



VEJLEDNING OM AT EVALUERE INNOVATIVE TILTAG

PETER DAHLER-LARSEN

MED FORORD AF COI

**VEJLEDNING OM AT EVALUERE
INNOVATIVE TILTAG**

2016

Publikationen kan frit refereres
med tydelig kildeangivelse

COI

Center for Offentlig Innovation
Købmagergade 22
1150 København K
61 81 31 10
coi@coi.dk
www.coi.dk

Udgivet af

Center for Offentlig Innovation

Forfatter

Peter Dahler-Larsen
med forord af COI

Design

BGRAPHIC

Tryk

Rosendahls A/S

ISBN Elektronisk

978-87-997555-7-8

ISBN Trykt

978-87-997555-6-1

COI'S FORORD

Innovation skal vise og bevise sit værd – her er systematisk *evaluering* en vigtig ingrediens. Derfor vil Center for Offentlig Innovation (COI) understøtte, at evalueringer bliver mere attraktive, tilgængelige og overkommelige for de innovationsaktive i den offentlige sektor.

Evaluering

En systematisk bedømmelse af koncipering (idé), gennemførelse, organisering, præstationer og udfald af et tiltag. Bedømmelsen er retrospektiv i den forstand, at der ses på noget, der er fastlagt eller foregået og på resultaterne heraf – bedømmelsen er ikke en forhåndsvurdering af de enkelte elementer. Bedømmelsen sigter på at spille en rolle i praktiske handlingssituationer undervejs i udviklingen og efterfølgende i anvendelsen af tiltaget. Samt i vurderinger af, om der er skabt værdi, der står mål med den indsats, som realiserede tiltaget.

Skal man så evaluere, før man overhovedet kan tale om en innovation? Ja, for ellers ved man ikke, om der er skabt værdi. Men som i andre forhold må også ressourcerne brugt til evaluering stå mål med gevinsten.

Proportionalitetsprincip

En vis sammenhæng mellem ressourcerne til evalueringen og ressourcerne anvendt til det potentielt innovative tiltag, der evalueres. Et mindre omfang er ikke lig en lavere kvalitetstandard. Er der få ressourcer til rådighed, er der muligt endnu større grund til at anvende dem godt.

Innovation

Offentlig innovation er at gøre noget nyt (bryde med det vante i organisationen) i den offentlige sektor, som skaber værdi. En ny eller væsentligt ændret måde at forbedre arbejdspladsens aktiviteter og resultater på. Det behøver blot være nyt for organisationen selv – andre kan have udviklet innovationen eller taget den i brug før.

Innovation er defineret ved tre faktorer: Nyt, gennemført og værdifuldt. Idéer alene tæller ikke. *Værdien* kan vise sig på flere måder og ofte i kombination. Eksempelvis kan øget effektivitet optræde side om side med højere kvalitet. Uden ibrugtagning og uden værdi – ingen innovation.

Værdi

I den offentlige sektor skabes værdi på mange forskellige bundlinjer. Det kan være forbedret kvalitet, forøget effektivitet, forøget medarbejldertilfredshed eller forøget indsigt i eller indflydelse på opgaveløsningen for borgerne.

Derfor må evalueringens omfang selvfølgelig tilpasses opgaven – der er behov for et *proportionalitetsprincip*.

Potentiel innovation

Noget, der kan vise sig at opfylde kravene til at være en innovation – nyt, gennemført og værdifuldt.

Innovationer er resultater af processer. På et tidspunkt bliver aktørerne bevidste om, at arbejdet handler om en *potentiell innovation*. Allerede hér er det værdifuldt at begynde evalueringsarbejdet. Så kan man bedre lære undervejs og måske tilpasse processen – eller selve målet. Og siden bedre vurdere, om den ibrugtagne løsning har skabt værdi. Men ofte kommer evaluering sent eller slet ikke om bord i *innovationsprocessen*.

Som led i at få innovation og evaluering til at nå hinanden, har COI bedt professor ved Københavns Universitet, Peter Dahler-Larsen, om at skrive denne *Vejledning om at evaluere innovative tiltag*. Initiativet er således COI's. Forfatterskabet er Peter Dahler-Larsens. Hensigten er at give flere lyst til og mod på at lave evalueringer af *potentielt innovative tiltag*.

Innovationsproces

Proces, der bevidst sigter på at skabe en innovation. Processen kan i nogle tilfælde have et forudgående forløb, hvor denne bevidsthed ikke var til stede.

Potentielt innovative tiltag

De konkrete handlinger, organiseringsformer eller produkter, der er anvendt for at opnå den potentielle innovation.

Med "potentielt" signaleres, at der endnu ikke kan tales om en innovation. Med "innovative tiltag" signaleres, at der tegner sig noget konkret i form af handlinger eller produkter, der er anvendt for at opnå den potentielle innovation. Der skal være noget at evaluere. Sondringen skaber begrebsmæssig plads til, at en evaluering af og til må fastslå, at det ikke lykkedes at skabe en innovation. I vejledningen bruges for det meste termen "innovative tiltag".

Peter Dahler-Larsen behandler nogle af de udfordringer, som opleves, når potentielt innovative tiltag skal evalueres: Hvordan fastholdes en systematisk tilgang, når tiltaget er komplekst, og når målet ikke står klart fra begyndelsen? Hvordan forholder man sig, når borgere og virksomheder skal inddrages? Hvilke evalueringsmodeller kan man støtte sig til, og hvordan sørger man for, at evalueringen ikke er ude af proportion med tiltaget?

Denne vejledning er kun et startskud. COI samler teoretikere og praktikere inden for evaluering og innovation for at udvikle konkrete og praktiske forslag, der kan gøre det lettere for de innovationsaktive i det offentlige at evaluere potentielt innovative tiltag.

God evalueringslyst!

Center for Offentlig Innovation

**VEJLEDNING OM
AT EVALUERE
INNOVATIVE TILTAG**

PETER DAHLER-LARSEN



FORFATTEREN

Peter Dahler-Larsen, Ph.D, er professor på Institut for Statskundskab, KU. Han er leder af CREME (Center for Resultater, Effekter, Måling og Evaluering).

