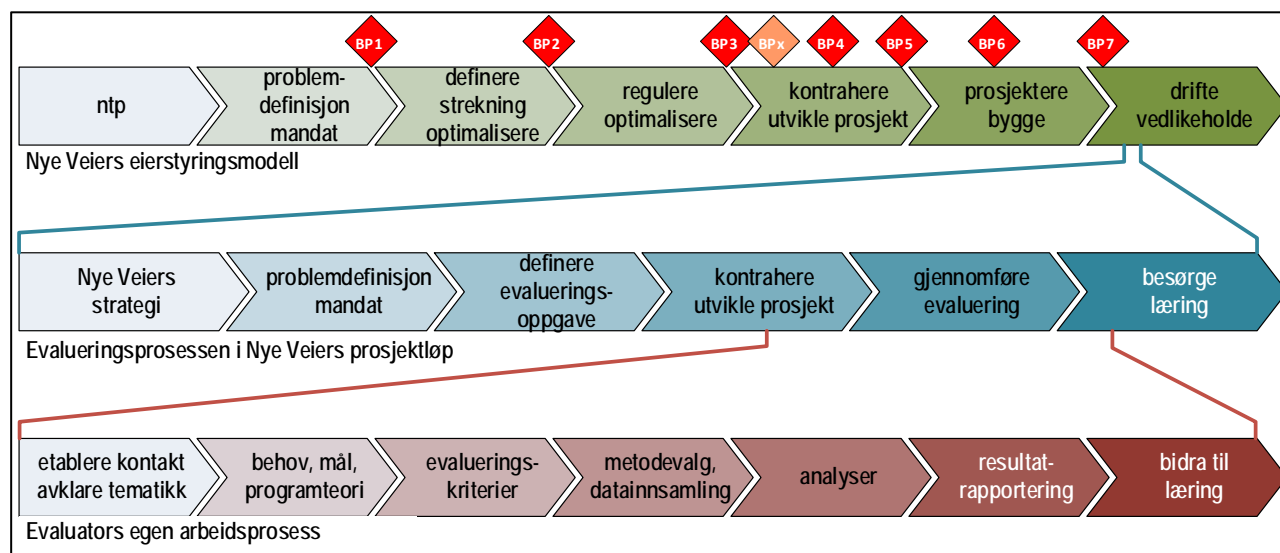
**Tabell 5: Faser og hovedansvar i evalueringsprosjekter**

Fase/trinn	Hovedansvar
Etablering av evalueringsobjekt og -formål	Nye Veier
Utarbeiding av behovsanalyse, mål og programteori	evaluator
Valg av evalueringskriterier og evalueringsspørsmål	evaluator
Valg av design, metoder og datagrunnlag, gjennomføring av datainnsamling	evaluator
Gjennomføring av analyser	evaluator
Utarbeiding av resultater og anbefalinger	evaluator
Implementering av læring	Nye Veier

At arbeidet går gjennom faser som angitt, betyr ikke alltid at de følger etter hverandre i tid. Det kan godt være overlapp. Det hender også at det er hensiktsmessig å gå tilbake og revurdere valg gjort i en tidligere fase i evalueringsprosjektet, selvstøtt underlagt en enkel kost-/nyttevurdering av å gjøre ekstra arbeid.

At en aktør har et hovedansvar, betyr heller ikke at det er vanntette skott. Tvert imot er det hensiktsmessig å ha dialog mellom aktørene for å trekke på kompetanse fra begge sider.

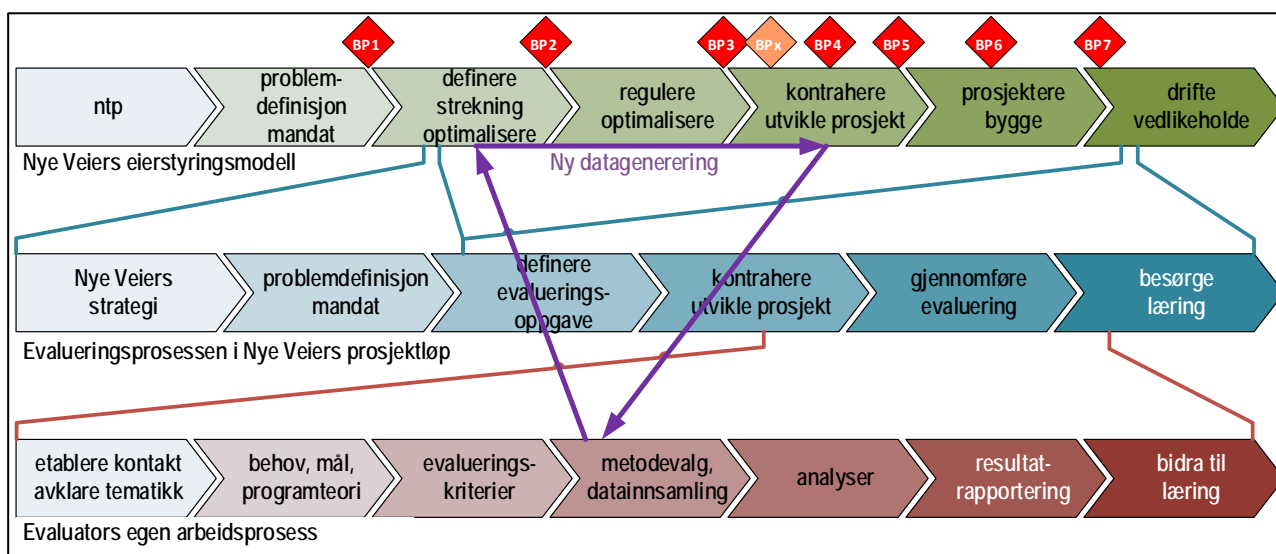
Planlegging og gjennomføring av et evalueringsprosjekt kan også fremstilles som en normal foretelse i Nye Veiers arbeid med vegstrekninger, så snart man får innarbeidet en rutine på det. I Figur 11 viser vi hvordan prosessforløpet for et standard etterevalueringprosjekt etter denne modellen (jfr Tabell 5) kan inngå i Nye Veiers arbeidsforløp.



**Figur 11: Evalueringsprosess i Nye Veiers eierstyringsmodell**

Øverst i figuren er Nye Veiers standard eierstyringsmodell for arbeidet med en vegstrekning, med beslutningspunkter (BP) for videreføring av det enkelte prosjektet inntegnet.<sup>20</sup> I midten ser vi vårt forslag til Nye Veiers prosesskart for arbeidet med en evaluering som de bestiller. Forarbeid, gjennomføring og etterarbeid forløper i normalt tilfellet samlet gjennom en forholdsvis kort periode på et tidspunkt der det foreligger tilstrekkelige data til å gjøre en empirisk etterevaluering. Merk også vår anvisning i kapittel 2.7 om at en læringsorientert oppdragsgiver har ansvar for å arkivere prosessdata som er relevante for fremtidige evalueringer, slik at de blir tilgjengelig. Nederst vises de trinnene som evaluators går gjennom, som en leverandør til Nye Veiers evalueringsoppdrag. Merk at evalueringsmodellen legger opp til en enkel samspillmodell med evaluators, slik at evalueringsoppdraget kan gjennomføres på en faglig forsvarlig måte. Modellen gir også mulighet for en interaktiv arbeidsform etter levering av evalueringsrapport, med tanke på å implementere læringspunkter.

Evalueringsmodellen kan også brukes til å evaluere kortere faser av et vegprosjekt, og de kan gjennomføres på andre tidspunkt. I Figur 11 svarer dette til å flytte den turkise ekspansjonslinjen mot venstre, til et annet prosessstrinn i vegprosjektet. Ulike faser av et evalueringsarbeid kan også være fordelt over to eller flere tidsperioder, som illustrert i Figur 12.



**Figur 12: Eksempel på alternativ plassering av evaluering**

Eksempelet i figuren viser et tilfelle der det er aktuelt å evaluere utprøving av et nytt element i arbeidet med å definere en strekning, med henblikk på hvilken betydning det har for de neste fasene. Da bør evalueringsskjemaet defineres før utprøving av det nye elementet starter, slik at evaluators kan bistå med å definere hvilke data som Nye Veier (eller evaluators) trenger å samle gjennom hele forløpet for at selve evalueringen, som vil foregå etter at det foreligger tilstrekkelige data for en empirisk evaluering, skal kunne besvare de viktigste evalueringsspørsmålene.

Arbeidet med trinnene i Tabell 5, under ulike utforminger av et evalueringsskjema, blir gjennomgått i kapitlene 3.2 til 3.8.

Evalueringsteamet må ha kompetanse på evaluering generelt, på evalueringsskjemaet spesielt, og på de substansielle forhold som skal evalueres, herunder både i prosjektfagene og i samfunnsøko-

<sup>20</sup> Som den forelå sommeren 2022. Eierstyringsmodellen revideres med mellomrom etter behov, men hovedtrekkene ligger ganske stabilt.

nomisk analyse i vegsektoren. Det innebærer også at dersom et hovedformål med evalueringen for eksempel er av geoteknisk art eller dreier seg om regulering etter plan- og bygningsloven, bør kompetanse på disse områdene være representert. I et vegprosjekt er det imidlertid så mange fagspesialiteter som bidrar at det i et normalt etterevalueringsprosjekt ikke vil være verken nødvendig eller hensiktsmessig å ha alle disse kompetansene i evalueringsteamet.

### 3.2 Trinn 0 Forarbeid: Avgrense evalueringsobjekt, formål og omfang

Det er oppdragsgivers (Nye Veiers) oppgave, som problemeier, å velge evalueringsobjekt og definere formålet med evalueringen, før evaluator velges. Begge deler bør imidlertid diskuteres med evaluator i oppstarten av arbeidet, ikke minst for gjensidig forventningsavklaring, håndtering av eventuelle faglige problemer knyttet til kombinasjonen av avgrensning og formål, og ressursfordeling mellom ulike deler av evalueringen. Nye Veiers erfaring med samspillsprosesser kan komme til nytte i slike avklaringer. Erfaringsmessig vil et altfor snevert mandat kunne begrense evaluators mulighet til å adressere spørsmål som kan vise seg å være viktige, noe som igjen kan slå tilbake på evalueringens troverdighet.

En full utnyttelse av evalueringsmodellens kontrollfunksjon, med god belysning av alle dens seks evalueringskriterier, forutsetter i utgangspunktet at evalueringsobjektet er planlegging og gjennomføring av et avgrenset investeringsprosjekt (en vegstrekning), og at evalueringen skjer noen få år etter at vegen er trafikksett, slik at bruken av vegen (trafikk, vedlikehold osv.) har satt seg, slik at det kan finnes data om faktiske effekter og virkninger.

