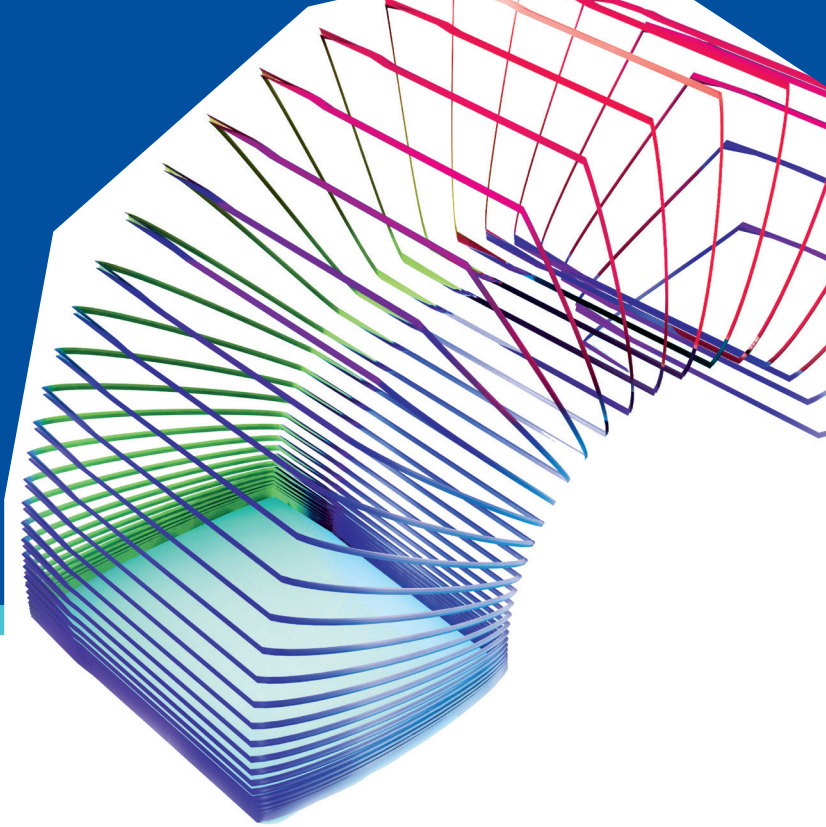




Lone-Eirin Lervåg

Teoribasert evaluering

Rammeverk for evaluering av innovative
og komplekse tiltak



CONCEPT TEMAHEFTE **NR. 18**

Om forfatteren:

Lone-Eirin Lervåg er seniorforsker i SINTEF ved avdeling Mobilitet og samfunnsøkonomi. Hun har 20 års erfaring med forskningsoppdrag knyttet til utprøving, implementering og evaluering av innovative mobilitetsløsninger i transportsystemet. Hun fikk tildelt en doktorgrad fra NTNU i april 2021, etter å ha forsvart avhandlingen «*Evaluering av intelligente transportsystemer. Utvikling av evalueringsmetodikk basert på tiltakets programteori*». Målsettingen med dette forskningsarbeidet har vært å bidra til bedre forståelse og kunnskap om et egnet metodegrunnlag for evaluering av intelligente transportsystemer (ITS). Problemstillingen er studert gjennom innledende litteraturstudier og med praktisk utprøving av en teoribasert evalueringsmetodikk i to konkrete case-studier. Erfaringene fra dette arbeidet gir grunn til å tro at en teoribasert tilnærming til ITS-evalueringer vil bidra til å skape et mer helhetlig kunnskapsgrunnlag og en bedre utnyttelse av evalueringsressurser innenfor fagfeltet.

Kontakt: lone.lervag@sintef.no

Rettighetene til innholdet i dette skrevet er forfatterens. Heftet er utgitt med støtte fra forskningsprogrammet Concept.

Adresse:

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

ISBN: 978-82-8433-014-3 (papirversjon)

ISBN: 978-82-8433-015-0 (nettversjon)

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



Om doktorgradsarbeidet

Doktorgradsavhandlingen (Lervåg, 2021) gjør rede for det metodiske grunnlaget for evaluering av IKT-baserte løsninger i transportsystemet, såkalte *intelligente transportsystemer* (ITS). Formålet med forskningsarbeidet har vært å undersøke om det finnes evalueringsmodeller som kan øke nytten av ITS-evalueringer i forhold til dagens praksis. Første del av avhandlingen bygger på litteraturstudier som gir innblikk i relevante evalueringsperspektiver, metodiske styrker og svakheter, samt tradisjoner for kunnskapsbygging innen ulike fagfelt. Dette arbeidet har gitt grunnlag for å rette søkelyset mot en *teoribasert tilnærming* til evaluering av ITS-tiltak. Den teoribaserte fremgangsmåten adresserer flere typiske utfordringer med ITS-evalueringer: Den kan håndtere sammensatte og komplekse tiltak, den inkluderer situasjonsavhengige faktorer (kontekst) i evalueringen, og den åpner for å kombinere flere metoder og analyseteknikker i situasjoner med begrenset datagrunnlag. Problemstillingen er utforsket med praktisk utprøving av en teoribasert evalueringsmetodikk i to konkrete case-studier. Erfaringene fra dette arbeidet viser at fremgangsmåten er anvendelig og kan bidra til å styrke kunnskapsgrunnlaget for implementering av innovative og komplekse virkemidler i det moderne transportsystemet.

Hovedresultatet fra dette doktorgradsarbeidet er et rammeverk for organisering og gjennomføring av en teoribasert evalueringsprosess, som også kan være relevant utover ITS-tiltak. Et viktig element i denne fremgangsmåten, er beskrivelsen av tiltakets *programteori*, som uttrykker logikken for innføringen av tiltaket – altså hvordan tiltaket skal føre til et ønsket resultat. Programteorien danner utgangspunkt for videre utforming av relevante forskningsspørsmål og evalueringskriterier. Rammeverket gir et vitenskapelig grunnlag for hensiktsmessige metodevalg og analyser, som legger til rette for relevant kunnskapsbygging gjennom hele beslutningsprosessen, fra tidlig konseptutvikling til tiltaket er implementert i ordinær drift.

Innenfor transportsektoren finnes det få eksempler på eksplisitt bruk av teoribasert metodikk i evalueringsstudier. Bruk av rammeverket gir transportmyndighetene større mulighet til å evaluere teknologi på et lavere modenhetsnivå, samt bedre innsikt i mekanismene som styrer hvorvidt man lykkes eller ikke med iverksettelsen av nye ITS-tiltak. Tidlig involvering og aktiv tilrettelegging av rammebetingelser kan bidra til å fremskynde innføringen av bærekraftige og fremtidsrettede tjenester. Samtidig kan man unngå tapsinvesteringer, dersom feilslått teknologi velges bort på et tidligere tidspunkt i innovasjonsløpet. På denne måten kan en teoribasert evalueringstilnærming bidra til å skape mer helhetlig kunnskap og en bedre utnyttelse av evalueringsressurser.

Innhold

1. Innledning	1
Bakgrunn og begrepsforklaring	1
Hvordan evaluere innovative og komplekse tiltak?	1
2. Hva er teoribasert evaluering?	4
Utvikling av tiltakets programteori	4
Teoribaserte evalueringsdesign	8
Sammenfatning av det teoribaserte evalueringsperspektivet	12
3. Når er teoribasert evaluering særlig egnet?	14
4. Erfaringer med en teoribasert fremgangsmåte for evaluering av intelligente transportsystemer (ITS)	17
5. Rammeverk for en teoribasert evalueringsprosess	21
Trinn 1: Tiltakets teknologiske modenhet	21
Trinn 2: Behovsanalyse	22
Trinn 3: Programteori	23
Trinn 4: Evalueringskriterier	25
Trinn 5: Evalueringsdesign	25
Trinn 6: Datainnsamling	28
Trinn 7: Analyser	29
Trinn 8: Rapportering av resultater	30
Referanser	32

1. Innledning

Bakgrunn og begrepsforklaring

I dette temaheftet presenteres fremgangsmåten i en teoribasert evalueringstilnærming. Begrepet *teoribasert evaluering* er i evalueringslitteraturen en etablert benevnelse på en evalueringsmodell som innebærer at evalueringen er strukturert rundt tiltakets *programteori*¹ og med en *mekanistisk forståelse av kausalitet* (årsakssammenheng). Ordet *teori* henspiller i denne sammenhengen på tiltakets underliggende logikk og ikke på en overordnet teoretisk forankring.

Konseptet med bruk av programteori eller årsaksvirkningskjeder bygger på en lang evalueringstradisjon, som gjerne omtales med begreper som *resultatkjeder*, *målstruktur* eller *logiske rammeverk*. Det som skiller den teoribaserte evalueringen fra øvrige tilnærminger, er at tiltakets kontekst og underliggende mekanismer trekkes inn som en viktig del av programteorien. I dette ligger en erkjennelse av at virkningene av et tiltak eller program ikke nødvendigvis er allmenngyldige, men kan oppstå i spesifikke (fordelaktige) situasjoner, og for spesifikke deltakere eller målgrupper.

Doktorgradsarbeidet som dette temaheftet bygger på, har fokusert på metodelitteratur og erfaringer med evaluering av *intelligente transportsystemer*. Det teoribaserte evalueringsperspektivet synes imidlertid å være relevant også for innovative, komplekse tiltak innenfor andre samfunnsområder. Teoribasert evaluering er også omtalt i Concept-rapport 52 (Volden og Samset, 2017), hvor teoribaserte evalueringsdesign anbefales som supplement til empiriske effektevalueringer, og hvor det påpekes at fremgangsmåten sjelden benyttes i dag.

Hvordan evaluere innovative og komplekse tiltak?

Digitalisering og teknologisk innovasjon forventes å spille en vesentlig rolle i utviklingen av et bærekraftig og fremtidsrettet samfunn. For å realisere ønskede virkninger, er det nødvendig å bygge kunnskap som kan fremskynde innføring av nyttig og lønnsom teknologi, samtidig som man sikrer at investeringene kommer samfunnet og befolkningen til gode. En viktig del av dette kunnskapsgrunnlaget fremskaffes gjennom effektstudier.

Tradisjonelt er idealet for transportfaglige effektstudier basert på et (kvasi-) eksperimentelt metodeoppsett med statistisk analyse av effekter før og etter innføring av tiltaket. Samtidig skal resultatene sammenlignes med en kontrollgruppe, for å dokumentere kausalitet (kontrafaktisk analyse). Når de metodiske betingelsene er oppfylt, gir slike studier et sterkt vitenskapelig effektbevis.

Dette er imidlertid vanskelig å oppnå i praksis, så det gjøres som regel en rekke forenklinger i metodeoppsettet. De fleste evalueringsstudier har begrenset tilgang til data,

¹ Tiltakets programteori uttrykker den overordnede logikken for innføringen av et tiltak, og beskriver hvordan et sett mer virkemidler (tiltaket) forventes å føre til et ønsket resultat.

og man ender gjerne opp med varianter av før- og etterundersøkelser med ulik grad av kontroll. Med stadig økende digitalisering og teknologisk innovasjon utfordres evalueringsarbeidet ytterligere:

- Mange innovative tiltak er gjenstand for kontinuerlig teknologiutvikling som gjør at løsningen forandrer seg raskt og fortløpende gjennom sin levetid. I praksis betyr dette at tiltaket ofte innføres gradvis og stadig forbedres, noe som gjør det vanskelig å definere klare før- og etter-perioder i en evalueringsstudie.
- Innovative tiltak består gjerne av sammensatte systemer og virkemidler, og ofte med kompliserte årsakssammenhenger og et situasjonsavhengig effektbilde som kan være vanskelig å fange opp med tradisjonell evalueringsmetodikk.

Dette medfører i mange tilfeller at man sitter igjen med evalueringsresultater som er beheftet med stor grad av usikkerhet, og som ikke er gyldige over tid eller mellom ulike installasjoner.

Urban, Hargraves og Trochim (2014) argumenterer for at evalueringer av komplekse, sosiale tiltak må inneha et evolusjonsperspektiv som tar høyde for at både tiltakene og omgivelsene rundt er i stadig endring. Tiltaket er dermed ikke en statisk størrelse som skal bevises for å oppdage "den perfekte tilstanden", men en del av en livssyklus med prøving, feiling og stadig forbedring. Som en følge av dette er det urealistisk å forvente at tiltaket skal fungere på et modent nivå fra start, og det argumenteres for at metodevalg og evalueringsdesign må baseres på hvor tiltaket er i sin utvikling. Når evolusjonsperspektivet legges til grunn for evalueringen, blir konsekvensen at evalueringsmetodikken må kunne inkludere flere former for validitet og kausalitetsteori. Marchal, Van Belle, De Brouwere og Witter (2013) argumenterer også for at forskning på komplekse intervensjoner krever et dynamisk forskningsdesign, som kan tilpasses etter hvert som ny kunnskap og uforutsette problemstillinger avdekkes.