Han er tidligere præsident for European Evaluation Society og forfatter til "Evaluering af Projekter – og andre ting, som ikke er ting" (Syddansk Universitetsforlag 2013) samt "The Evaluation Society" (Stanford University Press 2012).

INDHOLD

1 VEJLEDNING OM AT EVALUERE INNOVATIVE TILTAG	8
Hvorfor evaluere innovative tiltag?	8
Læsevejledning	9
2 LEDETRÅDE	10
Undgå rygsæksyndromet	10
Tag evaluering ombord tidligt	10
Beskriv problemerne og indsatsen indgående	11
Brug hovedideen i det innovative tiltag som ressource i evalueringen	12
Tænk anvendelse tidligt	12
Viden forudsætter organisering	13
Effekt er et kompromis	13
Hvis indsatsen er kompleks og dynamisk, må evalueringen følge med	14
At evaluere er at argumentere	14
Opbyg evalueringskapacitet	15
3 EVALUERING: SYSTEMATIK, ANVENDELSE OG DOSERING	17
Nulfejlskultur eller eksperimentkultur?	17
Hvad er evaluering?	17
Evaluators perspektiv	17
Fire overordnede hensyn i evaluering	18
Tænk anvendelse i stedet for opfølgning	19
Er evaluering altid på sin plads? Lav en evaluerbarhedsvurdering	21
Proportionalitet, skala og kvalitet	22
4 UDFORDRINGER VED EVALUERING AF DET INNOVATIVE TILTAG	23
Karakteristik af det innovative tiltag	23
Kompleksitet	24
Udfordringer ved at finde evalueringskriterier	26
Når effektvurdering bliver umulig: Evalueringernes Bermudatrekant	28
Evalueringsresistens som udfordring	29
Hvordan evaluering kan medvirke til spredning	30
Lær af erfaringer fra andre områder	30
5 HVILKE SPØRGSMÅL KAN BESVARES MED HVILKE EVALUERINGSMODELLER?	32
Målopfyldelsesevaluering	34
Variansbaseret effektevaluering	34
Virkningsevaluering	35
Implementeringsfejl eller teorifejl?	36
Brugerorienteret evaluering	38
Participatorisk evaluering	39
Ekspertbaseret evaluering	41
Systemisk evaluering	41
6 HVOR KAN MAN LÆRE MERE OM EVALUERING	44

1

VEJLEDNING OM AT EVALUERE INNOVATIVE TILTAG

HVORFOR EVALUERE INNOVATIVE TILTAG?

Der er store forhåbninger til, at innovation kan hjælpe den offentlige sektor med at håndtere velfærdsstatens store udfordringer: stigende forventninger i en tid med begrænset økonomisk råderum. Innovation i den offentlige sektor handler om danskerne, demokratiet, velfærdsstaten og dens fremtid. Evaluering af innovative tiltag har en mening i forhold til dette store billede.

Evaluering kan forbedre et innovativt tiltag ved at fremme læring og udvikling før, under og efter en innovativ proces. Evaluering kan skelne mellem succesrig og ikke succesrig innovation. Før vi spreder en god idé, skal vi vide, om den faktisk er god. Evaluering skaber et vidensfundament, der kan kvalificere arbejdet med innovation. Baggrunden for denne vejledning er en både optimistisk og realistisk tro på evaluering af innovative tiltag. Optimistisk, fordi evaluering kan kvalificere innovative tiltag. Realistisk, fordi tilgange og metoder i evaluering må forholde sig til, hvordan innovative tiltag faktisk udfolder sig i virkeligheden.

Når vi skal evaluere innovationen, er det ofte mest konstruktivt at fokusere på det enkelte konkrete tiltag. Derfor tales i denne vejledning om at evaluere et innovativt tiltag (i modsætning til innovation i bred forstand eller innovation som en generel organisatorisk funktion). Af konkrete innovative tiltag i offentlig sektor er kun 44 % evalueret, når man spørger på et givet tidspunkt, efter at innovationen er implementeret. Det har COI's analyse af innovationsbarometerdata vist (Center for Offentlig Innovation 2015a). De fleste af de evalueringer, der laves, sker internt i den offentlige organisation.

Den offentlige innovation i Danmark kan styrkes, hvis flere innovative tiltag kvalificeres via evaluering. Og det må gerne ske lidt mere systematisk og med en bredere vifte af evalueringstilgange. Måske behøver man ikke evaluere i 100 % af tilfældene. Og måske findes der ingen bestemt perfekt evalueringsmåde. Men hvis man er meget kritisk over for alle evalueringssløsninger, der ikke er perfekte, så ender man med lammelse. Så bliver alt for mange innovative tiltag slet ikke evalueret. Der skal i stedet skabes plads til at støtte det gode samspil mellem innovation og evaluering – uden dog at deponere den mulighed, at evaluering også af og til må fastslå, at ikke alt lykkes.

Denne vejledning skal inspirere, sætte i gang og give tiltro til, at det kan nytte at evaluere det innovative tiltag. Klarhed og struktur i evalueringsarbejdet anbefales. Forinden må man rydde sten af vejen og forebygge blindgyder. Det indgår også i en evaluators arbejde.

LÆSEVEJLEDNING

En vejledning er ikke et politisk dokument, hvor alle skal tilslutte sig brede formuleringer. En vejledning er ikke en metodebog eller en håndbog, der dækker alle eventualiteter. Den giver i stedet gode råd til en ven i nøden, nemlig den praktiker, der har til opgave at evaluere en innovativ indsats. Innovative indsatser er uensartede og finder sted i mange sammenhænge. Derfor kan ingen vejledning skrive sig alt for langt ind i den enkelte praktiske situation. Man kan i stedet give bud på ledetråde og modeller, der kan hjælpe mange ved at give klarhed og struktur i arbejdet.