Det er også klart mulig å gjennomføre en liknende evaluering rett etter avslutning av utbyggingsprosjektet. Dette kan være særlig aktuelt dersom det er kontroll av prosjektgjennomføringen i byggefasen som er hovedfokus. Da er det en fordel for datatilgang og datakvalitet, særlig for tilgang på intervjuobjekter med prosjektet i friskt minne, at man ikke venter for lenge etter prosjektavslutning<sup>21</sup>. En slik evaluering kan ikke si så mye mer om forventede effekter og virkninger enn det man kunne si i forkant av selve byggingen, med mindre man kan gjøre noen revideringer dersom den bygde vegstrekningen avviker vesentlig i kvaliteter fra det som var planlagt.

Det er selvsagt mulig å kombinere en tidlig evaluering av faktisk effektivitet i prosjektgjennomføringen med en senere evaluering av faktiske effekter og virkninger. Da er det fornuftig å ikke legge opp til å gjøre dobbelt arbeid. Hvor hører i så fall data om de tidlige fasene i prosjektet hjemme, det arbeidet som ligger før selve prosjektgjennomføringen? I et prosjekt som følger statens prosjektmodell bokstavelig, er effektiviteten i prosjektgjennomføringen (operasjonell vellykkethet) kun avhengig av entreprenøren og risikofordelingen mellom byggherre og entreprenør. For kontrollformål er det derfor ikke noe stort behov for å dra inn mye data fra tidlige faser. Entreprenør og prosjektleder kan ikke holdes til ansvar for byggherrens trasévalg, bestilt vegkapasitet og andre overordnede forhold. De har heller ikke hatt noen innflytelse på hvordan kostnadsbildet har utviklet seg fra prosjektet skisseres første gang til entreprisen er avtalt.<sup>22</sup>

---

<sup>21</sup> Det hender at prosjekter ikke avsluttes før lenge etter at leveransen er gjennomført og vegen trafikksett. Det kan for eksempel være konflikter mellom byggherre og entreprenør som må gjøres opp, med konsekvenser for samlet kostnad. En måte å håndtere dette på er å betrakte uavklarte kostnader i et usikkerhetsperspektiv, snarere enn å vente til alt er avklart for å få en offisiell sluttsum. På den andre siden kan et viktig tema nettopp være hvordan slike konflikter har oppstått og blitt avgjort, noe som kan tale til fordel for å vente.

<sup>22</sup> Denne utviklingen av kostnadsbildet frem til utbyggingskontrakt kartlegges imidlertid vanligvis, av praktiske grunner, sammen med den videre utviklingen av kostnadsbildet frem til sluttført leveranse.

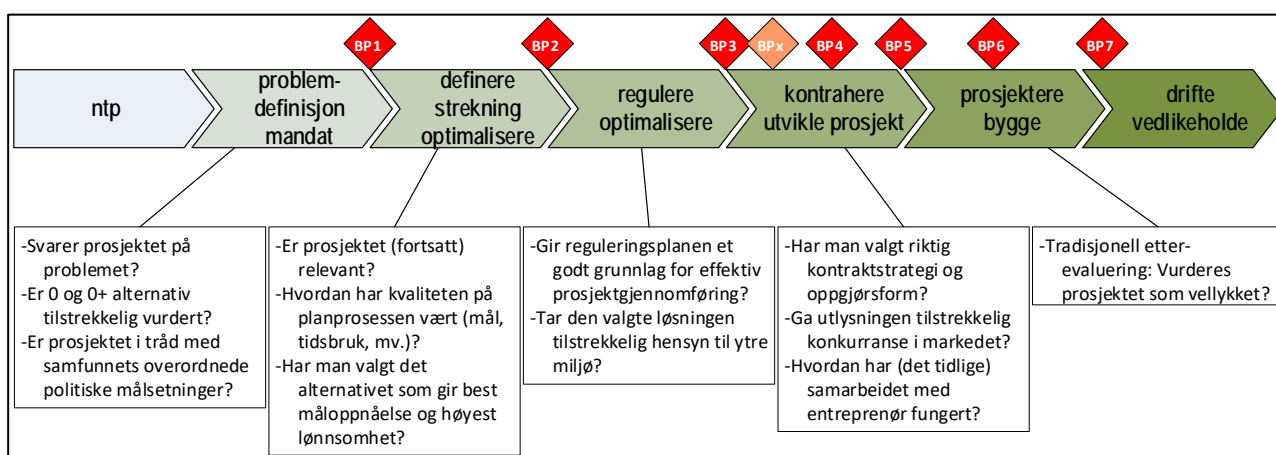


På den andre siden er det heller ikke strengt tatt påkrevd å gå langt bakover i tid dersom evalueringen av effekter og virkninger utelukkende har et kontrollformål. Imidlertid er det svært sjelden at evalueringer med et kontrollformål ikke også har et læringsformål, om ikke annet for å identifisere "hva som gikk galt" eller "hva som gjorde prosjektet vellykket". Evalueringsmodellen har dessuten overtatt fra Concept-modellen et fokus på å generere kunnskap om betydningen av valg i tidlige faser på hvordan prosjekter utformes og gjennomføres, og på hvilke effekter og virkninger de bidrar til. Dette omfatter også hvordan kostnadsbildet for prosjektet utvikler seg over tid.

Dessuten er det sjelden at arbeidsdelingen og faseinndelingen er så entydig som idealmodellene antyder. Samspillskontrakter mellom byggherre og entreprenør er et eksempel på at fasene tematisk sett glir inn i hverandre. Likeledes ser vi tendenser til at valg i prosjektgjennomføringsfasen nå forsøkes gjort ut fra et nytte-/kostnadsperspektiv, og ikke bare fra et kostnadsperspektiv. Dette innebærer at effektivitet i prosjektgjennomføring ikke nødvendigvis kan evalueres helt uavhengig av effekter og virkninger. Allikevel er det som nevnt gode argumenter for å gjøre ferdig en egen evaluering av effektivitet i prosjektgjennomføring ved leveransetidspunktet, forutsatt at den også inngår som en del av senere evaluering med vekt på taktisk og strategisk vellykkethet, og der sistnevnte også tar hovedansvaret for å studere prosjektets tidlige faser (og dets forløpere).

I det tilfelle at en vegstrekning gjennomføres som bygging av flere delstrekninger, kan det være hensiktsmessig å evaluere hver av delstrekningene hver for seg med tanke på effektivitet i prosjektgjennomføringen og samlet med tanke på effekter og virkninger. Hvis det går veldig lang tid før en siste delstrekning eventuelt bygges ut, kan det være hensiktsmessig å ikke vente med evalueringen av effekter og virkninger, og heller behandle den siste delstrekningen som en realopsjon.

Figur 13 illustrerer et utvalg av potensielle evalueringsspørsmål på ulike tidspunkter i Nye Veiers eierstyringsmodell. Utvalget er gjort på grunnlag av intervjuer i Nye Veier og kjennskap til vegprosjektevaluering mer allment, men det er verken komplett eller ment som noen begrensning. Betydelig mer detaljerte avgrensninger vil også være mulig. Evalueringsmodellen er robust i forhold til avgrensning av evalueringsskjemaer, så lenge evalueringsteamet kan rå over relevant og tilstrekkelig kompetanse og datagrunnlag.



**Figur 13: Utvalgte aktuelle evalueringstema på ulike tidspunkt**

I denne figuren er "tradisjonell etterevaluering" en forkortelse for den evalueringssjangeren som Concepts etterevalueringer gir seg inn på, nemlig en ex post-evaluering av hele prosjekter. Alle evalueringsskjemaer som berøres av evalueringsspørsmålene ovenfor, kan underkastes evaluering

både fra et kontroll- og et læringsperspektiv. Læringsbehovet kan være generelt og/eller knyttet til spesielle erfaringer gjort (eller planlagt) i et prosjekt. Evalueringsmodellen er lagt opp slik at det er kartleggingen av resultater, effekter og virkninger som ligger til grunn for vurdering etter de seks evalueringskriteriene, som i første rekke er kontrollorientert. Denne kartleggingen er også et godt utgangspunkt for videre undersøkelser av aktuelle læringsformål, ettersom den synliggjør de prosessene eller mekanismene som forbinder aktiviteter, resultater, effekter og virkninger med hverandre i den foreliggende konteksten. Ytterligere undersøkelser av utvalgte mekanismer fra denne kartleggingen danner så grunnlaget for ny læring fra prosjektet.

Når evalueringsobjektet og formålet er avgrenset på overordnet nivå, vil det også være behov for å utvikle konkrete evalueringsspørsmål. En første utgave av de viktigste evalueringsspørsmålene bør utvikles i Nye Veiers forarbeid. Det hender også at det er slike konkrete evalueringsspørsmål som utløser behovet for en evaluering og dermed også avgrensningen av det øvrige forarbeidet. Samtidig er det viktig å diskutere utformingen av evalueringsspørsmålene med evaluator i neste omgang, slik at de blir utformet slik at de kan besvares.

I det ovenstående har vi stilltiende gått ut fra at evalueringer gjøres etter at det arbeidet som skal evalueres, er gjennomført. For eksempel kan det være en etter- eller underveisevaluering av arbeidet med reguleringsprosessen.<sup>23</sup> Noen ganger er det erfaringene med dette arbeidet i seg selv som skal evalueres. Andre ganger er man mer opptatt av hvilken betydning dette arbeidet har for sluttproduktet av vegprosjektet. En kan da i prinsippet definere resultater, effekter og virkninger av dette arbeidet på senere deler av vegprosjektet, og vente med datainnsamling til slike utfall har satt seg. I eksempelet med reguleringsprosessen kan man da undersøke for eksempel hvorvidt reguleringen har tatt tilstrekkelig hensyn til selve byggeprosessen, og dermed til kostnadene. Jo tidligere man gjør evalueringen, jo mer aktuell blir læringen.

Hvis man ikke vil vente på faktiske data om utfall (resultater, effekter eller virkninger) av det konkrete arbeidet, kan man imidlertid evaluere arbeidet opp mot antatte utfall. Dette kan gi grunnlag for en tydeliggjøring av de mekanismer og forståelser av kontekstens betydning som de ansvarlige eksplisitt eller implisitt har lagt til grunn for de forventede utfall. Dermed beveger evalueringen seg ikke bare inn på læringsspørsmål, men også på å innta en forhåndsevalueringssposisjon. Dette kan gi opphav til interessante problemstillinger, som naturlig nok også kan etterprøves i form av etterevalueringer i senere faser.

Det er ikke noe i veien for å bruke evalueringsmodellen til forhåndsevalueringer på denne måten. Bruken av programteori som evalueringsmetode og de seks faste evalueringskriteriene som vurderingsstandard kan gi nyttig læring. En senere utsjekk mot faktiske prosjektutfall er naturligvis nødvendig.