Det teoribaserte evalueringsperspektivet adresserer flere typiske utfordringer med evaluering av innovative tiltak: Det kan håndtere sammensatte og komplekse systemer, det inkluderer kontekst som en relevant faktor i evalueringen, og det åpner for å kombinere flere metoder og analyseteknikker. Dette forutsetter imidlertid god kjennskap til de ulike metodenes styrker og svakheter, samt kompetanse til å vurdere resultatenes gyldighet og overførbarhet til andre situasjoner og implementeringer.

Teoribaserte evalueringer skiller seg fra klassiske effektstudier på to vesentlige områder:

- (1) Kausalitet betraktes som en iboende (latent) egenskap som kan aktiveres under gitte omstendigheter – det vil si at tiltaket øker sannsynligheten for at en virkning inntreffer, men at dette også avhenger av andre forhold (som rammebetingelser og kontekst). Kausalitet bestemmes derfor gjennom studier av årsaksvirkningskjeder som grunnlag for å vurdere i hvor stor grad ulike faktorer bidrar til det observerte utfallet. I denne mekanistiske forståelsen av kausalitetsbegrepet legges det vekt på at effektene som observeres er transfaktuelle (langvarige, ikke forbigående) istedenfor kontrafaktiske.

- (2) Kontekst (rammebetingelser, suksesskriterier og barrierer) inkluderes som en relevant faktor i evalueringen. I praksis innebærer dette at den teoribaserte evalueringsprosessen har stor forklaringskraft med hensyn til å vurdere hvilken kombinasjon av innsats og virkemidler som gir ønsket effekt. Dette er relevant kunnskap både for den som skal ta strategiske beslutninger om utvikling av fremtidige løsninger, og for den som skal implementere tiltakene i praksis.

Det teoretiske grunnlaget for disse forholdene er gjort nærmere rede for i kapittel 2.

2. Hva er teoribasert evaluering?

Den teoribaserte evalueringen har et systemperspektiv som innebærer at tiltaket settes inn i en kontekst med rammebetingelser og eksterne faktorer som påvirker hvorvidt man lykkes med iverksettelsen. Evalueringen struktureres rundt oppbygging og testing av tiltakets *programteori*, som beskriver hvordan tiltaket forventes å utløse effekter og virkninger i en gitt kontekst. På denne måten skiller den teoribaserte fremgangsmåten seg fra klassiske effektevalueringer, ved at den ikke bare skal gi svar på *om* et tiltak virker eller ikke, men også *hvorfor* og *hvordan* det virker.

Etablering av programteorien, danner utgangspunktet for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for modellen. Når evalueringsresultatene systematisk pløyes tilbake i programteorien, vil man både synliggjøre kunnskapshull (forskningsbehov) og bidra til en helhetlig forståelse av tiltakets potensiale.

Den teoribaserte tilnærmingen gir ikke føringer for spesifikt valg av analyseteknikk, men forutsetter at datainnsamling og analyser defineres på bakgrunn av hvilke forskningsspørsmål som ønskes besvart og praktisk mulighet for datainnsamling. Den teoribaserte evalueringsprosessen kan i praksis kombineres med de fleste analyser og forskningsdesign, og legger dermed til rette for fleksibilitet i metodevalg og datainnsamling. Dette innebærer for eksempel at en overordnet teoribasert evalueringsprosess også kan inkludere eksperimentelle analyser av enkelte variabler, dersom forholdene (datatilgang) ligger til rette for dette.

Utvikling av tiltakets programteori

Den teoribaserte evalueringstilnærmingen tar utgangspunkt i tiltakets *programteori* - en konseptuell modell av sammenhengen mellom innsatsen i tiltaket og de virkninger man forventer å oppstå på kort og lang sikt (Chen, 1990, 2005; Funnel og Rogers, 2011; Rossi, Lipsey og Freeman; Weiss, 1995). Programteorien² representerer tiltakets underliggende logikk: forventningen om hvordan og hvorfor planlagte virkemidler skal føre til ønsket resultat - altså den rasjonelle grunnen for iverksettelsen av et tiltak.

Programteorien etableres med utgangspunkt i prosjekteiers forventninger om effekter, men kan også involvere andre interessegruppers forventninger og antakelse. Det forutsettes videre at effektbildet inkluderer både ønskede og utilsiktede virkninger. Figur 1 viser en skjematisk fremstilling av en enkel lineær programteori. Når tiltakets kompleksitet og evalueringens detaljfokus tiltar, benyttes gjerne mer detaljerte modeller som også kan fremstilles hierarkisk eller på matriseform.

² Bruk av ordet *programteori* kan være litt misvisende, siden det er viser til tiltakets interne logikk, og ikke er knyttet til en overordnet vitenskapsteoretisk forankring.



Figur 1: Skjematisk fremstilling av programteorien; den underliggende logikken bak iverksettelsen av et tiltak.

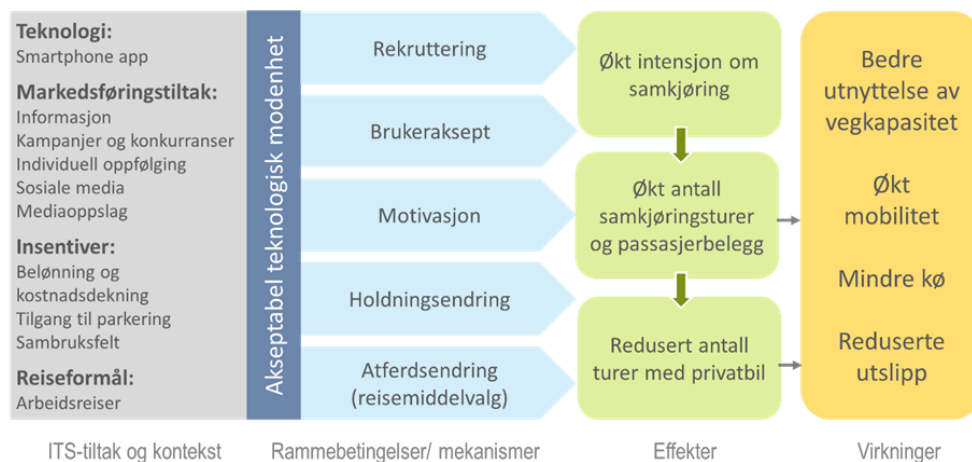
Programteorien er en sammenstilling av tiltakets endringsteori og implementeringsteori, som begge har betydning for tiltakets endelige utfall:

- *Tiltakets implementeringsteori:* En preskriptiv modell av hvilke betingelser (konkrete handlinger og prosesser) som må være oppfylt for at tiltaket skal avstedkomme en ønsket endring. Dette er illustrert med boksene i Figur 1. Implementeringsteorien viser hvordan prosjektet er konstruert med et logisk resonnement for å aktivere endringene i neste punkt.
- *Tiltakets endringsteori:* En deskriptiv modell av de underliggende mekanismene og kausale prosessene som må inntreffe for at endring skal oppnås. Dette er illustrert med pilene i figuren. Programteorien kan inneha ulike endringsteorier for ulike faser av tiltaket og for ulike målgrupper i prosjektet. Disse er gjerne tuftet på samfunnsvitenskapelige teorier (f.eks. psykologiske, sosiale eller økonomiske modeller).

Utforming av programteorien vil synliggjøre avstand i verdensbildet mellom de involverte aktørene, og til en viss grad også gi grunnlag for å oppnå felles forståelse basert på eksisterende kunnskap. Visualisering av programteorien gir videre en felles referanseramme for dialogen i prosjektet, og kan også være et nyttig instrument i kommunikasjon mot omverden.

Eksempel 1: Programteori for ITS-tiltaket «Samkjøring»

Figur 2 viser eksempel på en programteori for ITS-tiltaket *Samkjøring med turkobling via en mobilapplikasjon*, utviklet i evalueringsprosjekt på oppdrag fra Vegdirektoratet og presentert i Lervåg (2021). Formålet med evalueringsarbeidet var å avdekke hvorvidt samkjøring har potensiale som etterspørselsstyrende og trafikkdempende tiltak i Norge, med utgangspunkt i en pilotstudie i Bergen.



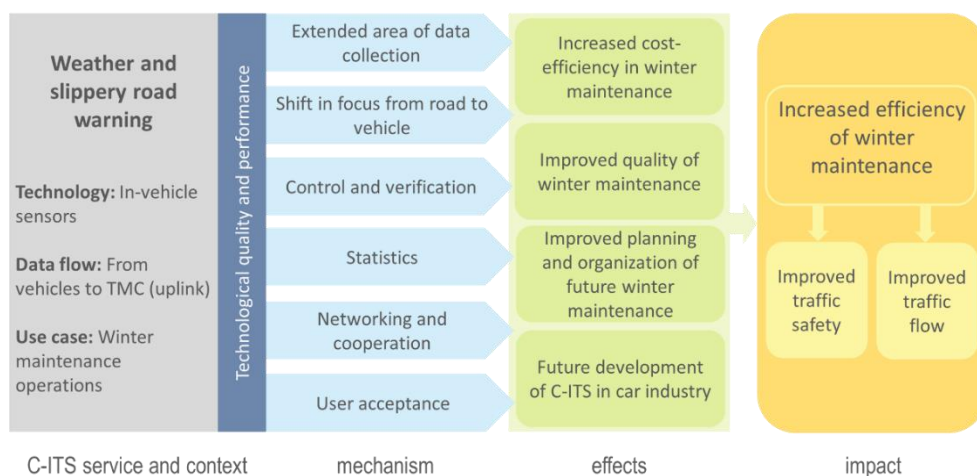
Figur 2: Programteori for ITS-tiltaket «Samkjøring» (Lervåg, 2021)

Implementeringsteorien utgjør her et logisk resonnement av hvilke betingelser som må være oppfylt for at samkjøringstiltaket skal avstedkomme de endringene man ønsker å oppnå. Dette er illustrert ved de reaksjonene som skjer i boksene i figuren: Virkningene som oppnås på sikt (gul boks) er betinget av at tiltaket innføres i en gitt kontekst (grå boks), at teknologien oppnår akseptabel teknologisk modenhet, og at det realiseres en rekke påvisbare effekter (grønne bokser).

Endringsteorien beskriver de underliggende mekanismer og kausale prosesser som sørger for at man oppnår ønsket effekter og virkninger. Dette er representert ved de blå pilene i figuren; *rekruttering, brukeraksept, motivasjon, holdningsendring og atferdsendring*. Samkjøringstiltakets endringsteori ble basert på en vitenskapsteoretisk atferdsmodell som kan forklare og predikere trafikantenes atferdsvalg (Fishbein og Ajzens, 2010).

Eksempel 2: Programteori for ITS-tiltaket «Varsling av glatt vegbane»

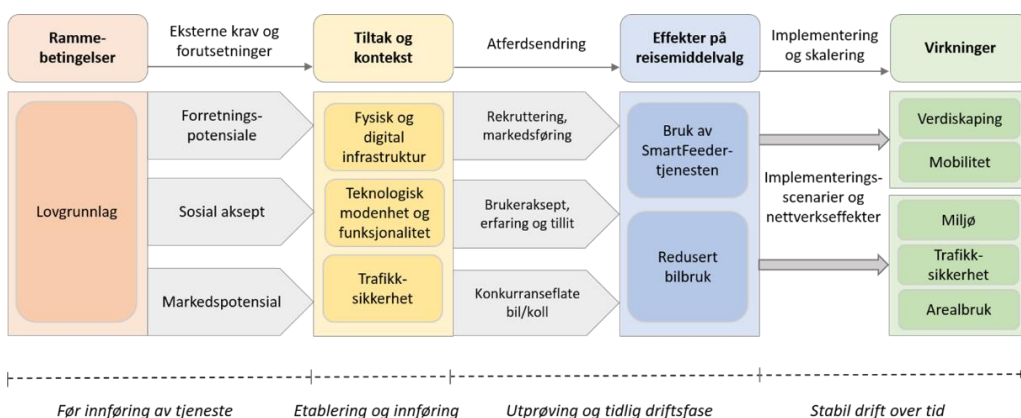
Figur 3 viser eksempel på programteori for ITS-tiltaket *Varsling av glatt vegbane*, utviklet i EU-prosjektet NordicWay og presentert i Lervåg (2021). Tiltaket innebar en tjeneste basert på kommunikasjonsteknologi og deling av sensordata fra kjøretøy i sanntid, som grunnlag for estimering av vegfriksjon. Som figuren viser, er tjenesten rettet inn mot overordnet beslutningstaking og kontraktsoppfølging av vinterdrift - som indirekte og over tid kan påvirke trafiksikkerhet og fremkommelighet.



Figur 3: Programteori for ITS-tiltaket «Varsling av glatt vegbane», utviklet i EU-prosjektet NordicWay (Lervåg, 2021).