En vejledning er ikke forskning, men formidling. Derfor finder du kun få, nødvendige litteraturhenvisninger – se endvidere litteraturlisten. Nogle passager og figurer findes også i Dahler-Larsen (2013), som der hermed henvises generelt til.

I del 2 findes mit bud på de ti vigtigste generelle ledetråde for evaluering af innovative tiltag. En ledetråd er et kort bud på, i hvilken retning man skal finde løsninger. Det er et kort bud med en overskrift, der er til at huske. Brug de ledetråde, du kan, og lad resten ligge.

I del 3 findes en oversigtlig fremstilling af evalueringsteori, du har brug for. Der forklares fire hensyn, man altid skal tænke igennem, herunder hvad evalueringen skal bruges til. Jeg giver også et bud på, hvor meget evaluering, der skal til.

I del 4 giver jeg mit bud på nogle typiske udfordringer ved at evaluere innovative tiltag. Det er blandt andet at finde frem til det innovative tiltags særlige karakteristika, og hvilke evalueringskriterier der skal bygges på.

I del 5 gives en oversigt over evalueringsmodeller. De svarer til typiske evaluerings-spørgsmål, man kan stille til et innovativt tiltag. På den måde kommer vejledningen tættere og tættere på det enkelte innovative tiltag. Det er fint, hvis du efter læsningen af del 5 kan beslutte dig for en evalueringsmodel i forhold til dit konkrete innovative tiltag.

I en kort vejledning har jeg prioriteret det stof, som du har mest brug for. Bagest finder du tips om, hvor du kan lære mere om evaluering (del 6) såvel som en litteraturliste.

2 LEDETRÅDE

Med ledetråd mener jeg et godt råd, der i de fleste tilfælde kan inspirere til et vellykket evalueringsarbejde. Men det kræver selvfølgelig dømmekraft at bruge en ledetråd på den bedste måde.

UNDGÅ RYGSÆKSYNDROMET

Rygsæksyndromet handler om, at man putter alt det i rygsækken, man måske får brug for på turen. Konsekvensen er, at rygsækken bliver alt for tung. På vandreture er reglen, at rygsækken inkl. sækkens egenvægt højst må svare til 10 % af vandrerens kropsvægt. På tilsvarende vis skal der være proportionalitet mellem indsats, ressourcer og evalueringskræfter i en organisation. Det kræver prioritering. Man skal ikke lade evalueringsarbejdet fylde op med brainstorm over, hvad det i øvrigt kunne være sjovt at få noget at vide om.

Hvis man ønsker en evalueringsmodel, der ikke kun fokuserer på udvalgte aspekter af evalueringsgenstanden, men i stedet får det hele med, så er det som at forvente et landkort i skalaen 1:1. Det vil være ressourcekrævende og upraktisk.

Det værste indhold i rygsækken er byrder, der er umulige at bære. Har man erkendt, at en bestemt evalueringsløsning er umulig, skal man tage konsekvenserne. Det kan gælde et design af evalueringen, som ikke kan lade sig gøre eller en kortlægning af effekter i fremtiden, som man pr. definition ikke kan skaffe data om. Man kan gøre evalueringsarbejdet en stor tjeneste ved at flytte den mentale energi væk fra det umulige og i stedet fokusere på de bedste løsninger blandt de mulige. Forestillingen om et evidenshierarki kan i den forbindelse være farlig. Det bedst mulige design mister retten til at være bedst, hvis det ikke kan lade sig gøre i situationen. Vælg det bedste af dem, der er mulige i den foreliggende situation. Netop det design skal i rygsækken, hverken mere eller mindre.

Stil et evalueringsspørgsmål, der faktisk kan besvares. Opstil prøvespørgsmål og find ud af, om der findes en egnet evalueringsmetode og et realiserbart design, der kan være en fornuftig ramme om evalueringsarbejdet ud fra det spørgsmål. Check om spørgsmålet kan besvares ved hjælp af data, du kan skaffe. Afklar formålet med din evaluering. Find herunder ud af, om nogen konkret person eller myndighed vil være interesseret i at handle på en måde, der afhænger af svaret på spørgsmålet.

TAG EVALUERING OMBORD TIDLIGT

Tidsfaktoren er afgørende for mange gode evalueringstiltag. Tænk evalueringfunktionen ind i selve det innovative tiltag. En evalueringstankegang kan være med til at indsnævre og fokusere hovedidéen i et projekt. Man kan formulere en tidlig programteori (forestillinger om, hvordan indsatsen virker, se del 5) og udsætte den for logiske spørgsmål. Hvorfor skulle det virke? Er der logiske huller? Kan man tænke sig til bivirkninger, som der bør tages højde for? Trykprøv teorien.

Evaluering kan medvirke til tidligt at sætte gode succeskriterier op, der er værd at styre efter. Derved kan man løbende justere den innovative indsats. Udnyt alle muligheder for tidlig afprøvning. I alt for mange tilfælde indfører man for eksempel et nyt heldækkende IT-system i en organisation på en bestemt dato i stedet for at sikre, at det virker i lille skala først. Man lukker endda ofte for det gamle system, inden det nye har overstået sine børnesygdomme.

Man kan sikre tidlig afprøvning og kvalificering af en innovativ idé ved en mock-up model eller en prototype. Lav forsøg i lille skala. Arbejd med løbende tilbagemelding fra relevante brugere og andre interessenter. Eksperimenter i lille skala har været en klassisk idé i meget evalueringsarbejde. Eksperimenter i lille skala fortjener igen at blive opdyrket.

Hvis evaluering er tidligt ombord, har man flere muligheder for at vælge et godt evalueringsdesign og indsamle gode data, inden tiden løber en af hænde. Mange evalueringsdesigns kræver tidlig planlægning. Selv en simpel før- og eftermåling kræver data før indsats.

Hvis de antagelser, virkemidler, mål og kriterier, som udgør et projekts kerne, ekspliciteres allerede i en tidlig udgave af projektbeskrivelsen, gives evaluering bedre plads.