### 3.3 Trinn 1: Kartlegge behov, målformuleringer og programteori

To av evalueringsmodellens strategiske vurderingskriterier gjelder løsningens relevans og levedyktighet. Begge forutsetter at prosjektets utløsende behov og andre behov som skal tilfredsstilles når vegen først bygges, er godt beskrevet. I den forbindelse må man også beskrive andre målgrupper som berøres, ikke bare de som har de behov som det knyttes mål til. Effektmålene skal også være nært knyttet til å tilfredsstille disse behovene. Prosjektets styringsdokumenter og samtaler med

---

<sup>23</sup> Noen vil kalle dette en underveisevaluering av vegprosjektet som helhet. Ut fra metodiske overveininger har vi i dette tilfellet valgt å snakke om en etterevaluering av en del av vegprosjektet.

nasjonale, regionale og lokale transportmyndigheter og interessegrupper gir vanligvis et godt bilde av behovene.

Gode målformuleringer er viktige for å kunne avgjøre graden av måloppnåelse. Typiske mål i vegutbyggingsprosjekter er vist i Tabell 6, der vi har utelatt konkrete måltall, som vil være prosjektspesifikke.

**Tabell 6: Eksempler på typiske mål i vegutbyggingsprosjekter**

Målnivå	Målformuleringer
<b>Samfunnsmål</b>	<ul style="list-style-type: none"> <li>• Binde sammen bo- og arbeidsmarkedet i prosjektets influensområde</li> <li>• Bidra til utvidet offentlig og privat tjenestetilbud</li> <li>• Styrke grunnlaget for bosetting og næringsutvikling</li> <li>• Bidra til en fremtidsrettet stedsutvikling</li> </ul>
<b>Effektmål</b>	<ul style="list-style-type: none"> <li>• Bidra til reduserte reisekostnader for trafikantene (f.eks. målsatt i form av redusert reisetid i antall minutter)</li> <li>• Fjerne ferje- og ulempekostnader for trafikantene</li> <li>• Redusere antall hardt skadde og drepte med (en gitt ulykkesfrekvens)</li> </ul>
<b>Resultatmål</b>	<ul style="list-style-type: none"> <li>• Vegen skal åpnes for trafikk innen (en gitt dato) iht. vedtatte reguleringsplaner</li> <li>• Ferdigstille prosjektet innenfor godkjent kostnadsramme/styringsramme</li> <li>• Leveranse med kvalitet som beskrevet i planer og håndbøker</li> <li>• Prosjektet skal gjennomføres med skadehyppighet (H-verdi) lik 0</li> </ul>

Prosjektevaluering etter Concept-modellen forutsetter ikke bare gode målformuleringer, men også en god målstruktur. Så også i denne modellen. En kritisk gjennomgang av målformuleringer og målstruktur i prosjektets styrende dokumentasjon er viktig av flere grunner:

- Vurderingen av operasjonell og taktisk vellykkethet skal normalt skje i det minste ved å sammenlikne oppnådde resultater og effekter mot avtalte resultatmål og effektmål
- Dersom effektmål er formulert kun som retning og ikke kvantifisert eller konkretisert på andre måter, kan vurderingen bli noe tilfeldig, selv om effektene i seg selv er målt
- For å kunne vurdere i hvilken grad innsatsen har bidratt til resultater og resultatene til effekter, er det nødvendig med en målstruktur som viser disse sammenhengene, enten de er godt eller ikke så godt underbygd
- Eventuelle målkonflikter kan også best oppdages dersom det er tydelig hvordan målene i målstrukturen antas eller forutsetter å samvirke
- Samfunnsmål er erfaringsmessig spesielt vanskelig å formulere godt, men samfunnsmål formulert som aggregerte brukernytteeffekter er ikke tilstrekkelig (og heller ikke alltid påkrevd)

Selv om mange målformuleringer og målstrukturer for store prosjekter i dag følger anvisningene i Statens prosjektmodell (FIN 2010), er det fortsatt mange prosjekter som ikke lar seg evaluere godt nok etter Concept-modellen og dens avleggere uten en reformulering og restrukturering av målene slik de er fremstilt i styrende dokumentasjon. Det er da evaluators oppgave å reformulere og restrukturere målene slik at de gjør evalueringen mulig (eller enklere), samtidig som relevante aktører går god for at de omarbeidede målene avspeiler intensjonene bak prosjektet på en gjenkjennelig måte.

Evaluering etter denne modellen forutsetter ikke bare vurdering av måloppnåelse som en kontroll-funksjon mot avtalte mål, men også som en identifisering av mulige årsakssammenhenger mellom innsats, resultater, effekter og virkninger. Målstrukturen, eventuelt i bearbeidet form, er et godt utgangspunkt for å videreutvikle en programteori som også synliggjør hvilke endringsmekanismer som er antatt å virke for at måloppnåelse på et lavere nivå skal styrke muligheten for måloppnåelse på et høyere nivå.

Programteorien utarbeides basert på resultater fra behovsanalysen og målformuleringer, samt tilgjengelig kunnskap og antagelser om årsakssammenhenger mellom innsats, resultater, effekter og videre virkninger. Den skal inkludere to sentrale elementer som har betydning for prosjektets endelige utfall:

- (1) Hvordan var vegprosjektet planlagt utformet og implementert? En beskrivelse av hvilke betingelser og eksterne faktorer (kontekst) som må være oppfylt for at prosjektet skal oppnå ønsket effekt og virkning.
- (2) Hvorfor, eller på hvilken måte, var prosjektet forventet å forårsake en ønsket endring? En beskrivelse av de underliggende mekanismene som gjør at prosjektet medfører en forandring i tråd med overordnede mål.

En første utgave av en programteori kan se ut som i Figur 7 på side 12. Hvorvidt man trenger en mer detaljert utgave for deler eller hele programteorien, som i Figur 16 på side 53, avgjøres etter hvert som man eventuelt oppdager sammenhenger som det er nyttig å studere i større detalj. Programteorien kan ha ulik grad av detaljering og kompleksitet, avhengig av prosjektets modenhet, eksisterende kunnskapsgrunnlag og evalueringens formål. Fremstillingen gir uttrykk for vurderings-teamets forståelse av hvordan det aktuelle prosjektet virker, og vil dermed være en felles referanseramme for alle aktører som er involvert i vurderingsarbeidet.

### 3.4 Trinn 2: Velge vurderingskriterier og vurderings spørsmål

Modellens seks hovedkriterier gir som sagt hver sin innfallsvinkel som er viktig for den samlede bedømmelsen av prosjektet. Det er allikevel ikke sikkert at alle seks kriteriene bør vies like stor oppmerksomhet i samtlige vurderingsprosjekter. I Concepts portefølje av vurderinger er det for eksempel lagt vekt på å prioritere de taktiske og strategiske kriteriene, og heller bruke mindre ressurser på de operasjonelle. En begrunnelse for dette er at prosjektgjennomføringen ofte underkastes egne vurderinger ved prosjektavslutningen, mens taktiske og strategiske vinklinger sjeldnere får noen stor plass. Concept bruker med andre ord sin vurderingsmodell til å fylle et viktig gap i vurderingspraksisen.

I Tabell 7 viser vi en del mulige vurderings spørsmål, relevante sammenligningsgrunnlag for tilstands- eller ytelsesmålinger, aktuelle indikatorer og relevante datakilder, alt sortert under vurderingsmodellens seks vurderingskriterier. Vi har ikke satt opp direktekoblinger mellom kolonnene i tabellen fordi mange av sammenhengene er mange-til-mange; en og samme datakilde kan for eksempel være til nytte for flere vurderings spørsmål. Innholdene i tabellen er langt fra tilfeldig valgt, men ikke alle elementer vil være aktuelle i alle vurderingsprosjekter, og det kan være mange andre forhold som ikke har fått plass i tabellen.<sup>24</sup>

<sup>24</sup> De to siste kolonnene (om datakilder) hører strengt tatt til i kapittel 3.5

**Tabell 7: Operasjonalisering av evalueringskriterier**

Effektivitet i prosjektgjennomføring				
Mulige evaluerings-spørsmål	Sammenlignings-grunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Har kostnadsestimatet endret seg over tid?</p> <p>Har prosjektet blitt gjennomført innenfor vedtatt styrings- og kostnadsramme?</p> <p>Ble prosjektet åpnet for trafikk i henhold til plan?</p> <p>Valgte prosjektet en gunstig entreprisstruktur gitt prosjektets natur og markedet?</p> <p>Er prosjektet bygget til forventet kvalitet?</p> <p>Hvordan ble SHA og ytre miljø i anleggsperioden planlagt og fulgt opp?</p> <p>Var rammene realistiske sammenliknet med andre prosjekter?</p>	<p>0-estimat</p> <p>Beslutningspunkt BP2 «Rammer, mål og Strekningsoppdeling»</p> <p>Beslutningspunkt BP3 «Reservere økonomirammer og klar for marked»</p> <p>Styringsdokument</p> <p>Stortingsproposisjon Meld. St. 25 (2014-2015)</p> <p>Eventuell KS1</p> <p>Andre vegprosjekter</p> <p>Vegnormaler</p> <p>Normal planleggings- og byggetid i tilsvarende prosjekter</p>	<p>Prosent kostnadsreduksjon mot hhv. 0-estimat og estimat ved BP2</p> <p>Sluttkostnad</p> <p>Kontraktsendringer</p> <p>Tidspunkt trafikkåpning</p> <p>Tidspunkt prosjektavslutning</p> <p>Løpemeterpris</p> <p>Byggetid</p> <p>Planleggingstid</p> <p>Samarbeid mellom byggherre og entreprenør</p>	<p>KVU og KS1-rapport</p> <p>Nye Veiers eget kostnadsanslag m/ usikkerhetsanalyse</p> <p>KS2-rapport</p> <p>Prosjektstyringsplan</p> <p>Kontraktstrategi</p> <p>SHA-plan</p> <p>Styringsdokument</p> <p>Tilbudsprotokoll</p> <p>Teknisk og økonomisk sluttrapport</p> <p>Prosjektregnskap med årsfordelt forbruk</p> <p>Reguleringsplan</p> <p>Evalueringsrapporter eller data om andre vegprosjekter</p>	<p>Befaring</p> <p>Intervjuer med prosjektledelse, styringsgruppe, prosjektdeltakere, leverandører, lokale/regionale transportmyndigheter og interessegrupper</p>
Realisering av effekter				
Mulige evaluerings-spørsmål	Sammenlignings-grunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Har prosjektet ført til økt kapasitet?</p> <p>Har prosjektet ført til redusert reisetid?</p> <p>Har prosjektet medført en reduksjon i generaliserte reisekostnader?</p> <p>Er trafikksikkerheten forbedret som følge av prosjektet (ny og evt. gammel veg)?</p>	<p>Reisetidsmålinger før åpning</p> <p>Antall registrerte ulykker før åpning</p> <p>Antall registrerte ulykker på tilsvarende veger</p>	<p>Endring i kapasitet</p> <p>Endring i reisetid (minutter)</p> <p>Endring i generaliserte reisekostnader (inkl. bompenger)</p> <p>Endring i antall drepte og hardt skadde</p> <p>Endring i generaliserte reisekostnader</p> <p>Oppetid</p>	<p>Styringsdokument</p>	<p>Reisetidsmålinger før åpning</p> <p>Reisetidsmålinger etter åpning</p> <p>STRAKS ulykkesregister</p> <p>Nasjonal vegdatabank (NVDB)</p> <p>Intervju med målgruppe</p>