Eksempel 3: Programteori for ITS-tiltaket «Smarte tilbringertjenester»

Figur 4 viser eksempel på programteori for ITS-tiltaket *Smarte tilbringertjenester med selvkjørende minibuss*, utviklet i forskningsprosjektet SmartFeeder (Lervåg m.fl., 2021). Hensikten med forskningsarbeidet var å bygge kunnskap om hvordan selvkjørende minibusser kan spille en rolle i fremtidens kollektivtransportsystem, med utgangspunkt i erfaringene fra de første pilotstudiene på norske veger. Tidslinjen i bunn av figuren, indikerer hvilke forskningstema som er hensiktsmessig å studere i de ulike fasene av innovasjonsprosessen.



Figur 4: Programteori for ITS-tiltaket «Smarte tilbringertjenester med selvkjørende minibuss», utviklet i forskningsprosjektet SmartFeeder (Lervåg m.fl., 2021)

Teoribaserte evalueringsdesign

Selv om grunnkonseptet med bruk av programteori og kausale årsakskjeder er felles for all teoribasert evaluering, er det ikke etablert et entydig forskningsdesign for teoribaserte evalueringsstudier. Det finnes flere varianter av evalueringsdesign, hvor spesifikke fremgangsmåter tilbyr bestemte strukturer for oppbygging av programteori og organisering av analyser. Felles for de teoribaserte fremgangsmåtene, er at det konkrete metodevalget foretas på bakgrunn av hva som er evalueringens hovedformål - altså hvilke spørsmål evalueringen skal besvare. Dette står i motsetning til tradisjonelle effektstudier, hvor fremgangsmåten er styrt av et predefinert metodeoppsett med statistisk analyse av før- og ettermålinger³.

I dette avsnittet presenteres to sentrale evalueringsdesign; *theory of change* og *realistisk evaluering*.

Theory of change-designet har stort fokus på implementeringen av tiltaket, med systematiske analyser av de lovmessige sammenhengene i årsaksvirkningskjeder og kontekst.

Det realistiske evalueringsdesignet vektlegger i større grad de bakenforliggende (psykologiske) aspektene som utløser individuelle atferdsendringer som en konsekvens av tiltaket (Blamey og Mackenzie, 2007; Hills og Junge, 2010; Pawson, 2013; Pawson og Tilley, 1997).

Theory of change-design

Theory of change-evalueringer tar utgangspunkt i en programteori som er strukturert i form av innsats, reaksjon, effekt og virkning som vist i Figur 1, med utfyllende beskrivelser og forklaringer av sammenhengen (de kausale relasjonene) mellom boksene (Funnel og Rogers, 2011; Government of Canada, 2012; Hills og Junge, 2010; Sverdrup, 2002; Weiss, 1995).

Programteorien etableres i en såkalt top-down-tilnærming, hvor man starter med målbildet og definerer hvilke virkninger man forventer å oppnå på lang sikt. Derfra jobber man seg stegvis bakover ved å identifisere hvilke betingelser som må være på plass for å oppnå resultater/effekter på de ulike nivåene. Prosessen involverer ideelt sett et bredt utvalg representanter fra prosjekter og interessegrupper, og det er en målsetting at det teoretiske rammeverket er basert på oppnådd konsensus blant deltakerne.

Når programteorien er etablert, vurderes denne med hensyn til plausibilitet (i hvilken grad modellen fremstår som logisk og rimelig), gjennomførbarhet og testbarhet (i hvilken grad det er mulig å måle/vurdere underliggende faktorer og effekter) (Blamey og Mackenzie, 2007; Bredgard, Salado-Rasmussen, og Sieling-Monas, 2016).

³ Det tradisjonelle metodeoppsettet gir svar på forskningsspørsmål av typen: *Virker tiltaket?* Det kan derimot ikke forklare *hvordan* eller *hvorfor* tiltaket virker.

Den som gjennomfører evalueringen, benytter så programteorien som utgangspunkt for valg av evalueringsdesign og analyser for testing av programteoriens ulike bestanddeler og kausale sammenhenger. Tilnærmingen legger ingen begrensninger på videre metodevalg, og det anbefales å kombinere både kvantitative og kvalitative analyser.

Tiltakets kausalitet betraktes som en generativ mekanisme hvor programteoriens årsakvirkningsskjeder testes med hensikt å finne de (kombinasjoner av) faktorer som har størst forklaringskraft. Kausaliteten bestemmes mekanistisk, på bakgrunn av følgende fire betingelser (Government of Canada, 2012):

- (1) Programteorien fremstår som plausibel og konsistent
- (2) Tiltakets innsats og aktiviteter ble implementert i henhold til plan
- (3) Underliggende antagelser ble bekreftet, effekter og virkninger ble observert
- (4) Kontekst og eksterne faktorer ble identifisert og vurdert

Evalueringer med utgangspunkt i *theory of change* har gjerne stort fokus på observerte effekter og beskrivelse av handlinger og prosesser som betinger resultater i de ulike fasene. Dette gir grunnlag for en helhetlig og overordnet forståelse for hvordan tiltaket fungerer i sine omgivelser. Resultatet av evalueringen er ofte en konkret og godt dokumentert *implementeringsteori* som forklarer hvordan kombinasjonen av ulike virkemidler og ressurser bidrar til tiltakets endelige virkning (Blamey og Mackenzie, 2007; Hills og Junge, 2010). Dette gir grunnlag for bred strategisk læring med nytte for fremtidig anvendelse av tiltaket.

Theory of change anses som en spesielt egnet fremgangsmåte i evalueringer hvor det fra før foreligger begrenset empiri og hvor tiltakene har høy innovasjonsgrad og kompleksitet (involverer mange aktører, består av pakker med flere virkemidler eller adresserer flere målsettinger). Fremgangsmåten gir svar på hvorvidt og hvordan et planlagt tiltak vil utløse en ønsket endring, kartlegger eksterne faktorer av betydning og sørger for involvering av relevante interessegrupper i prosessen (Blamey og Mackenzie, 2007).

Realistisk evalueringsdesign

Den realistiske evalueringen utvikler sin programteori i form av såkalte CMO-konfigurasjoner (context-mechanism-outcome) som definerer effekten av et tiltak som resultat av underliggende mekanismer og konteksten tiltaket opererer i (Pawson, 2013; Pawson og Tilley, 1997):

Outcome = Mechanism + Context

Til grunn for denne fremgangsmåten ligger en forståelse av at det samme tiltaket vil kunne utløse ulike effekter i forskjellige situasjoner og hos enkelte deltakere og målgrupper, avhengig av tiltakets egenskaper og den aktuelle konteksten. De underliggende mekanismene beskriver hva det er ved tiltaket som utløser en reaksjon eller endring i omgivelsene, ofte i form av bakenforliggende psykologiske eller sosiale forklaringsfaktorer. Tiltakets kontekst beskriver hvilke eksterne faktorer som bidrar til å utløse eller hindre en effekt. Det fundamentale spørsmålet som søkes besvart i realistiske evalueringer er: *Hva virker, for hvem, i hvilken kontekst og på hvilken måte?*

Den realistiske fremgangsmåten definerer en rekke CMO-konfigurasjoner med basis i tilgjengelig kunnskap og empiri om tiltaket. Dette skjer gjerne ved hjelp av intervju med prosjekteier og interessegrupper, og ved gjennomgang av forskningslitteratur og relevante dokumenter. På dette viset skiller realistisk evaluering seg fra *theory of change*-tilnærmingen, fordi målet er å identifisere de mest lovende teoretiske sammenhengene, uten at dette nødvendigvis skal skape konsensus blant involverte aktører.

Tabell 1 viser eksempler på relevante mekanismer (M) og kontekst (C) for ITS-tiltaket *Samkjøring*. Dette evalueringsgrunnlaget ble utviklet med utgangspunkt i tilgjengelig litteratur og erfaringer fra tidligere pilotprosjekter. Endringsmekanismene ble definert med utgangspunkt i en vitenskapsteoretisk atferdsmodell som forklarer/predikerer endring i atferd (reisemiddelvalg) basert på *individets personlige holdning, den sosiale normen* og *opplevd selvkontroll* (Fishbein og Ajzen, 2010). Videre empirisk testing av disse sammenhengene bidro til å identifisere hva som var de viktigste suksesskriteriene for vellykket gjennomføring av tiltaket (slik det senere presenteres i Tabell 2).

Tabell 1: Evalueringsgrunnlaget for utvikling av CMO-konfigurasjoner for ITS-tiltaket Samkjøring

Kontekst som påvirker effektoppnåelse (C)	Mekanismer som påvirker effektoppnåelse (M)	Ønsket effekt (O)
Tilfredsstillende teknisk løsning (mobil-app) og økonomisk godtgjørelse	Holdning og motivasjon: Fleksibilitet, tidsbesparelse, reisekostnader, sikkerhet, komfort, miljøhensyn, sosial image, teknologierfaring m.m.	Endret reisemiddelvalg: - Mer samkjøring - Redusert bilbruk
Informasjon og markedsføring	Rekrutteringsprosess (sosial norm): Påvirkning fra arbeidsgiver og kollegaer	
Lokalt trafikkbilde, tilgang til parkeringsplasser og sambruksfelt	Kontroll over atferdsvalg: Reisevaner, forpliktelser på arbeidsreisen som levering i barnehage, varehandel o.l.	
Demografiske bakgrunnsfaktorer i målgruppen		

CMO-konfigurasjonene utgjør hypoteser om hva som forventes å påvirke bestemte målgrupper i gitte omgivelser, og i neste steg velges de mest relevante hypotesene som utgangspunkt for empirisk testing. Det realistiske evalueringsperspektivet er metodenøytralt, men vektlegger vitenskapelighet i metodevalget. I praksis anbefales det å kombinere både kvantitative og kvalitative analyser i evalueringen.

Blamey and Mackenzie (2007) har oppsummert den realistiske fremgangsmåten i følgende fire punkter:

Steg 1: Forstå tiltakets natur og konteksten det opererer i

Steg 2: Utvikle relevante CMO-konfigurasjoner

Steg 3: Undersøke tiltakets virkninger for ulike målgrupper og omstendigheter

Steg 4: Justere CMO-konfigurasjonene basert på resultatene i steg 3, og velge ut kombinasjonene som best forklarer *hva som virker, for hvem, og i hvilken kontekst*.

Formålet med realistiske evalueringer er å finne kombinasjoner av målgruppe, innsats og kontekst som gir ønskede resultater, og ikke bevise en generell allmenngyldig effekt av tiltaket. Dette henger sammen med realistenes syn på kausalitet som er forankret i mekanistisk kausalteori og ikke kontrafaktisk gyldighet. I praksis betyr dette at evalueringen tar utgangspunkt i konkrete tilfeller hvor tiltaket har en åpenbar effekt, og anser sammenhengen mellom innsats og virkemidler som gyldig i den spesifikke konteksten (lokal kausalitet). Med dette perspektivet er overførbarheten mellom ulike implementeringer av tiltaket begrenset (lav ekstern validitet), men til gjengjeld muliggjør evalueringen en dypere forståelse av de underliggende mekanismene som trigger atferdsendring og forandring i enkelte situasjoner (høy intern validitet).

Når evalueringen rettes inn mot å avdekke hvilke kombinasjoner av kontekst og mekanismer som bidrar til en ønsket atferdsendring, gir resultatene grunnlag for å identifisere relevante suksesskriterier for innføring av tiltaket. Tabell 2 viser en oversikt hvilke faktorer som hadde størst betydning for effektoppnåelse i den nevnte studien av ITS-tiltaket *Samkjøring*.

Tabell 2: Suksesskriterier og potensielle effekter ved innføring av ITS-tiltaket samkjøring, basert på studie av pilotprosjekt i Bergen (Lervåg, 2021).

Kontekst	Mekanismer	Effekt
Rammebetingelser: Moden teknologisk løsning Fleksibilitet i arbeidstid Insentiver: Tilgang til sambruksfelt Dekning av reisekostnader Bakgrunnsfaktorer: Yngre aldersgrupper, kvinner, ikke småbarnsforeldre, høyere utdanning, teknologierfaring	Rekrutteringsprosessen: Aktiv prosjektorganisasjon Kjenne noen å samkjøre med Brukeraksept, holdninger og motivasjon for deltakelse: Opplevd tidsbesparelse Miljøhensyn Tillit til personvern Teknologioptimisme Endring i reiseatferd: Begrenset biltilgang Ikke daglige forpliktelser	Potensielle effekter i et optimalt scenario: Passasjerbelegg på 1,35 Redusert biltrafikk på 1-2 %

Den realistiske evalueringen vil typisk legge vekt på tiltakets *endringsteori*, med analyser av psykologiske og motivasjonsrelaterte variabler. Mens *theory of change-evalueringer* har sin styrke i overordnede evalueringer av komplekse tiltak på makronivå, passer den realistiske tilnærmingen bedre for tiltak som er tydelig avgrenset og med et definert influensområde på mikronivå (Blamey og Mackenzie, 2007). I et anvendt perspektiv, vil den realistiske evalueringen bidra til nyttig kunnskap om hvordan virkemidlene i et tiltak

kan skreddersys for å oppnå best mulig effekt i ulike situasjoner. Det er imidlertid fullt mulig å kombinere de to sentrale teoribaserte metodene for å oppnå en mer komplett forståelse av tiltakets fullstendige virkemåte og potensiale.