Jo tidligere, man får evaluering ombord i det innovative arbejde, jo større er chancen for, at evaluering medvirker til at kvalificere det innovative tiltag, jo mindre farlig er udsigten til efterfølgende evaluering, og jo bedre arbejdsvilkår får den, som skal foretage evalueringen.

BESKRIV PROBLEMERNE OG INDSATSEN INDGÅENDE

Innovative tiltag må kunne bidrage til løsningen af konkrete problemer. En indgående beskrivelse af indsatsen er vigtig for at forstå indsatsen bedre og for at formulere, hvilke særlige virkninger man forventer af indsatsen. Succes svarer til problemløsning. Hvis det er klart, kan man fokusere på det aspekt, der kræver evaluering.

Et eksempel: På et universitet udbydes sommerfag i august, hvor studerende arbejder intensivt med et enkelt fag en måned i stedet for at have tre fag over et semester. En beskrivelse af dette tiltag handler om pædagogik, koncentration og engagement. Et andet er, om den intensive arbejdsform frigør andre perioder til praktikophold eller udenlandske studier uden dog at forsinke studiet. Endelig kan man fokusere på fordele og ulemper ved at placere undervisningen i netop august. Ordet sommerfag betyder flere ting. Hver af betydningerne peger på, at det er forskellige problemer, man ønsker at imødegå.

Hvis man skal evaluere et sundhedshus, tænker nogen på bygningen og andre på organiseringen. Men bygningen og organisering er forskellige ting, der skal afhjælpe forskellige problemer.

I mange bureaukratiske organisationer giver man sprogligt sandheden en pænere kjole på. Når en indsats går dårligt, taler man om behovet for en yderligere styrkelse og målretning af indsatsen. Selv om det er godt at formulere positive mål, bør man ikke undlade konkret at beskrive de problemer, hvis afhjælpning udgør en motivation for indsatsen.

En projektbeskrivelse, der abstrakt og positivt taler om "bedre patientforløb" uden at komme med konkrete anvisninger eller succeskriterier, overlader både problemforståelse og målfortolkning til en evaluator, der har det svært nok i forvejen. Hvis evaluator er sat i den situation, kan det blive nødvendigt at spole tilbage til konkrete problemer med medicin-håndtering, akutte indlæggelser og tovholdere, der ikke ved, de er tovholdere. Og dernæst knytte virkemidlerne sammen med afhjælpning af disse problemer. En nødvendig forudsætning er at beskrive problemerne konkret.

Man har ikke altid lyst til at stille skarpe og klare evalueringsspørgsmål til en innovativ indsats, især hvis indsatsen er lidt uklar i kanten. På den anden side er der imidlertid store gevinster at hente: Entydigt positive evalueringresultater er jo nemmere at opnå, hvis indsatsen er klart defineret, og evalueringsspørgsmålene har været tydelige.

BRUG HOVEDIDEEN I DET INNOVATIVE TILTAG SOM RESSOURCE I EVALUERINGEN

For ofte holdes der brainstormmøder over, hvad man kunne ønske sig at vide i evaluering. For ofte skydes der for frit på mulige evalueringskriterier. I stedet bør det innovative tiltags hovedide stå så klart for alle, at den selvfølgelig bliver det afgørende udgangspunkt for evalueringsarbejdet.

Et godt forslag er at udtrykke denne hovedide i form af en programteori (se del 5 om virkningsevaluering), hvad enten man tester og udvikler denne programteori eller blot bruger den som en læringsramme for et komplekst og dynamisk projekt. Programteorien har det godt med mange verber. Aktører og ting er OK, men hvem skal gøre noget, for at det hele bindes sammen og lykkes?

Man bør udtrykke det innovative tiltags hovedide kort og koncist sprogligt og grafisk. Når man gør det, bør der følge logiske evalueringskriterier med. Formålet med nye rengøringsvogne på sygehuset i Ringkøbing er ikke at forbedre personalets livskvalitet, men at reducere antallet af belastningsskader. Formålet med rehabiliterende rygtræning er ikke at reducere antallet af sygedage i almindelighed, men præcist at reducere det antal af sygedage, der kan skyldes rygproblemer.

Selvfølgelig bør evalueringen også fokusere på det innovative tiltags hovedide og på de evalueringskriterier, der er logisk knyttede til denne hovedidé. Nogle innovative tiltag udvikler sig så uforudsigeligt, at det er nødvendigt at skifte hovedideen ud. Men også her gælder det, at både tiltaget og evalueringen af tiltaget bliver bedre, hvis man klart kan sige, hvori både den gamle og den nye hovedide består, og hvorfor det var nødvendigt at skifte hovedideen ud. Læring fremmes, når man er eksplicit.

TÆNK ANVENDELSE TIDLIGT

Opfølgning er vigtig, siger mange, for ellers laves der en masse evaluering, som ikke bliver brugt. Underforstået er, at man glemmer at samle evalueringen op, når den først er lavet. Men ideen om opsamling er en dårlig idé, hvis den forskyder ansvaret til ukendte beslutningstagere i fremtiden. Evalueringen må gerne bruges undervejs. Den må præge, forme, diskutere, oplyse og udfordre. Derfor må den indrettes på at være så stærkt tilstede i interessenters bevidsthed, at en egentlig opfølgning er overflødig. Man kan bedst arbejde for tilsigtet brug af evalueringen, hvis man har et klart billede af, hvad den tilsigtede anvendelse er.

Hvis du skal sende evalueringen ind til ministeriet, så find ud af, om de læser den derinde. Spørg dem. Hvis du skal betjene en styregruppe med en evalueringsrapport, så find ud af, hvordan styregruppen tænker, og hvordan den tager beslutninger. Hvis du konsulterer styregruppen, så find ud af, hvad du vil bruge svaret til. Forklar ikke kravene i dit eget hoved i passiv form med "det forventes". Forlad dig ikke på en uspecifik idé om opfølgning. Formuler tidligt tilsigtet anvendelse for tilsigtede brugere (Patton 2012) og gå efter den.