Andre virkninger				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Har prosjektet ført til økt konkurransekraft for lokalt næringsliv?</p> <p>Har prosjektet ført til at bo- og arbeidsmarkedet i regionen er lettere integrert?</p> <p>Har prosjektet ført til økt arbeidsmarkedsdeltakelse?</p> <p>Hvordan har prosjektet påvirket ytre miljø og klimagassutslipp?</p> <p>Hvilke avbøtende tiltak er eventuelt gjennomført?</p> <p>Har prosjektet påvirket nytte i andre transportprosjekter?</p>	<p>Situasjon i influensområde før åpning</p> <p>Utvikling i andre (tilsvarende) kommuner</p> <p>Miljø- og utslippsvirkninger i sammenliknbare prosjekter</p>	<p>Endring i antall sysselsatte</p> <p>Endring i antall pendlere</p> <p>Endring i antall nyetablerte bedrifter</p> <p>Antall bosatte innenfor pendlingsavstand (ca. 45 min)</p> <p>Endring i befolkning</p> <p>Antall støyutsatte boliger</p> <p>Antall dekar tapt vs. nyskapt (el. restaurert/kompensert) av ulike arealtyper, planlagt og gjennomført</p> <p>Netto antall dekar inngrep i naturområder med nasjonal eller vesentlig regional verdi</p> <p>Mengde gjenbruk (veg og masser)</p> <p>Mengde deponert avfall</p> <p>Barriereeffekter for vilt (ant. meter viltgjerder)</p> <p>Antall vilt påkjørsler</p> <p>CO2 utslipp i byggefase og driftsfase</p> <p>Virkninger på andre transportprosjekter</p> <p>Endring i antall reisende med kollektivtransport</p>	<p>Kommunedelplan og reguleringsplan (m/vedlegg)</p> <p>Miljøoppfølgingsplan</p>	<p>Data for sysselsetting, befolkning og pendling fra SSB</p> <p>Intervju med lokale myndigheter og interesseorganisasjoner</p>
Levedyktighet				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Hva er sannsynlig teknisk og økonomisk levetid av den nye vegen?</p>	<p>NTP</p> <p>Fremtidige politiske prioriteringer</p> <p>Prioritering av vedlikehold</p>	<p>Utvikling i reisetid</p> <p>Utvikling i kapasitetsutnyttelse</p> <p>Livsløpskostnader</p> <p>Befolkningsutvikling</p>	<p>KVU og KS1-rapport</p>	<p>Trafikkprognoser</p> <p>SSB</p> <p>Intervju med aktører innen bl.a. drift og vedlikehold</p>

Levedyktighet				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Vil vegens nytte vedvare over dens levetid?</p> <p>Vil vegen antas å være riktig løsning også på sikt?</p> <p>Er befolkningsveksten fortsatt ventet å være samme trend som ved bevilgnings-tidspunktet?</p>				
Relevans				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Representerer vegen den riktige løsningen på det identifiserte problemet?</p> <p>Er det behov for vegen (økt kapasitet, redusert reisetid etc.)?</p> <p>Ville andre konsepter, herunder en annen utbyggingsrekkefølge, ha vært mer egnet for å løse behovene?</p> <p>Hvordan passer investeringen med samfunnets overordnede politiske mål i dag?</p> <p>Er prosjektets mål i tråd med overordnede mål og strategier for transportpolitikken?</p>	<p>Politiske prioriteringer i dag</p> <p>NTP</p> <p>Evt. KVV</p> <p>Kommunedelplan</p>	<p>Endring i ÅDT</p> <p>Endring i samfunnets mål og prioriteringer siden investeringsbeslutning</p>	<p>Nasjonal transportplan</p> <p>Andre strategiske og politiske dokumenter</p> <p>Andre vurderte konsepter (herunder nullalternativet)</p>	<p>Trafikkdata</p> <p>Intervju med relevante aktører</p>
Samfunnsøkonomisk effektivitet				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
<p>Er prosjektet samfunnsøkonomisk lønnsomt med dagens beregningsforutsetninger?</p>	<p>Beregnet lønnsomhet på beslutningstidspunkt</p> <p>Beregnet lønnsomhet i andre prosjekter</p>	<p>Netto nytte, og tilsvarende lønnsomhetsindikatorer</p>	<p>KS1</p> <p>Sluttregnskap</p> <p>Oppdaterte prognoser</p>	<p>Tidligere EFFEKT-beregninger</p> <p>Oppdatert EFFEKT-beregning</p>

Samfunnsøkonomisk effektivitet				
Mulige evalueringsspørsmål	Sammenligningsgrunnlag	Mulige indikatorer	Relevante dokumenter	Andre datakilder
Er prosjektet samfunnsøkonomisk lønnsomt med beregningsforutsetninger som på beslutningstidspunktet?				

Ettersom evalueringmodellen er tilpasset bruk på ulike tidspunkt og med fokus på ulike delprosesser i vegbyggingsprosjekter, er det også hensiktsmessig å prioritere kriteriene etter formålet med den enkelte evalueringen. Det kan for eksempel ligge en betydelig læringsgevinst i å bruke evalueringmodellen som standard modell for evaluering av prosjektgjennomføringen ved prosjektavslutning, uten særlig hensyn til vurdering av taktiske og strategiske sider. Det kan også være at noen prosjekter er så enkle at en pålitelig konstatering av utfallene er tilstrekkelig, mens andre rommer så mye endringer i arbeidsmåte eller forutsetninger at en kanskje bare bør se detaljert på betydningene av endret arbeidsmåte for neste trinn i prosessen.

### 3.5 Trinn 3: Velge design, metoder og datakilder, og samle data

Valg av undersøkelsesdesign avgjøres av hva som er gjennomførbart og kostnadseffektivt for å gi svar på de respektive evalueringsspørsmål som er formulert. Det vil si at evalueringprosjektet som helhet kan romme flere design, se kapittel 2.7. Med mindre man står overfor evalueringsspørsmål som er besvart med hell i evaluering av tidligere prosjekter, på grunnlag av gode og kjente datakilder, eller man har designet et forsøk fra bunnen av, står man ofte overfor en situasjon der man må velge mellom evalueringsspørsmål (eller i alle fall formuleringer av disse), sammenligningsgrunnlag, indikatorer og datakilder samtidig, ut fra en mulighetsvurdering, og ikke ut fra en stringent utvalgsprosess ovenfra og ned i dette hierarkiet. Da blir oppgaven å prioritere de kombinasjonene som samlet sett gir best læringsutbytte (i tillegg til en tilfredsstillende kontrollfunksjon, som nesten alltid må være med). Å gjøre en slik avveining krever både evalueringskompetanse og forståelse av den type prosjekter eller aktiviteter som skal evalueres. Det skjer også at slike avveininger må gjøres underveis i datainnsamlingen fordi data ikke alltid er så gode som forventet, eller at det dukker opp ny kunnskap som kan gjøre det hensiktsmessig å utforske nye årsaker, eller også nye utfall, eller nye datakilder (særlig nye informanter).

Den teknologiske utviklingen med utbredt bruk av sensorer i både kjøretøy og veginfrastruktur gjør at transportsystemet i seg selv kan samle inn en rekke datatyper som tidligere ikke har vært tilgjengelig innenfor transportsektoren. Tilsvarende er det eksempler på bruk av mobiltelefondata for analyse av (endringer i) trafikantatferd. Dette gir mulighet for analyser som i større grad kan forklare og dokumentere hvordan vegprosjektet utløser en rekke mekanismer og effekter i transportsystemet. Det stiller imidlertid større krav til den som skal utføre analysene, fordi man står overfor muligheter og datatyper som man inntil videre har begrenset erfaring med å håndtere (f.eks. big data, usikker datakvalitet, personvern hensyn og behov for nye analysemetoder).

Innsamling av kvalitative og kvantitative data fra ulike kilder skjer som i et vanlig forskningsprosjekt. En forskjell kan være at en må ty til langt flere sekundærdata enn i andre prosjekter. Erfaring tilsier at selv om data finnes, kan det være vanskelig, tidkrevende og kostbart å få tak i dem. En



evalueringsorientert organisasjon vil over tid lære hvilke data som etterspørres for evalueringer, og sørge for at disse blir gjort tilgjengelig på en mer systematisk måte.

Selve datainnsamlingen er en fase der man, i tillegg til å samle inn data, reduserer usikkerheten i planen ved å oppdage om tilgjengelighet og kvalitet i data er som forutsatt, og om data gir opphav til nye eller reviderte evalueringsspørsmål. Det er derfor viktig å ha et kontinuerlig blikk på dette med tanke på korrektive tiltak.

Vi har forutsatt at det kan være mange aktuelle sammenligningsgrunnlag for de subjektive eller objektive data som kjennetegner det enkelte prosjekt. Da mener vi ikke bare som alternativer, men at det kan være aktuelt å sammenlikne for eksempel vegens realiserte løpemetertpris både med tidlige anslag, med et formulert kostnadsmål på dette detaljeringsnivået, med sammenlignbare vegger, og andre standarder. Det er viktig å være tydelig på hva det er man sammenligner med, og hva som er formålet med sammenligningen.

En av de situasjonene der det er aktuelt å sammenligne ett og samme funn mot ulike standarder, er når endringer i forutsetningene har endret seg etter at prosjektets styrende dokumentasjon er på plass. For kontrollformål er det viktig å forholde seg til avtalte mål som sammenligningsgrunnlag. Unntak kan gjøres der nye lover og forskrifter setter tidligere resultat-, effekt- eller samfunns mål til side; da må de nye reglene gjelde, så sant prosjektet ikke har kommet for langt til at det vil være mulig å rette seg etter de nye kravene. Dersom slike endringer derimot bare betyr at andre valg må gjøres, mens målformuleringene står, blir endringene å regne for endringer i konteksten som prosjektet må forholde seg til.

For læringsformål er det derimot viktigere i en slik situasjon nettopp å sammenligne med annen praksis. Et funn i en evaluering kan altså være at man har nådd målet ved å følge en viss praksis, men at målet nå er foreldet, slik at det vil være bedre å følge en annen praksis.