Sammenfatning av det teoribaserte evalueringsperspektivet

Dette kapittelet har presentert hovedpunktene i det teoribaserte evalueringsperspektivet. Bruk av en teoribasert tilnærming i evalueringsarbeidet, vil kunne styrke kunnskapsgrunnlaget for utvikling og innføring av komplekse og innovative tiltak og tjenester på flere måter:

- *Det bidrar til struktur og systematikk rundt metodevalg og prioritering av evalueringsressurser.*
- *Det utvider metodegrunnlaget ved å inkludere flere former for validitet og kausalteori.*
- *Det skifter fokus fra evaluering som verktøy for kontroll og styring, til evaluering som verktøy for innovasjon og kompetanseutvikling.*

Selv om mange tradisjonelle evalueringer i praksis kombinerer et bredt spekter av metoder, er det likevel forskjell på den overordnede evalueringsstrategien som igjen er forankret i litt ulike vitenskapsteoretiske utgangspunkt: Klassiske effektevalueringer undersøker om et tiltak virker gjennom etablering av en kontrafaktisk situasjon ved bruk av kontrollgrupper. Forsøksdesignet er innrettet mot å finne ett gyldig resultat: Effekten av tiltaket på et gitt tidspunkt (om tiltaket virker eller ikke). I denne tilnærmingen er de kvantitative analysene dominerende, og kvalitative tilleggsanalyser har ofte som formål å øke de kvantitative resultatenes validitet når metodiske betingelser ikke er oppfylt.

I det teoribaserte evalueringsperspektivet vurderes tiltaket etter flere kriterier enn ren effektmåling, og innsatsen styres i større grad mot å etablere kunnskap om suksessfaktorer som må styrkes eller barrierer som må overvinnes for å lykkes med tiltaket. Metodevalg og utvikling av (flere) forsøksdesign er en integrert del av evalueringsprosessen, som åpner for at hele eller deler av evalueringen kan baseres på en mekanistisk kausalitetsforståelse. I praksis betyr dette at kausalitet sees som en latent egenskap som kan aktiveres under visse omstendigheter: Det er altså ikke automatikk i at tiltaket vil utløse en gitt effekt, det påvirkes f.eks. av forhold rundt implementeringen eller respons fra målgruppen. Dette innebærer at evalueringen ikke nødvendigvis produserer *effektbevis*, men heller sannsynliggjør at tiltaket har hatt eller vil utløse en effekt under gitte forutsetninger. Samtidig gir den innsikt i tiltakets forbedringspotensial. Dette tilrettelegger for en dynamisk innovasjonskultur hvor man lærer av feil og hele tiden søker etter nye og bedre løsninger.

Den teoribaserte evalueringens evne til å forklare kausale årsakssammenhenger og inkludere kontekstuelle faktorer i vurderingen, bidrar til anvendbar kunnskap om hvilke virkemidler som har effekt, for hvilke målgrupper og i hvilke situasjoner. I tillegg har metodikken kraft til å forklare i hvor stor grad kombinasjonen av ulike virkemidler bidrar til det endelige resultatet, og hvilke eksterne forhold som styrker eller hindrer oppnåelsen av et ønsket resultat (Chen, 1990; Pawson og Tilley, 1997; Rossi, Freeman, og Lipsey, 2003; Sanderson, 2002). I motsatt utfall, dersom den forventede effekten uteblir, kan

evalueringen også forklare hvorfor tiltaket ikke virker. Det skilles i denne sammenhengen mellom *teorifeil* og *implementeringsfeil* (Bredgard et al., 2016; Dahler-Larsen, 2013). Dersom det foreligger en teorifeil, er det antakelsene som ligger til grunn for programteorien som ikke stemmer, og man kan anta at tiltaket ikke har den effekten man forventet i utgangspunktet. Foreligger det derimot en implementeringsfeil, er resultatet påvirket av andre forhold som kan fikses for å oppnå bedre resultater i en fremtidig implementasjon.

Figur 5 viser en sammenstilling av egenskaper ved en klassisk (eksperimentell) tilnærming til effektstudier og en teoribasert evalueringsmetodikk. I en overordnet teoribasert tilnærming kan det spesifikke teoribaserte metodeoppsettet (evalueringsdesign og analyser) være både et alternativ og et supplement til tradisjonelle eksperimentelle analyser.

KLASSISKE EFFEKTSTUDIER		TEORIBASERT EVALUERINGSMETODIKK
Metodestyrte og målorienterte evalueringprosesser	Tilnærming	Teoribasert evalueringsspross med utvikling av tiltakets programteori
Det vitenskapelige eksperimentet: (Kvasi-)eksperimentell og kontrafaktisk design	Forskningsdesign	Kombinerer flere design som besvarer ulike spørsmål. F.eks. eksperimentell og teoribasert (theory of change og/eller realistisk).
Kvantitativ datainnsamling og statistiske analyser	Metoder	Kvalitativ og kvantitativ datainnsamling og analyser
Gir vitenskapelig sterkt effektbevis når metodekrav er oppfylt	Resultat	Forklarer sannsynlige effekter, rammebetingelser, suksesskriterier og barrierer ved innføring

+
eventuelle
supplerende analyser

Figur 5: Sammenstilling av egenskaper ved klassiske effektstudier og en teoribasert evalueringsmetodikk. I praksis vil klassiske effektstudier ofte suppleres med tilleggsanalyser som kan bidra til å forklare resultatene eller øke studiens validitet når de underliggende metodekravene ikke kan oppfylles.

3. Når er teoribasert evaluering særlig egnet?

Det teoribaserte evalueringsperspektivet omfatter både effekt- og implementeringsanalyser, og kombinerer kvantitative og kvalitative evalueringsmetoder. På denne måten anses gjerne teoribaserte evalueringer som et bindeledd mellom klassiske, målorienterte effektevalueringer og kvalitative prosessevalueringer som har større fokus på gjennomføringsprosessen (Bredgard et al., 2016; Pawson og Tilley, 1997).

Sammenhengen mellom de nevnte evalueringsteoriene er vist i Tabell 3, gjengitt og (fritt) oversatt fra Bredgard et al. (2016).

Tabell 3: Typiske egenskaper ved det teoribaserte evalueringsperspektivet sett i forhold til klassisk effektevaluering og prosessevaluering.

	Klassisk effektevaluering	Teoribasert evaluering	Prosessevaluering
Forskningsspørsmål	Virker det?	Hva virker for hvem, hvorfor, og under hvilke omstendigheter?	Hvordan oppleves og oppfattes virkningene?
Fokus	Effekter	Effekt og prosess	Prosesser
Metode	Kvantitative metoder	Metodenøytralt; anvender både kvantitative og kvalitative metoder	Kvalitative metoder
Kausalteori	Kontrafaktisk kausalteori	Mekanistisk kausalteori	Logisk teori (resonnement)
Ideal for kunnskapsdannelse	Kontekstuavhengig og global viten	Kontekstavhengig, men overførbar viten	Kontekstavhengig og lokal viten
Beslutningsprosess	Antar at beslutningstakere foretar rasjonelle valg av de mest effektive virkemidler for å oppnå et mål	Informerer praktikere, deltakere og beslutningstakere, samt utvikler deres programteorier	Vektlegger lydhørhet overfor berørte interessenter, og informerer beslutnings-takere om hvordan tiltaket oppleves og anvendes

Den teoribaserte tilnærmingens egnethet i evalueringer diskuteres som regel i relasjon til klassiske effektevalueringer som følger det (kvasi-) eksperimentelle metodeoppsettet⁴, fordi dette i stor grad representerer normen for all evidensorientert metodikk. Forskningslitteraturen er forholdsvis entydig og konsistent når det gjelder de ulike evalueringsmodellenes fortreffelighet og mangler, og den teoribaserte tilnærmingen synes å ha sin styrke i følgende egenskaper (Blamey og Mackenzie, 2007; Bredgard et al., 2016; Chen, 2012; Funnell og Rogers, 2011; Government of Canada, 2012; Hills og Junge, 2010; Pawson og Tilley, 1997; Weiss, 1995):

⁴ Idealet for eksperimentell metodikk er randomiserte kontrollerte forsøk (RCT) med før- og etterundersøkelser og sammenligning mot en kontrollgruppe. Dette evalueringsdesignet er illustrert i Figur 10.

- *Den bidrar til kunnskap om hvorfor et tiltak virker eller ikke.* Dette gjør den spesielt egnet for evalueringer med et læringsperspektiv, når formålet er å skape kunnskap som grunnlag for strategiske valg og fremtidige implementeringer.
- *Den håndterer komplekse tiltak og kausale årsaksvirkningsskjeder.* Dette gjør den spesielt egnet for tiltak med høy innovasjonsgrad, med mange involverte aktører og virkemidler, samt når det foreligger begrenset empiri om tiltaket fra før. Videre er det en styrke at tilnærmingen inkluderer både planlagte og utilsiktede effekter i vurderingen, og at den tar høyde for potensielle virkninger av tiltaket i et lengre tidsperspektiv.
- *Den inkluderer kontekstuelle faktorer.* Dette gjør den spesielt egnet for tiltak hvor man antar at utenforliggende forhold spiller en viktig rolle for utfallet av tiltaket, eller hvor virkningen som oppnås avhenger av egenskaper ved målgruppen eller selve implementeringen.
- *Den er anvendelig.* Den teoribaserte evalueringen kan som regel benyttes i situasjoner hvor betingelsene for tradisjonelle metoder ikke er oppfylt. Dette bidrar til at man kan fremskaffe nyttig kunnskap også i situasjoner hvor muligheten til omfattende datainnsamling er begrenset, og der det ikke er praktisk mulig å gjennomføre eksperimentelle analyser. Mens klassiske effektstudier forutsetter målbare resultater i løpet av prosjektperioden, kan den teoribaserte tilnærmingen også fremskaffe relevant kunnskap i tilfeller med små, kvalitative eller indirekte effekter.
- *Den har formativ kapasitet.* Den teoribaserte evalueringen kan utføres i hele tiltakets livssyklus. Dermed kan den gi løpende innspill og tilbakemeldinger som grunnlag for å foreta korrigeringer og velinformerte beslutninger allerede før tiltaket implementeres. Dette bidrar til fokusert innsats og effektiv bruk av prosjektressurser.
- *Den bidrar til forventningsavklaring og bevisstgjøring hos involverte aktører.* Utviklingen av programteorien legger til rette for deltakelse fra interessenter og prosjekteier i evalueringsprosessen, og sikrer at involverte aktører utvikler en felles forståelse både for tiltakets og evalueringens formål og avgrensning. Både prosessen og den etablerte programteorien blir et kommunikativt instrument i prosjektet og mot omverdenen.

Den teoribaserte evalueringen anses derimot å være mindre egnet til å produsere kvantitative bevis med samme vitenskapelige styrke som eksperimentelle modeller (Blamey og Mackenzie, 2007; Hills og Junge, 2010; Weiss, 1998). Der hvor eksperimentelle analyser er basert på vitenskapens krav til stringens og validitet, benytter den teoribaserte modellen en praktisk, pragmatisk tilnærming basert på logikk og erfaring. Dette innebærer at teoribaserte evalueringresultater har større nytte i et læringsperspektiv, enn som grunnlag for kontroll og dokumentasjon av effektivitet. Dette kompenseres imidlertid av at den teoribaserte tilnærmingen åpner for en fleksibilitet i metodevalget, som gjør at eksperimentelle analyser fortsatt kan inkluderes i de deler av evalueringen hvor dette finnes hensiktsmessig. Derimot vil den tette kontakten mellom evaluator og prosjekteier, blant annet gjennom utviklingen av programteori og feedback av evalueringresultater

underveis i prosessen, kunne sette spørsmålstegn ved evalueringens objektivitet. Samarbeidet mellom oppdragsgiver og den som gjennomfører evalueringen kan dermed brukes som argument mot å benytte evalueringresultatene i et kontrollperspektiv.

Evalueringspraksis viser imidlertid at samfunnsfaglige evalueringer som foregår ute "i den virkelige verden" sjelden oppfyller de strenge metodiske kravene som følger av den eksperimentelle evalueringsmetodikken. Ofte er datagrunnlaget begrenset (f.eks. mangler gode data i før-situasjonen) eller det er vanskelig å etablere entydig sammenheng mellom tiltak og effekt (f.eks. påvirkes resultatene av det sosiale samspillet mellom menneske, teknologi og eksterne rammebetingelser). I slike tilfeller kan den teoribaserte tilnærmingen bidra til å øke studiens validitet gjennom logisk oppbygning og testing av årsaksvirkningskjeder.