VIDEN FORUDSÆTTER ORGANISERING

En evaluator kan sjældent skaffe viden alene. En god undersøgelse svarer til sociale aktiviteter mellem mennesker. At tilvejebringe en sammenligningsgruppe kræver samarbejde og organisering. At inddrage ekspertise kræver pleje af kontakter. At opbygge evalueringskapacitet kræver uddannelse. At indsamle data kræver aftaler og medvirken. Manglende forbindelse til mennesker med idéer, teorier, data og kompetencer er manglende evaluering.

Innovative tiltag kræver innovativ organisation. Mange simple forhindringer for innovation og for konstruktiv feedback på innovative tiltag kommer af, at klassiske organisationer fungerer bureaukratisk og siloagtigt. Man må i stedet organisere sig på tværs af linjeorganisationen for at få feedback på tværgående innovative tiltag. Det har man vidst, siden man udviklede idéen om kvalitetscirkler. Men vi er ved at glemme den idé igen, ikke mindst i en tid med kontraktstyring, opsplitting af organisationer, profitcentre osv.

Også i evalueringsarbejdet er verber om at organisere gavnlige. Hvem skal bidrage, formulere, levere, motivere, indsamle, analysere, foreslå, hjælpe? Metodiske værktøjer er verber frem for ting. Man kan trække begrebet organisering af evaluering ned på jorden ved at eksplicite, hvem der hjælper med hvad.

EFFEKT ER ET KOMPROMIS

Mange evalueringer består blot i at checke, om de planlagte aktiviteter er gennemført. Det har ført til kritik. Aktiviteterne er sjældent mål i sig selv, men bør føre til en eller anden form for langsigtet forandring. Man taler f.eks. om at se på outcome.

Mange ønsker sig et outcome, der ligger langt ude i fremtiden og er svært at påvise som en effekt af selve indsatsen. Dilemmaet er som regel, at forhold tæt på det innovative tiltag er nemme at påvise, mens de ønskede effekter er sværere at påvise, fordi de er længere væk i tid og rum. Et godt effektmål er et konkret og praktisk kompromis mellem disse to hensyn.

FIGUR 1
EFFEKT LIGGER I ET KOMPROMIS



Et eksempel: På et universitet gennemfører man et innovativt tiltag på IT-området for at styrke nye pædagogiske former med IT. Aktiviteter såsom lærernes kursusedtagelse i nye programmer er lette at dokumentere, men siger ikke meget om ønskede effekter. Til gengæld er de langsigtede effekter svære at dokumentere. En langsigtet ændring i brugen af IT er resultat af hundrede andre ting end selve det konkrete projekt. Man kan ikke engang måle på brugen af en konkret IT-plattform, for disse udskiftes løbende. Mange lærere søger selv egne nye IT-muligheder. Ændring af pædagogiske former er resultat af en endnu større vifte af inspirationskilder. Effektkriteriet må blive et konkret, praktisk kompromis. I den foreliggende situation vælger man at undersøge, hvor mange lærere der oplyser, at de i kraft af det særlige tiltag har fået inspiration til en vigtig ændring af deres undervisningspraksis, og om de kan give eksempler herpå.

Det er til gengæld et godt succeskriterium, der er værd at sigte efter. For det er nødvendigt at have lærerne med som aktive medspillere, når man skal forny undervisningen. Hvis man arbejder efter, at mange lærere skal sige, at de i kraft af innovationsprojektet har fået inspiration til at ændre deres undervisning på en vigtig måde, så er man direkte i gang med at realisere det innovative tiltags formål. Evalueringsskriteriet kan faktisk holde til, at man sigter efter det i det praktiske arbejde. Derfor er det et smukt kriterium, også selv om det er resultat af et kompromis.

HVIS INDSATSEN ER KOMPLEKS OG DYNAMISK, MÅ EVALUERINGEN FØLGE MED

Hvis det innovative tiltag er så komplekst og dynamisk, at effektvurdering og ansvarsplacering er irrelevant eller umulig, må evalueringen følge med på denne udfordring. Det samme gælder, hvis det innovative tiltag forandrer sig meget under evalueringsprocessen. Så må man organisere sig derefter.

Gode evaluatore anbefaler, at man som evaluator går i løbende samspil med det innovative tiltag og medvirker til hurtige forandringer og forbedringer i "real time". Man vil også søge at fremme et hurtigt og smidigt samarbejde mellem en indsats og dens implementerings-kontekst uden hensyntagen til, om det er indsatsens eller kontekstens skyld, hvordan det går. Det vil ofte ske på baggrund af data, som er hurtigt tilvejebragte, men ikke særligt solide. Det vil kræve, at en kreds af involverede interessenter har råderum til at udvikle det innovative tiltag, alt imens evalueringer giver løbende feedback. Det forudsætter også spillerum til evalueringen, så den kan være en aktiv og improviserende medspiller i dette rum. En sådan evalueringdagsorden er ikke altid accepteret af opdragsgivere, som måske fortsat ønsker en mere klassisk og lineær effektvurdering og ansvarsplacering. Hvilken vej, man går i evalueringsarbejdet, afhænger derfor både af forståelsen af indsatsens kompleksitet og af opdragsgiveres krav til evalueringsarbejdet. Hvis de to er uforenelige, må evaluator sige det højt.

Jeg har rubriceret tanker om evaluering af innovative tiltag under kompleksitet og dynamik under overskriften systemisk evaluering på side 41, f.eks. inspireret af Pattons idé om "developmental evaluation" (Patton 2012).

AT EVALUERE ER AT ARGUMENTERE

Evaluering af et innovativt tiltag indebærer en dialog med udveksling af argumenter. Hvis der ikke er klare mål for tiltaget, hvad er så evalueringsgrundlaget? Har man gode succeskriterier at arbejde med? Hvis man overtager et tiltag fra andre, er det så evidensbaseret

eller på anden måde evalueret? Hvor velbegrundede er de virkemidler, teknologier og organisationsformer, der skal bære det innovative tiltag? Hvis det innovative tiltag er meget specielt, i hvilken henseende er det så specielt? Hvori kan et evalueringsgrundlag bestå? Hvis man hverken har målsætninger eller programteori, hvad har man så?