### 3.6 Trinn 4: Analysere dataene

Valg av analysemetoder avgjøres i realiteten ved utforming av metodeoppsettet og planlegging av datainnsamlingen. Det vil som regel være nødvendig å kombinere både kvantitative og kvalitative analysemetoder for å forstå fullt ut hvordan prosjektet fungerer/har fungert. De kvalitative analysene bidrar gjerne til å identifisere relevante effekter (både de som er forventet og eventuelt utilsikket), og til å skape forståelse for de mekanismene som bidrar til å utløse disse effektene. De kvantitative analysene gir i større grad kunnskap om effektens størrelse og konfidensintervall, gjennom mulighet for statistisk testing.

Tabell 8 gir en oversikt over typiske analyser knyttet til de ulike evalueringskriteriene, men i praksis er det en fordel å kombinere flere analyser knyttet til hvert enkelt evalueringskriterium (metode-triangulering).

**Tabell 8: Eksempel på typiske analyser knyttet til overordnede evalueringskriterier**

Evalueringskriterier	Analyser
Effektivitet i prosjektgjennomføring	Analyse av prosjektets resultatmål med utgangspunkt i styringsdokumenter, dokumenterte resultater, observasjon og intervju
Realisering av effekter	Kvantitative effektstudier (+ kvalitative analyser)
Andre virkninger	Kvantitative effektstudier (+ kvalitative analyser)

Evalueringskriterier	Analyser
Relevans	Behovsanalyse med utgangspunkt i samfunnsutvikling og overordnede transportpolitiske mål
Levedyktighet	Analyse av trafikkprognoser og samfunnsutvikling, usikkerhets- og føl-somhetsanalyser
Samfunnsøkonomisk effektivitet	Samfunnsøkonomiske analyser

Det finnes et betydelig antall ulike analyseteknikker for kvantitative data innsamlet under ulike design. Se for eksempel Mohr (1995) for effektanalyser i programevaluering generelt og det britiske finansdepartementets veiledning til evalueringsdesign for en samling teknikker brukt i bl. a. vegsektoren i Storbritannia (HM Treasury 2020)<sup>25</sup>. Mange av disse forutsetter store datamengder for å kunne gi pålitelige resultater.

For kvalitative data er det etter hvert utviklet en rekke teknikker som skal styrke både konsistens og etterprøvrbarhet, men i situasjoner med overkommelige datamaterialer som i typiske evalueringsprosjekter er øvde forskere godt i stand til å trekke ut essensen av intervjuene uten formalisert og tidkrevende arbeid. Det er dog en fordel for påliteligheten at man arbeider to og to sammen om dette.

Samfunnsøkonomiske analysemetoder for denne typen prosjekter følger normalt anbefalingene fra Finansdepartementet (FIN 2014) om hvordan en av de følgende undersøkelser skal gjøres:

- *nytte-kostnadsanalyse*, som gir et tallmessig svar på om prosjektet er lønnsomt i kroner
- *kostnads-virkningsanalyse*, som tallfester kostnadene, men beskriver nytten kvalitativt (skjønnsmessig vurdering)
- *kostnads-effektivitetsanalyse*, som tar nyttesiden for gitt, men analyserer kostnadene og vurderer om prosjektet er kostnadseffektivt

*Det (kvasi-)eksperimentelle metodeoppsettet* legger kontrafaktiske analyser av testsituasjoner med og uten tiltaket til grunn for bestemmelse av kausalitet. Dette betyr i praksis at observerte nytteeffekter av prosjektet må sammenlignes med utviklingen i en kontrollgruppe (f.eks. representative vegstrekninger). Slike analyser inngår gjerne i studier av teknisk ytelse og kvalitet, samt målbare atferdsendringer og effekter på bl.a. reisetid og ulykkesutvikling. Betydningen av kontekst og situasjonsavhengige faktorer vil normalt ikke adresseres i slike analyser.

*Det ikke-eksperimentelle (teoribaserte) metodeoppsettet* har en mekanismeorientert tilnærming til kausalitet, hvor resultatenes validitet må sikres gjennom en systematisk testing av de ulike sammenhengene i programteorien. Kausaliteten ivaretas gjennom følgende betingelser:

- Programteorien fremstår som plausibel og konsistent
- Prosjektets innsats og aktiviteter ble implementert iht. plan
- Underliggende antakelser ble bekreftet, effekter og virkninger observert
- Kontekst og eksterne faktorer ble identifisert og vurdert

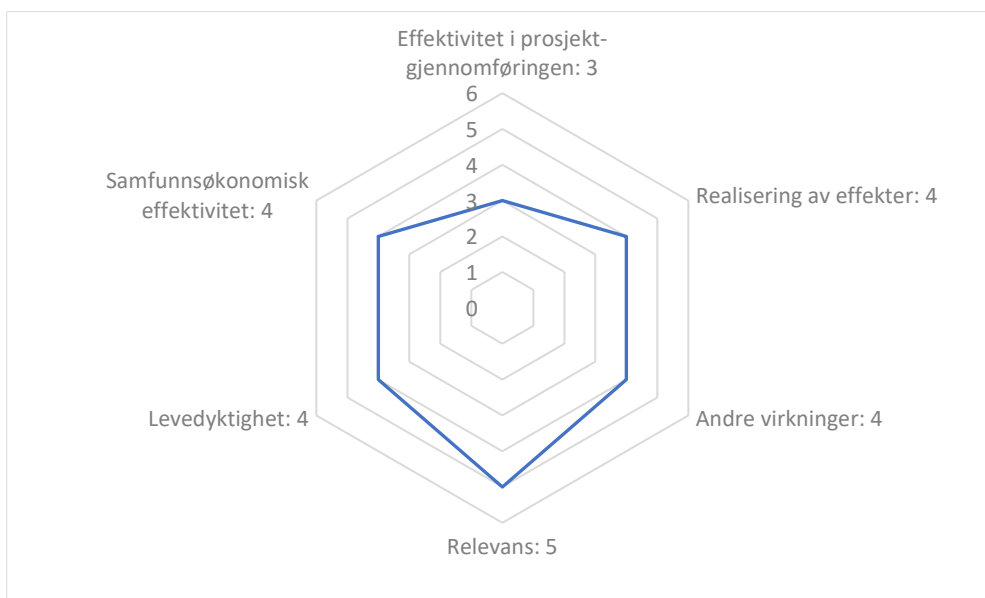
Den teoribaserte fremgangsmåten benyttes gjerne for å fremskaffe kunnskap om årsakssammenhenger, rammebetingelser og eksterne faktorer som påvirker i hvilken grad man vil/har lyktes med

<sup>25</sup> Se <https://www.gov.uk/government/publications/the-magenta-book> for flere detaljer.

prosjektet. Slike analyser kan også gi grunnlag for å vurdere hvorvidt det hadde vært mulig å oppnå større effekter ved å justere på enkelte innsatsfaktorer eller gjennomføringsstrategi i prosjektet.

Vurdering av prosjektets levetid bør også inkludere analyser av usikkerhet og følsomhet for endringer på vegen eller omgivelsene: Hva betyr det om antakelsene som ligger til grunn for programteorien er feil? Hvilket utslag gir det om trafikkvolum endres vesentlig? Hvor fleksibel er løsningen med tanke på teknologitrender i transportsektoren?

Hvert av de seks overordnede evalueringskriteriene skal gis en poengscore som angir evalueringsteamets vurdering av prosjektets vellykkethet, på en skala fra 1 (helt mislykket) til 6 (svært vellykket). Grunnlaget for vurderingen vil være både kvantitativ og kvalitativ informasjon, og følger retningslinjer utviklet i Concept-modellen (Concept-programmet 2021).



**Figur 14: Sammenstilling av vurdering av prosjektets vellykkethet**

Poengsettingen gir et kvantitativt mål på prosjekters vellykkethet og kan gi grunnlag for sammenlikning av resultater på tvers av prosjekter og sektorer.

Tabell 9 gir en oversikt over retningslinjer for hvordan poengfastsettelsen skal skje for hvert av evalueringskriteriene.

**Tabell 9: Retningslinjer for vurdering av prosjektets vellykkethet**

Evalueringskriterier	Vurdering av vellykkethet
<b>Effektivitet i prosjektgjennomføring</b> Vurdering av hvorvidt resultatmålene er realisert, og hvor effektivt ressursene er omsatt til leveranser.	<b>Score 5-6:</b> Prosjektet har levert innenfor styringsramme, tidsplan og med meget god kvalitet, og kommer i tillegg godt ut på en referansesjekk. <b>Score 3-4:</b> Prosjektet har levert innenfor kostnadsrammen, har ikke større forsinkelser, og med akseptabel kvalitet på nivå med andre sammenlignbare prosjekter. <b>Score 1-2:</b> Prosjektet har en betydelig overskridelse av kostnadsrammen (20 % eller mer) og leverer på et uakseptabelt nivå tidsmessig og kvalitetsmessig i forhold til sammenlignbare prosjekter.
<b>Realisering av effekter</b> Vurdering av i hvilken grad effektmålene er realisert, og om prosjektet har bidratt til dette.	<b>Score 5-6:</b> Prosjektet har svært god eller overoppfyllelse av effektmålene. Prosjektet fremstår som et treffsikkert virkemiddel for å realisere de planlagte effektene. <b>Score 3-4:</b> Prosjektet har akseptabel effektmåloppnåelse. <b>Score 1-2:</b> Effekter uteblir eller er klart lavere enn hva som er forventet.



Evalueringskriterier	Vurdering av vellykkethet
<b>Andre virkninger</b> Vurdering av hvorvidt prosjektet har andre virkninger utover mål-oppnåelse som resultat av prosjektet.	<b>Score 5-6:</b> Prosjektet har betydelig andre positive virkninger (utover effektmåloppnåelsen) og ingen eller kun marginale negative virkninger. <b>Score 3-4:</b> Prosjektet har få virkninger (utover effektmåloppnåelsen), og få eller ingen vesentlige negative virkninger. <b>Score 1-2:</b> Prosjektet har vesentlige negative virkninger. Lavest score gis dersom de negative virkningene er vesentlig større enn de positive virkningene av prosjektet.
<b>Relevans</b> Vurdering av hvorvidt prosjektets mål er i samsvar med brukernes og samfunnets mål og prioriteringer (over tid).	<b>Score 5-6:</b> Effekt- og samfunns mål er i samsvar med sentrale og høyt prioriterte behov, og tiltaket fører ikke til skjevprioriteringer eller konflikter mellom sentrale interessegrupper som berøres av tiltaket. <b>Score 3-4:</b> Det er akseptabelt samsvar med behov og prioriteringer. <b>Score 1-2:</b> Målene er ikke i samsvar med sentrale prioriteringer i samfunnet, og/eller fører til konflikter eller vesentlig skjevfordeling mellom sentrale interessenter.
<b>Levedyktighet</b> Vurdering av hvorvidt de positive effektene av tiltaket kan vedvare over hele tiltakets levetid.	<b>Score 5-6:</b> Prosjekter der en finner at det offentlige og sentrale interessenter både har evne og vilje til å videreføre de prosessene som prosjektet har gitt opphav til over hele levetiden. <b>Score 3-4:</b> Prosjekter der dette er usikkert. <b>Score 1-2:</b> Prosjekter der det ikke er tilfellet, og hvor sannsynligheten for at det vil skje i fremtiden er lav.
<b>Samfunnsøkonomisk effektivitet</b> Vurdering av tiltakets nytte i forhold til ressursbruken.	<b>Score 5-6:</b> Lønnsomme prosjekter ( $NNV >> 0$ ) <b>Score 3-4:</b> Lønnsomheten er lav eller nærmere null. <b>Score 1-2:</b> ulønnsomme prosjekter ( $NNV << 0$ ).