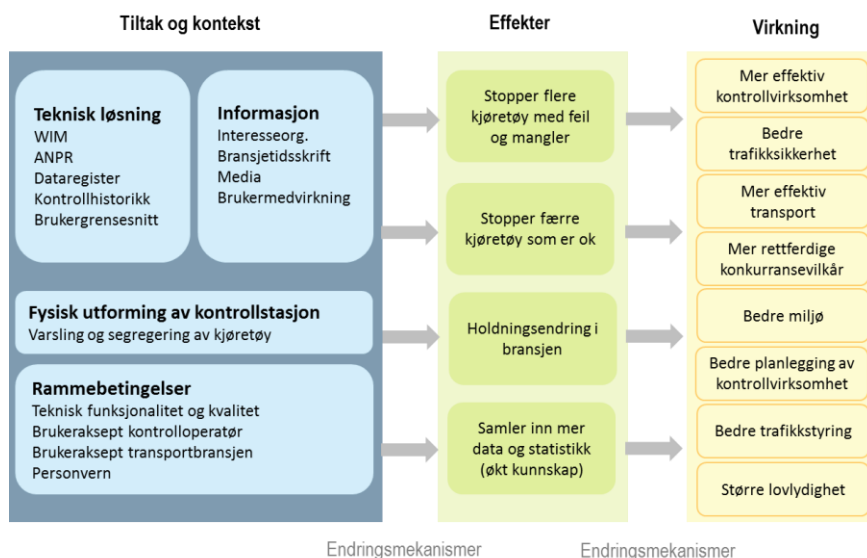
Med styrker og svakheter tatt i betraktning, synes den teoribaserte tilnærmingen å være spesielt egnet når en eller flere av følgende betingelser inntreffer (Blamey og Mackenzie, 2007; Hills og Junge, 2010);

- Evalueringens formål er læring og kunnskapsbygging
- Tiltaket som evalueres er komplekst
- Det foreligger lite empirisk kunnskap om tiltakets kausale sammenhenger og effekter
- Tiltakets utfall vil påvirkes av eksterne effekter eller kontekst
- Det forventes at tiltaket eller dets omgivelser vil utvikles over tid

4. Erfaringer med en teoribasert fremgangsmåte for evaluering av intelligente transportsystemer (ITS)

En teoribasert evalueringsprosess ble utviklet og benyttet i to evalueringsprosjekter med svært forskjellige ITS-tiltak, som del av doktorgradsarbeidet presentert i Lervåg (2021).

Det første case-studiet omhandlet evaluering av ITS-tiltaket *NonStop klarering av kjøretøy for tungbilkontroll*. Tiltaket består av et automatisk system for målrettet utvelgelse av tunge kjøretøy i forkant av kontrollstasjonen, basert på registrering av vektdata mens kjøretøyet holder normal hastighet i en trafikkstrøm. Tiltakets formål er å effektivisere kontrollprosessen, og således spare tid for transportnæringen, ved at kjøretøy uten indikasjoner på lovbrudd får fordelen av å passere kontrollstasjonen uten videre forsinkelser. NonStop-tjenesten viste seg imidlertid å ha begrenset funksjonalitet og vesentlig lavere teknologisk modenhet enn forutsatt ved prosjektets oppstart.



Figur 6: Programteori for ITS-tiltaket *NonStop klarering av kjøretøy for tungbilkontroll*, utviklet i forskningsprosjektet *NonStop* og presentert i Lervåg (2021).

Det andre case-studiet omhandlet evaluering av ITS-tiltaket *Samkjøring med turkobling via mobilapplikasjoner*, med utgangspunkt i et pågående pilotprosjekt i Bergen. Se tidligere presentasjon av programteori for tiltaket i Figur 2.

Samkjøringstiltaket var preget av høy kompleksitet med mange involverte aktører og ulike virkemidler knyttet til markedsføring og insentiver for deltakelse i samkjøringsordningen. Formålet med evalueringsstudien var å avdekke hvorvidt samkjøring har potensiale som etterspørselsstyrende og trafikkdempende tiltak i Norge, samt å identifisere relevante suksesskriterier og barrierer for innføring av tiltaket.

En kort beskrivelse av de to ITS-tiltakene er oppsummert i Tabell 4.

Tabell 4: ITS-tiltakenes ulike innretning i de to casestudiene

ITS-tiltak	NonStop	Samkjøring
ITS-applikasjon	NonStop beslutningsstøtte på kontrollstasjon	Samkjøringsapplikasjon på mobil
ITS-tjeneste	Forhåndsklarering av kommersielle kjøretøy i forkant av kontrollstasjonen (uten forsinkelse)	Etterspørselsbasert samkjøring basert på styring av delte reiser i privatbil
IKT infrastruktur	Nettbrett, sensorer langs veg, samt dataregistre og beslutningsstøttesystem	Mobiltelefon og koblingsalgoritmer i et sentralsystem
ITS-bruker	Myndighetene (trafikk-kontrollør)	Privatpersoner (arbeidsreisende)
Innføringsløp	Obligatorisk	Frivillig
Teknologisk modenhet	Tidlig fase, teknologisk umoden	Kommersiell løsning, antatt teknologisk moden
Markedspenetrasjon	Effekt fra første bruker	Forutsetter kritisk brukermasse
Antatt nyttevariabel	Redusert reisetid for transport, økt deteksjonsrate på kontrollstasjon	Endring i reiseatferd, reduksjon i antall bilreiser
Situasjonsavhengige faktorer	Fysisk utforming kontrollstasjon, bruksområder for tjenesten og trafikkvolum	Rammebetingelser, insentiver og bakgrunnsfaktorer i målgruppen

Tabellen viser at det er stor variasjon i ITS-tjenestene, både med hensyn til teknologi, målgrupper og virkningene det tas sikte på å oppnå. I tillegg påvirkes effektbildet av rammebetingelser og situasjonsavhengige faktorer rundt iverksettelsen av tiltaket. Det har eksempelvis stor betydning hvorvidt ITS-tjenesten er obligatorisk, som i tilfellet med NonStop-tiltaket (all næringstrafikk som passerer kontrollpunktet blir påvirket), kontra Samkjøringstiltaket som er frivillig, og dermed i mye større grad er avhengig av brukeraksept, betalingsvilje og eksterne insentiver. Til tross for at de to ITS-tiltakene har svært ulik innretning og egenskaper, har den teoribaserte evalueringsprosessen vært anvendbar og bidratt med nyttig kunnskap i begge tilfellene.

De to casestudiene har gitt en rekke sammenfallende metodiske erfaringer. Begge evalueringene hadde som formål å fremme kunnskapsbygging og økt forståelse for hvordan tiltaket kan fungere i en fremtidig anvendelse. I begge tilfellene avdekket den innledende litteraturgjennomgangen at det var store utfordringer med kvantitativ innsamling og analyse av effektdata i tidligere studier av tilsvarende ITS-tiltak. Det forelå derfor lite publisert forskning på faktiske virkninger. Likevel var evalueringspraksisen for tilsvarende tiltak i stor grad basert på klassiske effektstudier, uten særlig forklaring av antakelsene som lå til grunn for analysene. De fleste studiene manglet kontrollgrupper, og årsaksvariabelen var i realiteten en miks av ulike tiltak (ITS-applikasjon, holdningskampanjer, reguleringer og opplæring m.m.). Kvalitative analyser var i liten grad satt i system, og som regel kun gjennomført sekundært når konklusjoner ble fattet på et sparsomt datagrunnlag.

Begge tiltakene har vært forholdsvis komplekse, enten i form av en omfattende IKT-infrastruktur (som i NonStop-tiltaket) eller med en kombinasjon av flere tjenester og virkemidler (i begge tiltakene). I begge casene har den teoribaserte evalueringsprosessen

vist seg å være praktisk gjennomførbar, til tross for relativt små datautvalg. Metodikken har også vist stor fleksibilitet med hensyn til tidspunktet for evaluering. NonStop-tiltaket var fortsatt i en tidlig fase av innovasjonsprosessen, og dermed preget av lav teknologisk modenhet og ustabil pilotering. Samkjøringstiltaket befant seg i motsatt ende av skalaen, med en ferdig kommersiell løsning som allerede var på markedet. Likevel ble det erfart at begge applikasjonene var gjenstand for en kontinuerlig teknologiutvikling, som endret egenskaper og stabilitet underveis i evalueringsprosessen. Den teoribaserte evalueringsprosessen har likevel lyktes med å frembringe nyttig kunnskap, både om tiltakenes kapasitet som trafikkstyrende virkemiddel og om viktige rammebetingelser for å lykkes med tiltakene. Således har de teoribaserte metodene gitt økt innsikt og merverdi, selv i situasjoner der det ikke er mulig å bevise en umiddelbar effekt.

I begge case-studiene ble teknologisk funksjonalitet, kvalitet og modenhet behandlet som en rammebetingelse i forbindelse med etablering av programteorien. Temaet var imidlertid svært sentralt, både med hensyn til valg av evalueringsdesign og realisering av tiltakets effekter. Det ble erfart at teknologien ga føringer for hvilke forskningsspørsmål som var relevante å stille, samt datainnsamlingsmetodene som ble benyttet i gjennomføringsfasen. Det ble derfor konkludert med at tiltakets teknologiske modenhet kunne vært løftet frem som et eget punkt tidlig i evalueringsprosessen.

Det er verdt å merke seg at evalueringer i en tidlig fase av tiltakets innføringsløp gir mulighet for å utnytte resultatene underveis, fordi man tidlig kan avdekke kritiske barrierer og gjøre nødvendige justeringer (Bredgard et al., 2016; Chen, 2005; Karlsen og Jentoft, 2013). I så måte kan teoribaserte evalueringer bidra til mer effektiv ressursbruk både for teknogitilbyder og transportmyndighet. Eksempelvis ble det i NonStop-evalueringen tidlig avdekket at den fysiske utformingen av kontrollområdet hadde stor betydning for hvorvidt nytteeffekten i form av redusert ventetid for næringstransporten kunne realiseres. Det ble dermed iverksatt fysiske tilpasninger av omgivelsene parallelt med teknologiutviklingen. I tillegg ble det i en tidlig fase av evalueringen avdekket at brukergrensesnittet ikke hadde oppnådd tilstrekkelig funksjonalitet. Dette la føringer for både gjennomføringen av piloten og evalueringsdesignet, slik at man unngikk feilslått bruk av evalueringsressursene.

Fleksibiliteten i den teoribaserte evalueringsprosessen gjorde det mulig å skille mellom det som Dahler-Larsen (2013) benevner som *teorifeil* (tiltaket virker ikke) versus *implementeringsfeil* (ytre faktorer hindrer at tiltaket virker) i tiltakets effektivvurdering. Erfaringene fra casestudiene gir altså støtte til litteraturen som beskriver den teoribaserte tilnærmingen som et godt virkemiddel for proaktiv kunnskapsproduksjon (Chen, 2005; Karlsen og Jentoft, 2013) og hensiktsmessig både i planleggingsfasen av tiltaket, som styringsredskap og som evalueringsmodell (Bredgard et al., 2016).

Etablering av tiltakets programteori ble opplevd som en nyttig øvelse i begge casestudiene. Allerede tidlig i evalueringsprosessen, ble det mulig å avdekke logiske brister og mismatch mellom ulike aktørers forventninger (og behov) og hva som er realistisk å oppnå. I NonStop-caset hadde interessegruppene ulikt syn på hvilke lovbrudd som skulle inkluderes i beslutningsstøttesystemet, mens i Samkjørings-caset var det til dels urealistiske forventninger til miljøeffekten av tiltaket. Programteorien ble i begge

tilfellene et nyttig verktøy for forventningsavklaring, og bidro til å skape en felles forståelse av det teoretiske grunnlaget for innføringen av ITS-tiltaket.

Erfaringene samsvarer med tidligere teoretiske betraktninger som peker på den teoribaserte evalueringsmodellens styrke som kommunikativt verktøy (Kaplan og Garrett, 2005; Marchal et al., 2013; van der Knaap, 2004). Det finnes ikke en generell oppskrift på hvilke elementer som skal inngå i en programteori, og programteorien må utvikles og tilpasses det konkrete tiltaket i hvert tilfelle. Dette krever en viss innsats fra prosjektgruppa i en tidlig fase av evalueringsprosessen. En mulig utfordring som trekkes frem i metodelitteraturen, er at det kan være krevende å etablere programteorien med mange involverte aktører og motstridende interesser (Berger og Vrangbæk, 2011). Dette var ikke en fremtredende erfaring i de to case-studiene. Tvert imot ble det opplevd som en styrke at evalueringsprosessen var grundig forankret i et felles teorigrunnlag.