Evaluerer er argumentatorisk arbejde. Man etablerer præmisser. Man ekspliciterer, forklarer og begrunder. Man drager konsekvenser af præmisserne. Man behandler indvendinger. Man tager højde for deltageres forudsætninger i den konkrete situation. Man bygger bro til noget, som publikum kender og accepterer, samtidig med at man søger at overbevise om noget nyt. Som i retorik afhænger sandhedens øjeblik af den kunstneriske udførelse.

Meget evalueringsarbejde kan fremmes, hvis man ikke bruger energien på at lede efter et optimalt evalueringsværktøj, men i stedet tager på sig at argumentere godt for, at man har valgt det bedste evaluerings tiltag blandt de mulige. Hvis man vil have evaluering anvendt, må man som regel engagere sig i at overbevise andre.

Bliver folk overrasket over resultaterne? Er der konflikter på området i forvejen? Stiller man spørgsmål ved evalueringens kvalitet? Til gengæld kan et stort forandringspres på området hjælpe med til, at selv uperfekte evalueringer kan blive brugt. Alle disse forhold spiller sammen (Lederman 2012). For eksempel øges krav til evalueringens kvalitet, hvis dens resultater er meget overraskende. Man kan mindske elementet af overraskelse, hvis evaluering er i løbende dialog med det innovative tiltag. Hvis man via argumenter fremmer accepten af evalueringen, øger man chancen for, at den bliver brugt.

OPBYG EVALUERINGSKAPACITET

De færreste evaluatore arbejder alene. Evaluering er et organisatorisk anliggende. En organisations evne til at producere og bruge god evaluering kaldes evalueringskapacitet. Mange aspekter indgår heri. Man kan hyre kvalificerede evaluatore eller uddanne dem, man har. Man kan trække på bøger, kurser, foreninger, konferencer. Man kan organisere sig sådan, at evaluatore oplever ledelsesstøtte og nærhed i forhold til beslutninger. En ledelse kan vise, at den tager evaluering alvorligt ved at udvise interesse for evalueringer og ved selv at tage beslutninger ud fra evalueringsresultater.

Sproget i en organisation er vigtigt som basis for evaluering. Man må give plads til et sprog, hvor problemer omtales klart og tydeligt. Man må give plads til, at tiltag kan have forskellige udfald med mere eller mindre succes. Man skal ikke indbygge et positivt udfald i selve det ord, der betegner tiltaget.

En evaluator har tre typer af ressourcer at trække på: Ekspertise, positionsmagt og personlig integritet. Organisationen har en stor indflydelse på evaluators ekspertise gennem ansættelsespolitikken, gennem efteruddannelse, kurser og gennem deltagelse i relevante netværk. Organisationer har indflydelse på evaluators positionsmagt ved at placere evaluatore på et handlekraftigt sted i organisationsstrukturen og ved at tildele evaluering ressourcer og ledelsesmæssig opmærksomhed. Gode arbejdsforhold for evaluatore er også en vigtig del af den organisatoriske basis for evalueringsarbejdet. Konkret betyder det for eksempel, at man også internt i en organisation arbejder på baggrund af et skriftligt kommissorium, så det er klart for alle, hvad en evaluering handler om, og hvad en evaluator forventes at gøre – således at evaluator ikke kommer i klemme ved uklarhed om opgaven.

Evaluators personlige integritet er tillige en ressource i sig selv. Men organisationer har en rolle at spille i forhold til at give gode evaluatore rimelige spilleregler for arbejdet. Det gælder om ikke at skjule væsentlige dagsordener. Det skaber nemlig tvivl om, hvorvidt evalueringen fokuserer på de reelle spørgsmål.

Evalueringskapacitet omfatter også evalueringskultur, herunder holdninger og vaner i forhold til evaluering. Den bedste tilgang til at påvirke holdninger til evaluering er ikke ved at tale meget om evalueringskultur, men ved at udføre selve evalueringerne på en måde, der giver respekt. Det kræver systematik i arbejdet.

3

EVALUERING: SYSTEMATIK, ANVENDELSE OG DOSERING

NULFEJLSKULTUR ELLER EKSPERIMENTKULTUR?

Vi lever i en tid med et spændingsforhold mellem en nulfejlskultur og en eksperimentkultur. I en nulfejlskultur søger man at undgå fejl. I en eksperimentkultur lærer man af de fejl, som uundgåeligt sker. Spændingen mellem de to viser forholdet mellem en offensiv og en defensiv kvalitetsopfattelse. Offensiv kvalitet indebærer søgen efter nye og bedre løsninger for fremtiden under en vis usikkerhed og en vis risiko. Defensiv kvalitet sikrer et vist minimumsniveau via standardisering og kontrol.

Eksperimenter har igennem tiderne haft en stor rolle at spille i forhold til at skaffe ny viden. Eksperimenter giver mulighed for feedback fra den praktiske virkelighed. I evalueringernes barndom var der en tæt alliance mellem eksperimenter og evaluering.

I lyset af spændingsforholdet mellem nulfejlskultur og eksperimentkultur, mellem offensiv og defensiv kvalitet, er udfordringen at sikre et handlerum for innovative tiltag, der muliggør nye og kreative løsninger uden at blive viklet ind i for stramme og rutineprægede krav om procesreguleringer, standarder og afrapportering. Men dydens sti er smal, for de nye tiltag er ikke pr. definition gode, men må dokumentere deres duelighed gennem evaluering.

HVAD ER EVALUERING?

Evaluering er en systematisk retrospektiv bedømmelse af koncipering (idé), gennemførelse, organisering, præstationer og udfald i forbindelse med et offentligt tiltag. Det er en bedømmelse, som tilsigtes at spille en rolle i praktiske handlingssituationer. Definitionen er inspireret af Vedung (2009).