### 3.7 Trinn 5: Oppsummere resultater og anbefalinger

Resultatene fra evalueringsstudien bør forklares og presenteres i en form som beskriver hvordan evalueringen bidrar til utviklingen av ny kunnskap, samt også bidrar til å gjøre kunnskapen anvendbar for fremtidige prosjekter. Det anbefales at evalueringsrapporten som et minimum inneholder en beskrivelse av grunnlaget for beslutninger i de ulike trinnene i evalueringsmodellen:

- Beskrivelse av prosjektet med rammebetingelser, innsatsfaktorer og leveranser, inkl. styrings- og kostnadsramme, kontraktstrategi, organisering og styring av prosjektet, samt risikofaktorer
- Beskrivelse av prosjektets kontekst, omgivelser og mål, inkl. målgrupper, relevante problemstillinger og behov, samt ulike interesser som berøres av prosjektet
- Beskrivelse (og illustrasjon) av programteorien som ligger til grunn for evalueringen; prosjektets hensikt og hvordan det virker (forventning om sammenheng mellom innsats og effekter/virkninger)
- Beskrivelse av evalueringsspørsmål og indikatorer som ligger til grunn for analysene
- Beskrivelse av evalueringsdesign, metodevalg og datainnsamling, inkludert eventuelle avvik og begrensninger
- Beskrivelse av analyseteknikk og beviskrav (f.eks. statistisk signifikans), samt analyseresultater og vurdering av usikkerhet og følsomhet
- Beskrivelse av prosjektets vellykkethet (poengscore) fordelt på de ulike evalueringskriteriene
- Beskrivelse av viktige erfaringer og læringspunkter
- Overførbarhet til andre situasjoner og implementeringer

Resultatene av evalueringen kan gjerne visualiseres i form av en ny og oppdatert programteori som synliggjør hvilke sammenhenger som har blitt styrket gjennom empirisk testing, hvilke antakelse som eventuelt ikke er blitt oppfylt, og hvilke kunnskapshull som krever videre oppfølging.

Drøfting av resultatene bør eksplisitt reflektere over hvordan prosjektet påvirker **klima, miljø og naturmangfold** i et bærekraftperspektiv. Dette er et sentralt tema som gjerne behandles under flere av evalueringskriteriene; eksempelvis i form av krav til prosjektgjennomføring (reduere klimagassutslipp i byggeperioden), andre virkninger (støy, inngrep i matjord, utslipp i vassdrag), eller også i form av levetidsvurderinger (klimavennlige materialvalg). Videre må drøfting av resultatene gi grunnlag for å vurdere om prosjektet er vellykket på ulike styringsnivå:

- **Operasjonelt:** Er prosjektet effektivt? Hvordan kan eventuelle justeringer i prosjektgjennomføring eller rammebetingelser bidra til forbedring?
- **Taktisk:** Realiserer prosjektet de tiltenkte effektene for målgruppen? Har man avdekket negative virkninger? Hvilke betingelser og eksterne faktorer påvirker graden av måloppnåelse?
- **Strategisk:** Er prosjektet relevant, nyttig og levedyktig i et samfunnsperspektiv?

Evalueringsrapporten må gi en oppsummering av hvilke forbedringspunkter evalueringen har avdekket, samt beskrive hvordan dette kan brukes til forbedring av Nye Veiers arbeidsprosesser. Så langt det er mulig, bør det også gis et forslag til prioritering ut ifra forventet innsats og effekt. Et forslag til mal for rapportering av forbedringspunkter er presentert i Tabell 10.

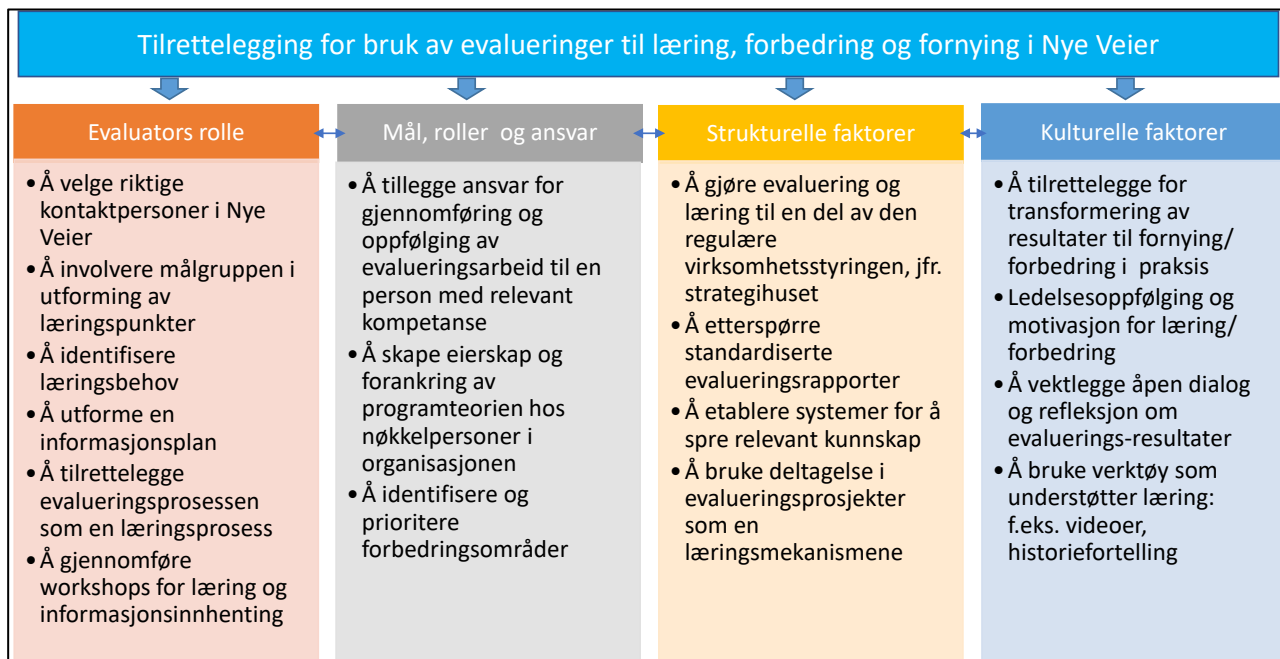
**Tabell 10: Mal for rapportering av forbedringspunkter**

Evalueringskriterier	Hva har fungert bra?	Hva kan forbedres?	Prioriterte forbedringer
Effektivitet i prosjektgjennomføringen			
Realisering av effekter			
Andre virkninger			
Relevans			
Levedyktighet			
Samfunnsøkonomisk effektivitet			

### 3.8 Trinn 6 Etterarbeid: Implementere læring

Hovedansvaret for dette etterarbeidet ligger som sagt hos Nye Veier. Grunnlaget består av både det arbeidet som er gjort i trinn 0, der utvalgte læringsområder er innsirklet, og evalueringsresultatene slik de er rapportert i trinn 5.

Dersom implementering viser seg vanskeligere enn antatt, kan utfordringen ligge i anbefalingene, særlig dersom de skulle vise seg å stride mot organisasjonens oppfatning av saksforholdet. Utfordringene kan også ligge i organisasjonen, se Figur 15 for en liste over kulturelle, strukturelle og organisatoriske forhold som kan fremme læring (eller hindre den dersom de ikke er tilstrekkelig godt til stede). Implementeringen kan også vise seg å være for kostbar å implementere i forhold til nytten av å gjøre det. Uansett årsak kan det være fornuftig å involvere evaluator i en diskusjon dersom implementering viser seg å være vanskelig.



**Figur 15: Tilrettelegging for bruk av evaluering til forbedring og fornying i Nye Veier**

Det er viktig å understreke at faktorene i ulik grad påvirker hverandre og spiller sammen innen den aktuelle kontekst, og derfor bør vurderes i en helhetlig sammenheng. Under forklares de fire hovedkategorier av faktorer på et overordnet nivå.

Evaluators rolle er sentral som tilrettelegger av evalueringsprosessen som en læreprosess. Det er et betydelig potensial for læring gjennom involvering, aktiv deltagelse og samhandling mellom evaluator og virksomheten. Dette gjelder for etterevalueringer, men særlig for følgeevalueringer, som har et skreddersydd design for kontinuerlig læring og forbedring ved gjennomføring av prosjekter (faser av prosjekter). Ved etterevalueringer er bl.a. deltagelser i workshops egnet for erfaringsutveksling, refleksjon og læring.

Mål, roller og ansvar: Det er viktig at virksomheten (Nye Veier) definerer hvem i egen organisasjon som har ansvar for forarbeidet (formål og omfang), oppfølging av gjennomføring og ikke minst av evalueringresultatene for å sikre intern læring av evalueringer. Det anbefales at en stabsenhet (for eksempel Virksomhetsstyring) får et tydelig definert ansvar for dette, siden den vil kunne utvikle og bidra med nyttig kompetanse på tvers av prosjekter og enheter i Nye Veier.

Strukturelle faktorer er de tiltak og systemer som virksomheten (Nye veier) kan designe og bestemme seg for å innføre, for å støtte opp under bruk av evalueringresultater, læring og forbedringsarbeid. Ofte er det nødvendig å etablere prosedyrer og rutiner som sikrer at viktige tema får tilstrekkelig oppmerksomhet, og at de inngår i organisasjonens beslutningsstrukturer. Det anbefales at sentrale resultater fra evalueringer blir en integrert del av Nye Veiers virksomhetsstyring, for å sikre ledelsesoppfølging, identifisering og forankring av forbedringstiltak. Læring og forbedring kan tas inn som et tema i virksomhetens strategiprosesser, og hvor anvendelse av resultater fra evalueringer inngår som en KPI i Nye Veiers strategihus.