Konteksten for iverksettelsen av tiltaket har hatt stor betydning i begge case-studiene. Det er allerede nevnt at den fysiske utformingen av kontrollstasjonen hadde stor innvirkning på nytteverdien av NonStop-tiltaket. Tilsvarende ble samkjøringstiltaket i sin tid iverksatt i omgivelser som var forventet å være spesielt gunstige med tanke på å rekruttere brukere, både med hensyn til problemopplevelse (store trafikkutfordringer) og insentiver (blant annet tilgang til sambruksfelt for kjøretøy med flere passasjerer). Tidligere evalueringer av slike tiltak, har i stor grad behandlet kontekst og eksterne faktorer som en svakhet i datamaterialet som det er vanskelig å kontrollere (Gjerstad og Bayer, 2012; Griffin et al., 2015). I den teoribaserte tilnærmingen ble disse faktorene tatt inn som en integrert del av evalueringsstudien.

I begge ITS-studiene ble det teoribaserte evalueringsdesignet *theory of change*, benyttet som fremgangsmåte for å vurdere tiltakets evne til å utløse ønskede effekter. For samkjøringstiltaket ble det i tillegg benyttet et *realistisk evalueringsdesign* som grunnlag for å vurdere hvilken kombinasjon av mekanismer og kontekst som hadde størst betydning for reisemiddelvalget (se eksempel i Tabell 2). Den sterkeste kritikken mot teoribaserte metoder er knyttet til manglende evidensorientering; altså liten evne til å produsere kvantitative bevis med samme vitenskapelige styrke som eksperimentelle modeller (Blamey og Mackenzie, 2007; Hills og Junge, 2010; Weiss, 1998). Det er imidlertid viktig å være klar over at en *overordnet* teoribasert tilnærming ikke er avgrenset til bruk av teoribaserte evalueringsdesign. Tvert imot åpner fremgangsmåten for å kombinere flere hensiktsmessige metoder og innfallsvinkler for testing av de ulike sammenhengene i programteorien. I evalueringen av samkjøringstiltaket ble det for eksempel inkludert en kvantitativ analyse av reisemiddelvalg før og etter innføring av tiltaket. NonStop-caset er også et eksempel på at en teoribasert evalueringsprosess kan gi input til tradisjonelle beslutningsprosesser med bruk av nyttekostnadsanalyser.

Erfaringene fra case-studiene har vist at en overgang fra klassiske effektanalyser til en teoribasert evalueringstilnærming ikke vil begrense kunnskapsproduksjonen. Tvert imot vil den i mange tilfeller bidra til økt forståelse omkring rammebetingelser og eksterne faktorer som må hensyntas for å oppnå en vellykket implementering.

5. Rammeverk for en teoribasert evalueringsprosess

I dette kapittelet presenteres en teoribasert evalueringsprosess for oppbygging av en hensiktsmessig evalueringsmetodikk, som tilpasses det aktuelle tiltaket og evalueringens formål i hvert tilfelle. Rammeverket gir en stegvis veiledning som støtter forskeren i prosessen med å definere relevante forskningsspørsmål, etablere en programteori og metodevalg, samt gjennomføre datainnsamling og analyser. Den teoribaserte evalueringsprosessen kan benyttes for evalueringer gjennom hele tiltakets livsløp (både ex ante og ex post studier), og den gjør det mulig å kombinere flere ulike evalueringsdesign og analysemetoder.

Et sammendrag av evalueringsprosessen er gjengitt i Figur 7 og nærmere beskrevet i avsnittene som følger.



Figur 7: Rammeverk for en teoribasert evalueringsprosess (Lervåg, 2021)

Trinn 1: Tiltakets teknologiske modenhet

Digitalisering og rask teknologiutvikling gjør at IKT-systemer og andre teknologibaserte tjenester ofte gjennomgår en gradvis forbedring gjennom hele sin levetid. Dette gir vesentlige implikasjoner for evalueringsprosessen: De ulike fasene i teknologiens livsløp gir grunnlag for å besvare forskjellige forskningsspørsmål, og tiltakets teknologiske funksjonalitet og modenhet har dermed avgjørende betydning for senere valg av evalueringsmetode og analyser.

Vurdering av teknologisk modenhet kan eksempelvis ta utgangspunkt i Technology Readiness Levels (TRL) som benyttes som felles grunnlag for vurdering av teknologistatus i EUs Horizon-program. TRL-konseptet adresserer hele innovasjonsprosessen med en detaljert oversikt over ulike forsknings- og utviklingssteg fra tidlig konseptutvikling til ITS-tjenesten rulles ut på markedet, som vist i Figur 8.

TRL 1	TRL 2	TRL 3	TRL 4	TRL 5	TRL 6	TRL 7	TRL 8	TRL 9
Basic principles observed	Technology concept formulated	Experimental proof of concept	Technology validated in lab	Technology validated in relevant environment	Technology demonstrated in relevant environment	System prototype demonstrated in operational environment	System complete and qualified	Actual system proven in operational environment
Concept development			Proof of principle			Proof of performance		

Figur 8: Technology Readiness Scale (TRL) for vurdering av teknologisk modenhet

Evalueringsdesignet (metodevalg og fremgangsmåte) må derfor tilpasses ITS-tjenestens modenhet og stadium i innovasjonskjeden. Evaluering av IKT-tjenester i en tidlig fase av innovasjonsprosessen (typisk pilotstudier på TRL 3-5) kan ikke besvare de samme forskningsspørsmålene som evaluering av ITS-tjenester med en høyere grad av teknologisk modenhet (typisk implementeringsstudier på TRL 6-9). Evalueringsdesignet (metodevalg og fremgangsmåte) må derfor tilpasses ITS-tjenestens modenhet og stadium i innovasjonskjeden.

Tilgangen til hensiktsmessige evalueringsmetoder og datagrunnlag endrer seg etter nivå av teknologisk modenhet. På lavere nivå benyttes ofte ekspertvurderinger, modellering og simulering, mens feltforsøk og effektmåling krever et høyere nivå av teknologisk modenhet. Ex-post-evaluering basert på den (kvasi-) eksperimentelle fremgangsmåten med før- og etter-målinger, er kun hensiktsmessig å gjennomføre når tiltaket har nådd en høy grad av teknologisk modenhet, når implementeringsfasen er fullført og tiltaket fungerer stabilt over en lengre periode.

Det finnes mange eksempler på tiltak som har vist seg å ha begrenset funksjonalitet og lavere teknologisk modenhet enn forutsatt ved prosjektets oppstart (Lervåg, 2021). Det er også verdt å merke seg at et IKT-tiltak kan være sammensatt av flere teknologier og tjenester som hver for seg innehar ulik grad av modenhet.

Trinn 2: Behovsanalyse

Som grunnlag for evalueringen må det gjøres en kartlegging av hva man søker (eller har forsøkt) å oppnå ved innføring av det aktuelle tiltaket; altså definere tiltakets bruksområde og målsettinger. Hvilke brukerbehov eller samfunnsutfordringer ligger til grunn for innføring av tiltaket? Dette er særlig relevant når tiltak iverksettes med den teknologiske løsningen som utgangspunkt, uten at denne adresserer et definert behov eller samfunnsmessig utfordring (technology push). Behovsanalysen skal gi svar på hvilke underliggende problemer eller behov tiltaket er ment å løse, samt en spesifisering av hvilke målgrupper som berøres av tiltaket. Definisjon av tiltakets bruksområde (kontekst) og målsettinger kan blant annet skje på bakgrunn av møter med transportmyndigheter og interessegrupper, ekspertvurderinger (f.eks. Delphi-undersøkelser), intervjuer med brukere og spørreundersøkelser. Behovsanalysen gir grunnlag for utarbeidelse av

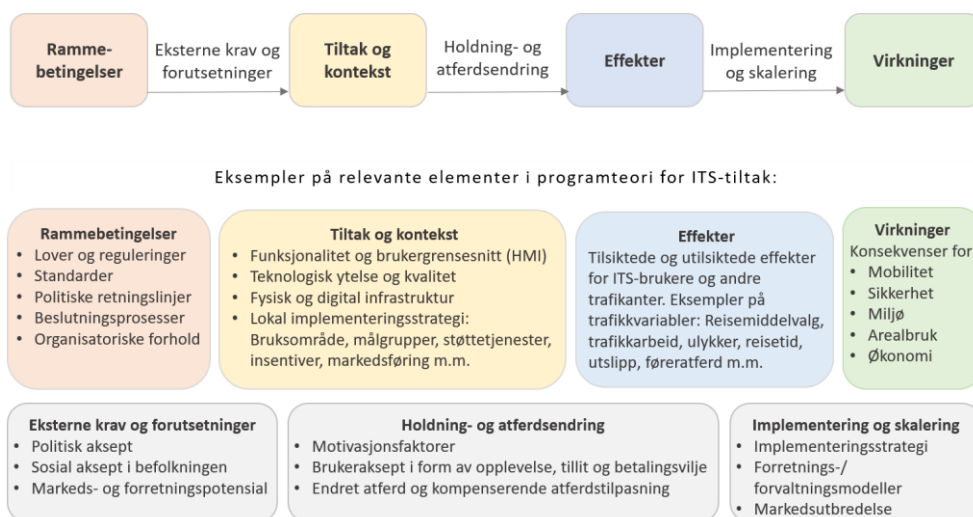
programteorien og evalueringskriterier senere i evalueringsprosessen, samt mulighet til å vurdere tiltakets relevans, for eksempel i et samfunnsperspektiv.

Trinn 3: Programteori

Utvikling av tiltakets *programteori* er en kritisk oppgave som danner grunnlaget for hele evalueringsprosessen. Programteorien er en konseptuell modell av sammenhengen mellom tiltakets innsats og virkninger, inkludert de underliggende mekanismene som utløser effekter på kort og lang sikt. Den beskriver altså *tiltakets underliggende logikk*; hvordan tiltaket er ment å operere og medføre endringer i en gitt kontekst. Dette representerer således prosjekteiers (ofte myndighetenes) forventninger og rasjonelle grunnlag for innføring av tiltaket. For evaluering av veg- og trafikktiltak inkluderer dette som regel effekter knyttet til trafiksikkerhet, trafikkavvikling, komfort, miljø, mobilitet og/eller økonomiske gevinster for bruker, næringsliv eller samfunnet for øvrig.

Programteorien utformes basert på resultater fra behovsanalysen og tilgjengelig kunnskap om tiltakets årsaksvirkningskjeder, forventede effekter og virkninger på lang sikt. Den konseptuelle modellen kan inneha ulik grad av detaljering og kompleksitet, avhengig av tiltakets beskaffenhet, eksisterende kunnskapsgrunnlag og evalueringsens formål.

Figur 9 viser en generell lineær programteori for intelligente transportsystemer, med eksempler på elementer som kan inngå i modellen, presentert i Lervåg (2021). Når tiltakets kompleksitet og evalueringsens detaljfokus tilar, kan denne utvikles til mer avanserte modeller som også kan fremstilles hierarkisk eller på matriseform.



Figur 9: Skjematisk fremstilling av programteori for ITS-tiltak, med eksempler på relevante elementer. Listen er ikke uttømmende, og forholdene kan være mer eller mindre relevante avhengig av tiltakets beskaffenhet og formålet med den enkelte evalueringen.

Den visuelle modellen gir grunnlag for å utdype en rekke betingelser som må oppfylles for å lykkes med innføringen og intensjonen med tiltaket (de fargede boksene), samt hvilke underliggende mekanismer som bidrar til at betingelsene oppfylles i de ulike fasene av innføringsprosessen (pilene/de grå boksene). For konkrete eksempler på programteorier, vises det til Kapittel 2 (Figur 2, Figur 3 og Figur 4).

Programteorien kan sees på som et dynamisk rammeverk som hele tiden justeres og oppdateres i løpet av ITS-tiltakets levetid. I et anvendt perspektiv, fungerer kunnskapsbyggingen optimalt, når resultater fra evalueringsstudien tas inn i beslutninger og prioriteringer underveis (feedback loops). Slik kan tiltaket justeres for å optimalisere effektene i de ulike fasene, basert på kunnskapen som genereres om de relevante årsakssammenhengene i programteorien.

Programteorien består av to sentrale dimensjoner som begge har betydning for tiltakets endelige utfall:

Implementeringsteorien – hvordan skal tiltaket innføres?

Implementeringsteorien er en anbefalende (preskriptiv) fremstilling av hvilke betingelser som må være oppfylt for at tiltaket skal avstedkomme ønsket effekt og virkning. Denne beskrivelsen er knyttet opp mot selve implementeringen av tiltaket, og viser hvordan prosjektet er konstruert med et logisk resonnement for å aktivere endringene i neste punkt. I Figur 9 illustreres dette ved de aktivitetene som skjer i selve boksene; rammebetingelser som må være oppfylt, tiltakets kapasitet (bruksområde, funksjonalitet og ytelse), umiddelbare effekter, samt de langsiktige virkningene man forventer å realisere. Et eksempel på implementeringsbetingelser kan være at IKT-tjenesten forutsetter pålitelig og rask nettilkobling og en brukervennlig mobilapplikasjon. Evalueringen kan da innebære en vurdering eller måling av om denne infrastrukturen er tilstrekkelig til stede for at tjenesten skal fungere optimalt.

Endringsteorien – hvorfor vil ITS-tiltaket forårsake en endring?