Evaluatoren kan være intern eller ekstern. De to har forskellige arbejdsvilkår. Men i begge tilfælde skal der være tale om godt og gennemtænkt evalueringsarbejde, der lever op til sit formål.

EVALUATORS PERSPEKTIV

Evaluator er betegnelsen på den, der udfører evaluering. Her tænkes på evaluator som en funktion, ikke som en stillingsbetegnelse. Du kan derfor også være evaluator, selv om din stillingsbetegnelse er chef, fagperson, udviklingskonsulent eller andet. Hvis innovative tiltag skal evalueres, må der være nogen, der tager det perspektiv og det medfølgende ansvar på sig.

Hvis du er projektleder på et innovativt projekt, som du selv evaluerer, er du både projektleder og evaluator. I denne dobbelte egenskab har du godt nok selv en interesse i at bruge evalueringen til udvikling af din egen innovative indsats, men andre kan også bruge dit evalueringsarbejde til at danne sig en mening ud fra deres synsvinkler. Som evaluator bringer du dig i spil både i forhold til det innovative tiltag selv og i forhold til andres vurdering af den.

En evaluator er selvfølgelig ikke alene om at få evalueringsarbejdet til at lykkes. Derfor er det også en god idé, hvis ledere, fagpersoner og andre læser denne vejledning for at se det innovative tiltag gennem en evaluators øjne. Det er nødvendigt at klargøre og udrede den organisatoriske situation, så den er klar til evaluering, lige som der skal ryddes op i huset, før man gør rent. Som bekendt kan det godt være en fælles opgave!

FIRE OVERORDNEDE HENSYN I EVALUERING

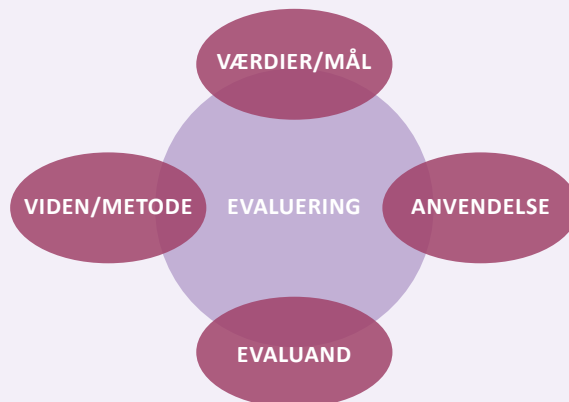
Vi husker, at evaluering er en systematisk retrospektiv bedømmelse af koncipering, gennemførelse, organisering, præstationer og udfald i forbindelse med et offentligt tiltag. Det er en bedømmelse, som tilsigtes at spille en rolle i forhold til praktisk handling. Der er fire dimensioner i denne definition.

- Systematisk betyder en systematisk vidensindsamling ved hjælp af en metode.
- Bedømmelse sker ud fra værdier og mål.
- Forskellige aspekter af offentlige tiltag er genstand for evaluering, i fagsprog kaldet "evaluanden." Vi har lov at evaluere hele vejen lige fra konciperingen (om tiltaget som idé er godt tilrettelagt) til udfaldet (positive og negative effekter).
- Endelig sker evaluering med et anvendeshensyn i sigte.

Figuren kan bruges til overordnet planlægning af en evaluering. En evalueringsopgave skal ikke løses ved, at man kun tænker på en enkelt ting, f.eks. metoden. En god evaluering forholder sig til alle fire dimensioner på en gennemtænkt måde.

Når jeg selv planlægger en evaluering, drøfter jeg derfor med min opdragsgiver de praktiske muligheder og bindinger inden for hvert af disse fire domæner, inden jeg foreslår en evalueringsmodel og en plan. Brug de fire domæner som en huskeliste, således at du ikke overser et væsentligt aspekt af evalueringsarbejdet.

FIGUR 2
FIRE HENSYN I EVALUERING



TÆNK ANVENDELSE I STEDET FOR OPFØLGNING

Blandt de fire hensyn i evalueringsarbejdet er der en tilbøjelighed til at skubbe tanker om anvendelse ud til senere eventuelt under navnet "opfølgning." Jeg anbefaler i stedet, at man tidligt gennemtænker formålet med evalueringen i form af den tilsigtede anvendelse.

Evalueringsarbejdet kan have fire officielle formål/anvendelsesformer:

- evaluering kan medvirke til kontrol og opsyn. Herunder hører politisk og ledelsesmæssigt opsyn med, at offentlige midler bruges i overensstemmelse med, hvad de er bevilget til.
- evaluering kan støtte læring, udvikling og forbedring af innovative tiltag.
- evaluering kan støtte vidensudvikling. Det indebærer, at man på tværs af tid og rum (f.eks. på tværs af kommunegrænser) får viden om, hvordan forskellige innovative tiltag har vist sig at virke.
- sidst, men ikke mindst medvirker evaluering til at sikre, at skatteborgere, for hvis penge det innovative tiltag er sket, bliver informeret om, hvordan det så er gået. I et demokrati er stillingtagen til innovative tiltag vigtig. Innovation er et stort emne i forbindelse med velfærdsstatens fornyelse. Der findes også en del innovationer, som giver anledning til debat af bredere etiske problemer. Det gælder medicinsk teknologi. Det gælder også velfærdsteknologi i plejeopgaver eller til løsning af ensomhedsproblemer.

Der kan være stærke spændinger mellem de forskellige formål/anvendelsesformer, typisk f.eks. mellem kontrolhensynet og læringshensynet.

Hvis der er et tilstrækkeligt råderum til innovation, spiller evaluering gerne med. Evaluering vil gerne være aktive medspillere i dynamiske læreprocesser. De vil også gerne undersøge, f.eks. om nye løsninger er kreative, og hvorvidt de reelt er udviklet, så de imødekommer brugerens behov. Evaluering kan være meget tæt sammenflettet med hele den innovative proces. Men hvis der politisk og økonomisk er et stærkt ønske om kontrol samt håndfaste krav om f.eks. effektivitet, så kan evaluering ikke i sig selv ophæve spændingsforholdet mellem kontrol og læring.