Kulturelle faktorer er menneskelige og organisatoriske forhold som utvikles ved påvirkning og læring, og som forutsetter deltagelse og involvering av ansatte. Dette omfatter kunnskaper, ferdig-



heter og holdninger, som ligger til grunn for handlinger og praksiser, og ikke minst holdninger til og motivasjon for læring og forbedring. Kontinuerlig læring refererer også til begrepet lærende organisasjon, som kan forstås som en organisasjon som tilrettelegger for læring for alle medlemmer, og som gjennom det evner å transformere seg kontinuerlig. Det betyr at Nye Veier må tilrettelegge for læring for de ansatte som forventes å være involvert i utvikling og forbedring basert på evalueringresultatene. Kollektive læreprosesser kan skje på to måter: ved prosesser som karakteriseres som enkelt-sløyfet læring, og som handler om effektivitet (gjør vi ting riktig?), og som dobbelt-sløyfet læring, som handler om endring og innovasjon (gjør vi de riktige tingene?).

Når det gjelder forutsetninger for implementering av forbedringspunkter er dette noe evaluator må ta opp med den som har ansvaret for å følge opp evalueringresultatene i Nye Veier. Det anbefales å innføre en rutine for evaluator og Nye Veier, hvor det legges opp til at evaluator og ansvarlig i Nye Veier gjennomfører en dialog med den som skal følge opp og ta ansvar for implementeringen i virksomheten.

## 4 Evalueringsmodellens styrker og svakheter

### 4.1 Viktige egenskaper ved evalueringsmodellen

Utarbeiding av evalueringsmodellen har vært underlagt en del føringer (mål, krav og forventninger) fra oppdragsgiver som vi i det ovenstående har avveid mot faglige og ressursmessige muligheter og begrensninger. Modellen har dermed fått en del trekk (eller **egenskaper**), som gjør at den

- kan brukes på enkeltprosesser og enkeltfaser i Nye Veiers gjennomførings- og eierstyringsmodell og på ferdigstilte prosjekter og strekninger (den er **fasenøytral**)
- kan i en viss grad brukes til både etterevaluering, underveisevaluering<sup>26</sup> og forhåndsevaluering (den er **tidspunktnøytral**)
- omfatter planlegging og gjennomføring av evalueringer og bidrar til tilrettelegging av selskapets (og sektorens?) læring fra dem (den er **læringsorientert**)
- gir rom for perspektivene til Nye Veier som uttrykk for nasjonale politiske beslutninger, for perspektivene til ulike interessentgrupperinger, for perspektivene til lokaldemokratiske organer, og for ulike faglige perspektiver (den er **multi-perspektivert**)
- spesifiserer hvilke data og annen dokumentasjon som Nye Veier bør besørge registrert og arkivert underveis for å muliggjøre effektive evalueringer (den er **virksomhetsdata-orientert**)
- prioriterer evalueringsaktiviteter etter deres nyttebidrag til evalueringsformål og læringsutbytte (den er **kostnadseffektiv**)
- er tilstrekkelig standardisert til å kunne brukes til å sammenlikne mellom Nye Veiers egne prosjekter og mellom Nye Veiers og Concepts etterevalueringer av vegprosjekter (den er **standardisert** og **kompatibel** med Concept-modellen)
- er allikevel tilstrekkelig fleksibel til å ivareta de enkelte evalueringobjektene særpreg og spesielle formål med de enkelte evalueringprosjektene (den er **formåls-fleksibel**)<sup>27</sup>
- er dokumentert slik at den kan benyttes av kvalifiserte evalueringsteam (den er **dokumentert**)
- kan etableres som et standardgrunnlag for evalueringstjenester lagt inn som opsjon ved flere beslutningspunkter i Nye Veiers eierstyringsmodell (den er **tjeneste-standardisert**)
- har mange elementer som kan bygges inn i Nye Veiers interne mål- og resultatstyringssystem og også brukes med hell til intern læring og utvikling, dog uten at intern bruk av disse elementene kan forventes å få den samme troverdighet som en ekstern og uavhengig evaluering (dens metoderepertoar er **internaliserbart**)
- lar seg tilpasse for andre byggherrer i vegsektoren (den er **klientuavhengig**).

Disse egenskapene lar seg ikke nødvendigvis kombinere like godt i alle enkeltevalueringer som gjennomføres etter modellen. Formålet med den enkelte evaluering og dens bidrag til den akkumulerte kunnskapsbasen bør derfor avgjøre prioriteringen mellom disse egenskapene, dersom det er aktuelt. I utgangspunktet er imidlertid modellen først og fremst tilpasset – og gjennom Concepts evalueringssportefølje validert for – etterevaluering av avsluttede vegprosjekter der vegen har vært i drift en tid.

<sup>26</sup> Underveisevaluering kan skille seg fra følgeevaluering, som ofte er mye mer interaktiv. Evalueringsmodellen dekker ikke interaktiv følgeevaluering og følgeforskning, selv om mange elementer kan gjenfinnes også der. Se for øvrig kapittel 4.2.

<sup>27</sup> Se allikevel kapittel 4.2 for avgrensninger mot beslektede former for kunnskapsgenerering.



## 4.2 Hva modellen er, og ikke er, egnet til

Denne evalueringsmodellen er i utgangspunktet beregnet på etterevaluering av Nye Veiers arbeid med vegstrekninger som de har ansvar for. Den er også i noen grad tilpasset underveisevaluering av dette arbeidet. Dette omfatter også muligheten til å gjøre en vurdering av fremtidige prosjektutfall og sammenlikne med det som ble planlagt av Nye Veier selv på et tidlig stadium.

Slike evalueringer overlapper i noen grad med aktiviteter som evalueringsmodellen ikke er egnet til å bruke som faglig grunnlag for. De viktigste av disse er vist i Tabell 11, som viser hvordan evalueringsmodellen overlapper med andre aktiviteter for kunnskapsgenerering, hvor ulikhetene er mellom dem, og hvordan vedkommende aktivitet mottar bidrag fra evalueringsmodellen og vice versa. Verken aktivitetslisten eller de enkelte egenskaper er uttømmende.

**Tabell 11: Evalueringsmodellens forhold til beslektede aktiviteter**

Beslektet aktivitet	Overlapp	Ulikhet (aktivitet vs modellen)	Bidrag fra modellen	Bidrag til modellen
Evaluering av Nye Veiers oppfylging av sitt samfunnsoppdrag	Tematisk	Hele virksomheten vs enkeltprosjekter	Data	Organisasjonens kultur og struktur som kontekstuell forklaring Politiske prioriteringer og vurderinger som kontekstuell forklaring
Evaluering av Nye Veiers arbeid med hele porteføljer	Tematisk og kriteriemessig	Porteføljenivå vs strekningsnivå	Strekningsvise evalueringer	Porteføljestyling som kontekst for å forstå relativ prioritering av fremdrift osv. på enkeltstrekninger
Evaluering av organisasjonen Nye Veier	NV som "produksjonsmekanisme" for veier	NV som "reproduksjonsmekanisme" for seg selv	Evalueringsresultater	Organisasjonens kultur og struktur som kontekstuell forklaring
Følgforskning av enkeltprosesser	Tematisk	Betydelig ulikhet i interaksjon gir store metodiske forskjeller	Kriteriene	Utdyping av enkeltproblemstillinger og betydning av kontekst
Forskning på ulike prosessers bidrag til transportsektoren og dens samfunnsnytte	Fokus på årsaksforklaringer	Bredere tilfang vs enkeltprosjekter	Data fra enkeltprosjekter	Problemstillinger verdt å undersøke i evaluering
Ex ante-evaluering av nye tiltak	Tematisk	Uten vs med empiri	Årsaksforklaringer	Målformuleringer, målstruktur
Eksperimentell utprøving av nye prosesser og løsninger	Tematisk	Kontrollert eksperiment vs naturlige case	Evalueringskriterier	Velfunderte endringsmekanismer
Kvalitetssikring mot etablerte normer	Tematisk	Kriterier, metode, formål	Nye usikkerhetsmomenter	Forhåndsundersøkelse av viktige felt



Beslektet aktivitet	Overlapp	Ulikhet (aktivitet vs modellen)	Bidrag fra modellen	Bidrag til modellen
Evaluering av vegsektorens utvikling	Tematisk	Sektor vs enkeltprosjekter	Nærbilder av vegbygging	Viktig kontekstforståelse
Selskapets egenvurdering av egne meritter	Tematisk, delvis metode	Egenselektiv vs uavhengig utvalgs-kriterium	Strukturerte datakilder og årsaksforklaringer	Nærhetsforståelse