Endringsteorien er en beskrivende (deskriptiv) modell av de underliggende mekanismer som må inntreffe for at tiltaket skal medføre en endring. I Figur 9 illustreres dette ved de reaksjoner som skjer i pilene mellom boksene. Programteorien kan inneha ulike endringsteorier for ulike faser av tiltaket og for ulike målgrupper, men her inngår typisk faktorer som sosial aksept og brukeropplevelse. Endringsteorien er gjerne tuftet på samfunnsvitenskapelige teorier (f.eks. psykologiske, sosiale eller økonomiske modeller). Dersom man skal evaluere en IKT-tjeneste som forventes å påvirke trafikantenes reisemiddelvalg, vil man for eksempel benytte allerede kjente og velprøvde atferdsteorier for å vurdere sannsynligheten for at man oppnår en faktisk endring hos trafikantene.

Når programteorien er etablert, danner den grunnlaget for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for modellen. Rammeverket gir således en struktur for evalueringen, ved å identifisere kausale relasjoner og synliggjøre hull i kunnskapsgrunnlaget. Innledningsvis kan programteorien være forholdsvis lite detaljert og med mange kunnskapshull. Programteorien er imidlertid en foreløpig læringsramme som justeres og oppdateres etter hvert som sammenhengene i modellen

bekreftes eller avkreftes. Resultatet av evalueringen blir dermed en mer sofistikert programteori som *forklarer hvordan tiltaket virker* basert på økt kunnskap og forståelse av systemet som studeres.

Trinn 4: Evalueringskriterier

Evalueringskriteriene gjenspeiler formålet med evalueringen (*hva man ønsker å etablere kunnskap om*) og hvordan nytten av tiltaket skal vurderes. Dette inkluderer typisk kriterier knyttet til relevans, måloppnåelse, bærekraft, produktivitet og effekter/virkninger. I denne fasen defineres hva evalueringen skal gi kunnskap om, med utgangspunkt i tiltakets programteori. Den visuelle modellen (programteorien) synliggjør kunnskapshull og vil også være et hjelpemiddel i prioritering av evalueringsressurser mellom ulike problemstillinger.

Evalueringskriteriene må reflektere hvordan tiltaket er forventet å påvirke ulike målgrupper i henhold til behovsanalysen. Eksempler på relevante evalueringskriterier⁵:

- *Produktivitet*: Hvor effektivt kan man omsette ressurser til leveranser i form av resultatmål for kostnad, tid og kvalitet (herunder både teknisk kvalitet og tjenestekvalitet)?
- *Måloppnåelse*: I hvilken grad vil tiltaket bidra til at man oppnår ønskede effekter og oppfyller målsettinger innen eksempelvis trafiksikkerhet, trafikkavvikling, komfort, miljø og mobilitet?
- *Andre virkninger*: I hvilken grad vil tiltaket forårsake utilsiktede effekter (positive og negative) for ulike aktører?
- *Relevans*: I hvilken grad er tiltaket relevant med hensyn til politiske prioriteringer, etterspørsel i markedet og interessegruppers behov?
- *Levedyktighet*: I hvilken grad vil de positive effektene vedvare i hele tiltakets levetid med utgangspunkt i økonomisk, miljømessig og politisk (fordelmessig) aksept?
- *Samfunnsøkonomisk effektivitet*: I hvilken grad vil tiltaket bidra til samfunnsøkonomisk lønnsomhet eller formålseffektivitet (sammenheng mellom måloppnåelse og ressursbruk).

Evalueringskriteriene uttrykkes i form av forskningsspørsmål eller hypoteser som videre brytes ned til et sett målbare indikatorer, og som i neste omgang legger føringer for valg av evalueringsmetode og nødvendig datainnsamling. Det er viktig at formålet med evalueringen er avklart mellom involverte aktører og at det stilles forskningsspørsmål som faktisk kan besvares med hensyn til datatilgang, egnet metode og gjennomførbart evalueringsdesign.

Trinn 5: Evalueringsdesign

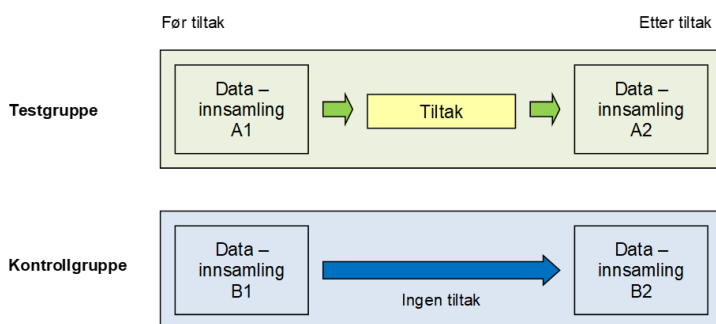
Evalueringsdesignet beskriver evalueringens *fremgangsmåte* med hensyn til metodevalg, gjennomføring av datainnsamling, og analyser. Den teoribaserte evalueringsprosessen er metodenøytral og gir ikke føringer for spesifikt valg av metodeoppsett og analyseteknikk.

⁵ Eksempelene er basert på OECDs evalueringsmodell, videreutviklet av forskningsprogrammet Concept ved NTNU.

Evalueringsdesignet defineres derfor med utgangspunkt i de metodene som er best egnet til å besvare de relevante forskningsspørsmålene, ut ifra formålet med evalueringen og tiltakets kompleksitet og teknologiske modenhet. Evalueringen kan i utgangspunktet inkludere og kombinere flere metoder og analyser (både kvalitative og kvantitative), og det forutsettes at evalueringsdesignet utformes på bakgrunn av kunnskap om de ulike metodenes styrker og svakheter.

Det finnes i utgangspunktet tre hovedtyper evalueringsdesign, som alle kan benyttes i deler av den teoribaserte evalueringsprosessen:

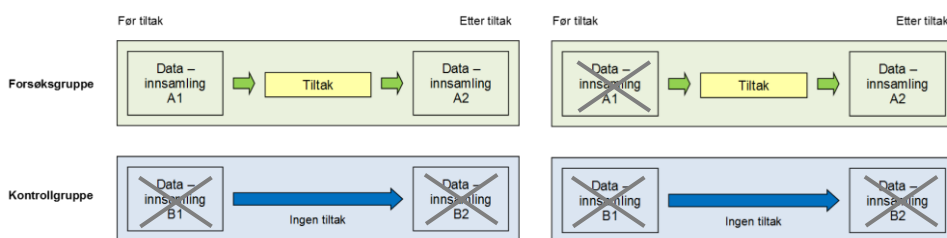
Randomisert eksperimentelt design: Innebærer at forsøkspersoner fordeles tilfeldig i en testgruppe og kontrollgruppe som sammenlignes både før og etter innføring av et tiltak, som vist i Figur 10. Dette er en rigid metode som blant annet benyttes i kontrollerte laboratorie-eksperimenter innenfor medisinsk forskning, og som ofte refereres til som "gullstandard" for metodeoppsett.



Figur 10: Det eksperimentelle evalueringsdesignet

Kvasi-eksperimentelt design: Følger det eksperimentelle metodeoppsettet med testgruppe og kontrollgruppe, men uten at forsøkspersoner fordeles tilfeldig i de to gruppene.

Ikke-eksperimentelle design: Innebærer at tiltaket 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 eller fordi tiltaket berører hele befolkningen. Andre varianter omfatter rene etterundersøkelser, uten tilstrekkelig data om før-situasjonen.



Figur 11: Eksempler på ikke-eksperimentelle forskningsdesign, i form av hhv. en enkel før- og etterundersøkelse (til venstre) og en ren etterundersøkelse (til høyre).

Et ikke-eksperimentelt evalueringsdesign krever at det etableres teoretisk kunnskap om og forståelse av hvordan tiltaket forventes å forårsake effekter i en målgruppe (årsakssammenheng), for eksempel ved at evalueringen struktureres rundt en definert programteori. *Theory of change* og *realistisk evaluering* er eksempler på slike teoribaserte fremgangsmåter.

Innenfor samfunnsfaglige forskningsprosjekter benyttes randomiserte eksperimenter blant annet i kjøresimulator eller intervjustudier, da det er vanskelig å oppnå dette designet i evalueringer som foregår "i den virkelige verden". Effektevalueringer er derfor gjerne preget av ulike kvasi- eller ikke-eksperimentelle studier, hvor den indre validiteten svekkes jo lengre bort man beveger seg fra det ideelle randomiserte kontrollerte designet. I tilfellene hvor grunnlaget for eksperimentelle analyser ikke er til stede, vil den teoribaserte tilnærmingen øke validiteten i et ikke-eksperimentelt design, gjennom å knytte forskningen opp mot teori (programteorien).

Tiltakets årsakssammenhenger vil kunne testes empirisk med ulike metoder og på ulike tidspunkt i innføringsprosessen. Eksempelvis vil måling av faktiske konsekvenser kreve at tiltaket har fungert stabilt over en lengre driftsperiode. Det presiseres at selv om den overordnede evalueringsmetodikken er basert på en teoribasert prosess, innebærer dette likevel en fleksibilitet i metodevalg som åpner for å kombinere eksperimentelle og ikke-eksperimentelle forskningsdesign, og både kvantitative og kvalitative metoder, for analyse av ulike forskningsspørsmål og elementer i programteorien.

Tabell 5 på neste side viser en oversikt over relevante betingelser ved valg av evalueringsdesign og analysemetoder, med utgangspunkt i de tre fremgangsmåtene som er presentert i dette heftet; (kvasi-) eksperimentelt, *theory of change* og realistisk design.

Tabell 5: Grunnlag for valg av evalueringsdesign og metode

Evaluerings-design	(Kvasi-) eksperimentelt	Teoribasert (ikke-eksperimentelt)	
		Theory of change	Realistisk
Metode	Kontrafaktisk analyse med statistiske metoder	Analyse av programteori	Analyse av endringsmekanismer
Formål	Bevise effekter	Finne effekter	Forklare effekter
Eksempel på forsknings-spørsmål	<i>Har man oppnådd forventet effekt i løpet av prosjektperioden?</i>	<i>Ex ante: Hvordan kan tiltaket bidra til å utløse effekter?</i> <i>Ex post: Har man oppnådd effekter og var det tiltaket som førte til disse?</i>	<i>Ex ante: Hva virker, for hvem, i hvilken kontekst og på hvilken måte?</i> <i>Ex post: Hvilke endringer har skjedd og hvordan har de oppstått?</i>
Tidspunkt for evaluering	Etter implementering	Gjennom hele innovasjonsprosessen: Før, underveis eller etter innføring.	
Egenskaper ved tiltaket og omgivelser	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, få andre forklaringer på endring og robuste hypoteser	Langsiktig perspektiv på realisering av effekter Forventer små, kvalitative eller indirekte effekter Tiltaket er teknologisk umodent og/eller forventes å utvikle seg over tid Tiltaket innebærer komplekse årsakssammenhenger og/eller situasjonsavhengige effekter Dynamiske omgivelser og/eller variasjoner i implementeringen av tiltaket	
Metodiske begrensninger og muligheter	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).

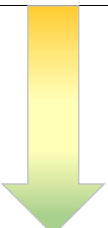
Trinn 6: Datainnsamling

Datainnsamlingen danner grunnlaget for analysene som skal gjøre det mulig å besvare forskningsspørsmålene, gjennom å teste antakelsene (hypotesene) som ligger til grunn for programteorien i trinn 3. Datainnsamlingen må derfor planlegges i relasjon til *evalueringskriteriene* som er definert i trinn 4 og *evalueringsdesignet* som er beskrevet i trinn 5. I trinn 6 utarbeides dermed en operativ plan for datainnsamlingen som identifiserer *hvilke* data som skal samles inn, *hvordan* dataene skal samles inn, samt en *tidsplan* for datainnsamlingen.

I planleggingen av datainnsamlingen bør det inngå en vurdering av datakvalitet og egnethet for bruk i evalueringsdesignet som er beskrevet i trinn 5. Hva som er praktisk

mulig å innhente av data vil være nært knyttet til tiltakets teknologiske modenhet, slik den er vurdert i trinn 1, og ikke minst til evalueringens rammebetingelser i form av tilgjengelige ressurser. I praksis er det sjelden mulig å innhente et komplett datagrunnlag med alle datatyper og av det omfanget som man ideelt ønsker seg. Det er derfor nødvendig å foreta en kunnskapsbasert prioritering av de mest relevante datatyper med riktig kvalitet i henhold til evalueringsdesign og med kapasitet til å besvare de viktigste forskningsspørsmålene.

Figur 12 viser et eksempel på hvordan et utvalg datainnsamlingsmetoder kan relateres til ulike faser av innovasjonsprosessen. Eksempelet viser datainnsamling for evaluering av ITS-tjenester i EU-prosjektet NordicWay (Lervåg, 2016).