## 5 Litteraturreferanser

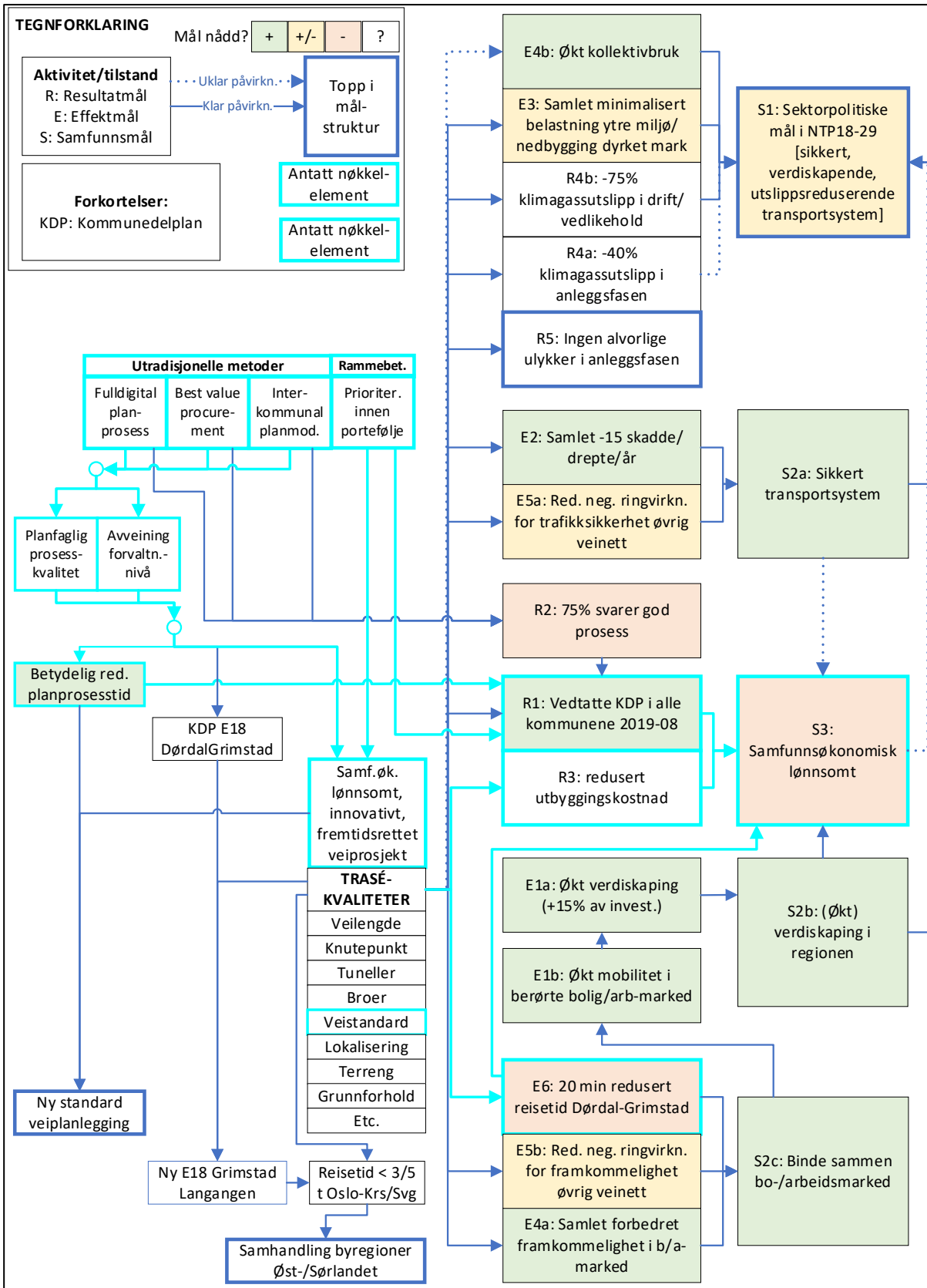
- Andersen, Bjørn og Tom Fagerhaug (2006): *Root cause analysis : simplified tools and techniques*. Milwaukee WI: ASQ Quality Press.
- Argyris, Chris og Donald A Schön (1978): *Organizational learning. A theory of action perspective*. Reading MA: Addison-Wesley.
- Bredgaard, Thomas, Julia Salado-Rasmussen og Stella Mia Sieling-Monas (2016): "Virkningsevaluering." I Thomas Bredgaard (red.) *Evaluering af offentlig politik og administration*: Hans Reitzels Forlag, s. 331-366.
- Bukkestein, Ingri, Morten Welde og Gro Holst Volden (2020): *Bruk og nytte av etterevalueringer av prosjekter*. Concept arbeidsrapport 2020-03, Trondheim: NTNU Concept.
- Chen, Huey-Tsyh (1990): *Theory-driven evaluations*. Newbury Park CA: Sage.
- Concept-programmet (2021): *Etterevaluering av statlige investeringsprosjekter. Retningslinjer for evaluator*. Trondheim: NTNU Concept.
- FIN (2010): *Kvalitetssikring av konseptvalg, samt styringsunderlag og kostnadsoverslag for valgt prosjektalternativ. Målstruktur og målformulering. Veileder nr. 10*. Oslo: Finansdepartementet.  
[https://www.regjeringen.no/contentassets/d16417cc990c457db3fb2965dcdcf3abe/veileder\\_nr10\\_malstruktur\\_og\\_malformulering.pdf](https://www.regjeringen.no/contentassets/d16417cc990c457db3fb2965dcdcf3abe/veileder_nr10_malstruktur_og_malformulering.pdf)
- FIN (2014): *Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv*. Rundskriv R 109/14, Oslo: Finansdepartementet.  
[http://www.dfo.no/Documents/FOA/publikasjoner/rapporter/r\\_109\\_2014.pdf](http://www.dfo.no/Documents/FOA/publikasjoner/rapporter/r_109_2014.pdf)
- FIN (2019): *Statens prosjektmodell - Krav til utredning, planlegging og kvalitetssikring av store investeringsprosjekter i staten*. Rundskriv R 108/19, Oslo: Finansdepartementet.  
[https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/rundskriv/faste/r\\_108\\_2019.pdf](https://www.regjeringen.no/globalassets/upload/fin/vedlegg/okstyring/rundskriv/faste/r_108_2019.pdf)
- FIN (2021): "Hva er statens prosjektmodell?" <https://www.regjeringen.no/no/tema/okonomi-og-budsjett/statlig-okonomistyring/ekstern-kvalitetssikring2/hva-er-ks-ordningen/id2523897/>
- Finne, Håkon (2006): *Trøndelags innovasjonserilja? Evaluering av Innovasjon Midt-Norge*. Trondheim: SINTEF Teknologi og samfunn, International Operations.
- Finne, Håkon, Morten Levin og Tore Nilssen (1995): "Trailing research. A model for useful program evaluation." *Evaluation* Vol. 1 No. 1, s. 11-31.
- Funnell, Sue C og Patricia J Rogers (2011): *Purposeful program theory: effective use of theories of change and logic models*. San Francisco CA: Jossey-Bass: XXVI, 550 s.  
<https://read.amazon.com/?asin=B004NSW9G8>
- HM Treasury (2020): *Magenta Book : Central Government Guidance on Evaluation*. London: HM Treasury.
- Ishikawa, Kaoru (1990): *Introduction to quality control*. London: Chapman & Hall.
- Johansson, Kersti, Verner Denvall og Evert Oskar Vedung (2015): "After the NPM wave. Evidence-Based practice and the vanishing client." *Scandinavian Journal of Public Administration* Vol. 19, s. 69-88.
- Klakegg, Ole Jonny (2006): *Målformulering i store statlige investeringsprosjekt*. Trondheim: Concept-programmet.
- Lervåg, Lone-Eirin (2021): *Evaluering av intelligente transportsystemer. Utvikling av evalueringsmetodikk basert på tiltakets programteori*. Trondheim: NTNU.
- Meld. St. 20 (2020-2021) (2021): *Nasjonal transportplan 2022-2033*. Melding til Stortinget 20 (2020-2021), Oslo: Departementenes sikkerhets- og serviceorganisasjon.
- Meld. St. 33 (2016-2017) (2017): *Nasjonal transportplan 2018-2029*. Melding til Stortinget 33 (2016-2017), Oslo: Departementenes sikkerhets- og serviceorganisasjon.



<https://www.regjeringen.no/contentassets/7c52fd2938ca42209e4286fe86bb28bd/no/pdfs/stm201620170033000dddpdfs.pdf>

- Mohr, Lawrence B (1995): *Impact analysis for program evaluation*. Thousand Oaks CA, London, New Dehli: Sage Publications.
- OECD (1991): *DAC principles for evaluation of development assistance*. Paris: OECD.  
<http://www.oecd.org/dac/evaluation/2755284.pdf>
- OECD DAC Working Party on Aid Evaluation (1998): *Review of the Dac Principles for Evaluation of Development Assistance*. Paris: Organisation for Economic Co-operation and Development.
- OECD/DAC Network on Development Evaluation (2019): *Better criteria for better evaluation. Revised evaluation criteria. Definitions and principles for use*. Paris: OECD.
- Olsen, Odd Einar og Preben Lindøe (2004): "Trailing research based evaluation; phases and roles." *Evaluation and Program Planning* Vol. 27 No. 4, s. 371-380.
- Pawson, Ray og Nick Tilley (1997a): "An Introduction to Scientific Realist Evaluation." I Eleanor Chelimsky og William R Shadish (red.): *Evaluation for the 21st Century: A Handbook*. Thousand Oaks CA: SAGE Publications, s. 405-418.
- Pawson, Ray og Nick Tilley (1997b): *Realistic evaluation*. London: Sage Publications.
- Ramstad, Lone Sletbakk, Morten Welde, Cecilie Flyen, Håkon Finne og Bjørn Andersen (2020): *Følgforskning av planprosjektet E18 Dørdal - Grimstad*. Trondheim: NTNU Concept.
- Rosenberg, Leon J, Lawrence D Postner og Edward J Hanley (1970): *Project evaluation and the project appraisal reporting system*. Washington DC: Fry Consultants.  
<https://www.alnap.org/system/files/content/resource/files/main/pnadw881.pdf>
- Samset, Knut Fredrik (2014): *I riktig retning : prosjekters mål og målstruktur*. Trondheim: NTNU Concept.
- van der Knaap, Peter (2004): "Theory-based evaluation and learning: Possibilities and challenges." *Evaluation* Vol. 10 No. 1, s. 16-34.
- Volden, Gro Holst og Bjørn Andersen (red.) (2019): *Ekstern kvalitetssikring av investeringsprosjekter i statlige foretak og statsaksjeselskaper. En gjennomgang av prosjektmodeller og kvalitetssikring for de regionale helseforetakene, Statnett (OED), Avinor, Bane NOR og Nye Veier*. Trondheim: NTNU Concept.
- Volden, Gro Holst og Knut Samset (2013): *Etterevaluering av statlige investeringsprosjekter: konklusjoner, erfaringer og råd basert på pilotevaluering av fire prosjekter*. Trondheim: NTNU Concept.

## Vedlegg



Figur 16: Programteori for forsert kommunedelplanarbeid



I figuren (fra Ramstad mfl. 2020) er formulerte samfunns mål (S), effektmål (E) og resultatmål (R) brutt ned og satt sammen med andre mål oppdaget underveis, og med de metoder som skulle prøves ut, i et kart der kausale mekanismer er markert med piler. Forventet eller faktisk måloppnåelse er markert ved trafikklysfargene på de respektive boksene. Påvirkningsveien fra de metodene som skulle prøves ut, til de mest relevante målene, er markert i turkis. På evalueringstidspunktet var målet om betydelig kortere kalendertid brukt på kommunedelplan (Betydelig red. planprosessetid & R1) nådd, og vektleggingen i denne prosessen på samfunnsøkonomisk lønnsomhet framfor å tilgodese lokale behov (Rammebet.), hadde ført til en trasé med et bedre kost-/nytte-forhold (en boks som mangler i figuren), men ikke tilstrekkelig til å gjøre prosjektet samfunnsøkonomisk lønnsomt (S3). Det lå en usikkerhet (ikke inntegnet i figuren) i forhold til senere innsigelser mot traséen, til tross for at man hadde gått bredt ut for å få mulige innsigelser på plass i løpet av den tidsforserte prosessen. Senere viste det seg at nettopp denne usikkerhetsfaktoren slo inn og gjorde det nødvendig å gjøre om noen planer, med ytterligere kostnadsøkning som konsekvens. Men analysen hadde ikke i dette tilfellet gjort noen scenarioanalyse av dette; hovedfokus var at analysen gjorde det enklere å undersøke hvordan momentene markert i turkis faktisk fungerte og hvorvidt de i seg selv bidro til effekter og andre virkninger.