TRL	Innovation Phase	Data collection method
1, 2, 3	Concept development	 Checklist Delphi/foresight Modeling Simulation Laboratory testing Field trials Impact monitoring
4, 5, 6	Proof of principle	
7, 8, 9	Proof of performance	
All	User needs and user acceptance	Interviews and surveys

Figur 12: Sammenheng mellom tiltakets teknologiske modenhet og aktuelle datainnsamlingsmetoder

Innovasjonsprosessen er her representert ved nivåene på TRL-skalaen, som tidligere presentert i Figur 8. Som indikert i figuren er det ikke et entydig forhold mellom TRL-nivå og datainnsamlingsmetode, derimot anbefales det å kombinere flere datatyper for å dekke alle evalueringskriterier og indikatorer.

Evaluering av IKT-tiltak kan i mange tilfeller dra fordel av at IKT-tjenestene i seg selv samler inn en rekke datatyper som tidligere ikke har vært tilgjengelig. Dette åpner for analyser og evalueringer som i større grad kan forklare og dokumentere hvordan IKT-tjenesten utløser en rekke mekanismer og effekter i sine omgivelser. Det stiller imidlertid også større krav til den som skal utføre analysene, fordi man nå står overfor muligheter og datatyper som man har begrenset erfaring med å håndtere (f.eks. big data, usikker datakvalitet, personvern hensyn og behov for nye analysemetoder eksempelvis basert på kunstig intelligens og maskinlæring).

Trinn 7: Analyser

Valg av analysemetoder avgjøres i realiteten ved utforming av evalueringsdesign og planlegging av datainnsamling. Hvilke analyser som prioriteres vil i stor grad avhenge av forskningsspørsmålene som er definert i en tidligere fase av evalueringsprosessen. For mange innovative tjenester har man i utgangspunktet begrenset kunnskap om effektbildet, og det vil som regel være nødvendig å kombinere både kvantitative og kvalitative analysemetoder for å forstå fullt ut hvordan tiltaket fungerer. De kvalitative

analysene bidrar gjerne til å identifisere relevante effekter og virkninger (både de som er forventet og eventuelt utilsiktet), og til å skape forståelse for de mekanismene som utløser disse effektene (f.eks. implementeringsstrategi, motivasjonsfaktorer og brukeraksept). De kvantitative analysene gir i større grad kunnskap om effektens størrelse og konfidensintervall, gjennom mulighet for statistisk testing.

Hvilket beviskrav som stilles med hensyn til vurdering av kausalitet avhenger av evalueringsdesignet som er valgt i trinn 5. Mens den eksperimentelle metodikken legger kontrafaktiske analyser av testsituasjoner med og uten tiltaket til grunn for tolkningen av nytteeffekter, har den teoribaserte evalueringen i større grad en mekanistisk tilnærming med fokus på å teste de ulike sammenhengene i programteorien. I praksis vil større evalueringer som regel kombinere både eksperimentelle metoder i studier av teknisk ytelse og kvalitet, atferdsendringer og effekter der dette er mulig å måle, og teoribaserte fremgangsmåter som gir kunnskap om årsakssammenhenger, rammebetingelser og eksterne faktorer som påvirker i hvilken grad man vil lykkes med innføring av tiltaket i ulike situasjoner.

Siden det foreligger lite empiri om nytten av innovative tiltak, bør analysene også inkludere vurderinger av usikkerhet og følsomhet for endringer i tiltaket eller omgivelsene: Hva betyr det om antakelsene som ligger til grunn for tiltakets programteori er feil? Hvilket utslag gir det om størrelsen på brukergruppen (markedsutbredelse) blir mindre enn forutsatt? Og hvordan endres lønnsomheten dersom teknologien forbedres vesentlig?

For de fleste evalueringsformål vil det være relevant å gjøre en vurdering av samfunnsøkonomisk lønnsomhet i en eller annen form. I offentlig sektor er dette tradisjonelt knyttet opp mot nyttekostnadsanalyser, men det finnes også andre tilnærminger som i større grad håndterer og inkluderer kvalitative variabler som ikke kan uttrykkes monetært, eksempelvis multikriterieanalyser. Den teoribaserte evalueringsmetodikken legger ingen begrensninger til valg av analysemetode, utover at den som gjennomfører nytteevalueringen må være kjent med de premisser og forbehold som ligger til grunn for ulike metodevalg.

Trinn 8: Rapportering av resultater

Resultatene fra evalueringsstudien bør forklares og presenteres i en form som beskriver hvordan evalueringen bidrar til utvikling av ny kunnskap om tiltaket som studeres, samt også bidrar til å gjøre kunnskapen anvendbar for fremtidige implementeringer. Det anbefales at resultatrapporten som et minimum inneholder en beskrivelse av grunnlaget for beslutninger i de ulike trinnene i evalueringsmetodikken:

- a) Beskrivelse av selve tiltaket og nødvendig infrastruktur, hvilke IKT-tjenester og virkemidler som inngår i tiltaket, teknologisk funksjonalitet, begrensninger og modenhet, teknisk kvalitet og ytelse, samt brukergrensesnitt.

- b) Beskrivelse av tiltakets kontekst og målsettinger, i form av rammebetingelser og lokal implementeringsstrategi, med blant annet bruksområde, målgrupper og problemstillinger/behov som skal dekkes.
- c) Beskrivelse av programteorien som ligger til grunn for evalueringen; tiltakets hensikt og hvordan det virker (forventning om sammenheng mellom innsats og effekter/virkninger).
- d) Beskrivelse av evalueringskriterier og forskningsspørsmål som ligger til grunn for analysene.
- e) Beskrivelse av evalueringsdesign, metodevalg og datainnsamling, inkludert eventuelle avvik og begrensninger.
- f) Beskrivelse av analyseteknikk og beviskrav (f.eks. statistisk signifikans), samt analyseresultater.
- g) Vurdering av usikkerhet og følsomhet.
- h) Overførbarhet til andre situasjoner og implementeringer.

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

Diskusjonen av resultatene må gi grunnlag for å vurdere om tiltaket er realistisk og hensiktsmessig innenfor det bruksområdet som er studert. Dette innebærer en drøfting av hvordan tiltaket fungerer på ulike organisatoriske nivå:

- *Operasjonelt*: Fungerer tiltaket effektivt? Hvordan kan eventuelle justeringer bidra til optimalisering av teknisk og funksjonell ytelse?
- *Taktisk*: Realiserer tiltaket de tiltenkte effektene for målgruppen? Har man avdekket uønskede effekter? Hvilke betingelser og eksterne faktorer påvirker graden av måloppnåelse?
- *Strategisk*: Er tiltaket relevant og nyttig i et samfunnsperspektiv?

Referanser

- Berger, N. P., & Vrangbæk, K. (2011). Kan MTV styrkes ved brug af realistisk evaluering? Udfordringer, barrierer og muligheder ved brug af realistisk evaluering til vurdering av komplekse behandlingsteknologier. *Tidsskrift for Forskning i Sygdom og Samfund*, 15, 55-76. Danmark.
- Blamey, A., & Mackenzie, M. (2007). Theories of Change and Realistic Evaluation. Peas in a Pod or Apples and Oranges? *Evaluation*, 13(4), 439-455.
- Bredgard, T., Salado-Rasmussen, J., & Sieling-Monas, S. M. (2016). *Virkningsevaluering. Evaluering af offentlig politik og organisationer*: Hans Reitzels Forlag.
- Chen, H. T. (1990). *Theory-Driven Evaluations*: Sage Publications.
- Chen, H. T. (2005). *Practical Program Evaluation. Assessing and Improving Planning, Implementation and Effectiveness*: Sage Publications.
- Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.), *Evaluation von Programmen und Projekten für eine demokratische Kultur*.
- Dahler-Larsen, P. (2013). Evaluering i Norge set udenfra. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Funnel, S. C., & Rogers, P. J. (2011). *Proposeful program theory: Effective use of theories of change and logic models*: ISBN: 978-0-470-47857-8.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*.
- Gjerstad, B., & Bayer, S. B. (2012) *Evaluering av pilotprosjekter for samkjøring i Oslo og Bergen*. Stavanger: International Research Institute of Stavanger (IRIS).
- Government of Canada. (2012). *Theory-Based Approaches to Evaluation: Concepts and Practices*. Treasury Board of Canada Secretariat.
- Griffin, G., Wood, N., Stoeltje, G., Simek, C., Jones, N., & Ettelman, B. (2015) *Real-Time Ridesharing Technology to Support Differential Tolling by Occupancy. Evaluation of 2014 Pilot and 2015 Extension*. Central Texas Regional Mobility Authority.
- Hills, D., & Junge, K. (2010). *Guidance for transport impact evaluations. Choosing an evaluation approach to achieve better attribution*. Tavistock Institute/ UK Department for Transport
- Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28, 167-172.
- Karlsen, T. K., & Jentoft, N. (2013). Programteori versus samfunnsvitenskapelig teori. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Lervåg, L.-E. (2016). *NordicWay Evaluation Handbook*. NordicWay.
- Lervåg, L.E. (2021). *Evaluering av intelligente transportsystemer. Utvikling av evalueringsmetodikk basert på tiltakets programteori*. Doktoravhandling hos NTNU, 2021:103. Fakultet for ingeniørvitenskap, institutt for bygg og miljøteknikk.
- Lervåg, L.E., Foss, T., Jenssen, G.D. og Lillestøl, P.J. (2021). *SmartFeeder: Hvordan kan smarte tilbringertjenester styrke kollektivtransporten? Erfaringer fra utprøving av selvkjørende minibusser i Norge*. SINTEF Rapport 2020:00149.
- Marchal, B., Van Belle, S., De Brouwere, V., & Witter, S. (2013). Studying complex interventions: reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco. *BMC Health Services Research*, 13(1), 469.
- Pawson, R. (2013). *The Science of Evaluation. A Realist Manifesto*. London: Sage Publications.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage Publications.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2003). *Evaluation: A systematic approach* (6th edition). Sage Publications.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2007). *Evaluation. A systematic approach* (7 ed.): Sage.

- Sanderson, I. (2002). Evaluation, policy and evidence-based policy making. *Public Administration*, 80(1), 1-22.
- Sverdrup, S. (2002). *Evaluering. Faser, design og gjennomføring*. (2 ed.): Fagbokforlaget Vigmostad & Bjørke AS.
- Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary Evaluation: Implication for evaluation, researchers, practioners, funders and the evidence-based program mandate. *Evaluation and Program Planning*, 45, 127-139.
- van der Knaap, P. (2004). Theory-based Evaluation and Learning: Possibilities and Challenges. *Evaluation*, 10(1), 16-34.
- Volden, G. H. & Samset, K. (2017). Statlige investeringstiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene. Concept-rapport 52.
- Weiss, C. (1998). *Evaluation: Methods for Studying Programs and Policies* (2 ed.): Prentice Hall.
- Weiss, C. (Ed.) (1995). *Nothing as practical as good theory; Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families*: The Aspen Institute.

Temahefter fra Concept-programmet

Nr.	Tittel	Forfatter	Utgitt
1	Fleksibilitet i prosjekter – et tveegget sverd	Nils Olsson	2009
2	På sporet av relevans og levedyktighet	Ole Jonny Klakegg	2010
3	Gjøre det selv eller betale andre for jobben – Byggherrens valg av kontraktstrategi i bygg- og anleggsprosjekt	Ola Lædre	2012
4	Kostnadsestimering under usikkerhet	Frode Drevland	2013
5	I riktig retning. Prosjekters mål og målstruktur	Knut Samset	2014
6	Prosess for kostnadsestimering under usikkerhet	Olav Torp Frode Drevland Kjell Austeng	2015
7	Mulighetsrommet. Utgangspunktet for et godt konseptvalg	Knut Samset	2016
8	Økonomisk verdsetting av liv og helse	Rune Elvik	2018
9	Fremsynsmetoder	Tore Sager og Knut Samset	2019
10	Mandater for KVVU-arbeid	Knut Samset	2020
11	Problemanalysen	Knut Samset	2020
12	Tidlig involvering av entreprenør	Paulos Abebe Wondimu	2020
13	Prosjektevaluering	Knut Samset	2021
14	Vernede bygninger i konseptvalgutredninger. Vurdering av nytte og kostnad for kulturhistoriske bygninger.	Mari Oline Giske Stendebakken	2021
15	Utredninger og rapporter. Bedre beslutningsunderlag med mindre detaljer	Knut Samset	2021
16	Insentiver og statlig finansiering. Hvordan unngå perverse insentiver i investeringsprosjekter som initieres lokalt	Knut Samset	2021
17	Grunnlaget for ex ante-evalueringer av nettoringvirkninger fra transportinvesteringer i Norge	Rasmus Bøgh Holmen	2021
18	Teoribasert evaluering. Rammeverk for evaluering av innovative og komplekse tiltak	Lone-Eirin Lervåg	2021

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

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

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



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

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

ISSN: 1891-5620 (papirversjon)
ISSN: 1891-5655 (nettversjon)
ISBN: 978-82-8433-014-3 (papirversjon)
ISBN: 978-82-8433-015-0 (nettversjon)