Det er en god idé at gøre sig den primære tilsigtede anvendelse klart. Ellers kan man ikke arbejde for at fremme den. Prøv at undgå ordet "opfølgning", hvis det indebærer, at ansvaret for at bruge evalueringen skubbes ud i fremtiden på en uspecifik måde. Mange organisationer ønsker at tilgodese mange anvendelsesformer samtidig og ender derved paradoksalt med så bredt fokus, at man opnår ingen af dem. Selv om det er svært at prioritere, er det bedre til gengæld at opnå en begrænset og fokuseret anvendelse.

Anvendelsen kommer sjældent af sig selv. En klassisk forestilling om, at man sender en evaluering, der som en billardkugle rammer et eller andet, så noget andet sker, er som regel forældet. Sjældent er det en fremsendelse af en evalueringsrapport via postvæsenet, der udløser en evalueringsanvendelse.

I stedet har forskningen peget på betydning af netværk, samspil og interaktion. Samspil er vigtigt undervejs for at skabe forståelse for evalueringens vigtighed, fokus, metode og tilsigtede anvendelse. Det baner vejen for større accept af eventuelt overraskende resultater såvel som af, at evalueringer kan være brugbare, uden at de behøver være metodisk perfekte. Man kan blive hjulpet på vej af, at der er et stort forandringspres på det område, hvor evalueringen finder sted (Ledermann 2012).

Men anvendelse kræver bearbejde. Hvis man vil fremme evalueringens anvendelse, må man knytte interaktive bånd over tid og man må gide at argumentere godt. Man må være i dialog med den, som skal anvende evalueringen og fastholde denne dialog, indtil anvendelsen er sket.

Hvis der er for mange anvendeshensyn på spil, kan man dele evalueringsarbejdet op i forskellige faser, som hver fokuserer skarpt på en enkelt form for anvendelse.

Her er et eksempel. I et innovativt tiltag om nye patientforløb findes forskellige roller og funktioner i tilknytning til tiltaget, f.eks. designe, finansiere, støtte, implementere og lære. Tillige yder nogen modstand.

Det giver mulighed for at opstille figur 3 over evalueringsomsættere, dvs. aktører, der kan omsætte evalueringen til handlinger af betydning for tiltaget. I dette skema kan man prioritere en bestemt form for tilsigtet anvendelse af en foreliggende fase af evalueringsarbejdet.

I et konkret tilfælde valgte man at prioritere læreprocessen i samspillet mellem hospitalspersonale, forløbskoordinatorer, praktiserende læger og kommuner (illustreret ved en boble). Det gav klare pejlemærker for, hvem der skulle involveres og hvordan i de umiddelbart foreliggende faser af evalueringsarbejdet.

I andre faser kan andre aktører selvfølgelig ikke ignoreres. For eksempel spiller politikere indlysende en stærk rolle i forhold til at finansiere og bakke op om innovative tiltag. Samtidig vil politikere ikke altid forpligte sig til at bruge evalueringsresultater på en bestemt måde.

Afklaring kan hjælpe et stykke ad vejen og kun dertil. Det er ikke evaluatorene, der i sidste ende træffer politiske og ledelsesmæssige beslutninger. Men evaluatorene bør gøre sig klart, hvad de selv mener, at de arbejder for. Og i alle de tilfælde, hvor det hjælper at afklare og fokusere, bør man gøre netop det.

FIGUR 3
INTERESSENER I EVALUERING AF PATIENTFORLØB

	Design	Finansiere	Støtte	Implementere	Lære	Yde modstand
Ministeriet						
Politikere						
Hospitalsledelse						
Afdelinger						
Personalegrupper						
Forløbskoordinatorer						
Praktiserende læger						
Kommuner						
Andre regioner						

ER EVALUERING ALTID PÅ SIN PLADS? LAV EN EVALUERBARHEDSVURDERING

Evaluering koster tid, penge og opmærksomhed. Derfor skal evaluering ikke være rutine. Man skal kun evaluere, hvis man kan sandsynliggøre, at evaluering vil gøre en positiv forskel. Proceduren hertil hedder evaluability assessment (Smith 2005), på dansk nok bedst oversat med evaluerbarhedsvurdering.

Formålet med evaluerbarhedsvurderingen er at bruge lidt ressourcer på at skaffe viden for at afgøre, om der bør bruges flere ressourcer på en mere omfattende evaluering. I stedet for at købe meget viden på en gang med store omkostninger, køber man billigt lidt viden ad gangen for på det grundlag at vurdere næste skridt.

Evaluerbarhedsvurderingen omfatter følgende spørgsmål:

- er den innovative indsats klart beskrevet og defineret? Ellers ved man ikke, hvad man evaluerer. Resultater og konklusioner bliver uklare.
- er der klare forestillinger om mål? Hvis der er en åbenhed i fortolkningen af mål, kan den så afgrænses? Hvis svaret er nej, vil mange af evalueringsressourcerne blive brugt til afklaring snarere end til evaluering.
- er der klare forestillinger om, hvorfor og hvordan det innovative tiltag tænkes at virke? Man får mere ud af evalueringsarbejdet, hvis man kan indkredse og teste sådanne forestillinger op mod virkeligheden. Hvis man har klare forestillinger om, hvorfor indsatsen virker, styrkes tiltroen til, at den bør fortsættes og spredes.
- er allerede kendte implementeringsproblemer luget ud? Hvis nej, bør disse håndteres af den almindelige ledelse og den almindelige drift i organisationen. Man behøver ikke bruge evaluering til at indsamle viden om problemer, der allerede er kendte.
- kan relevante data skaffes til de stillede evalueringsspørgsmål? Hvis data er umulige at skaffe eller ligger for langt ude i fremtiden, så er det en advarsel om, at evalueringskræfter måske spildes. På den anden side er relevante data nemmere at skaffe, hvis man i forvejen registrerer, hvordan det går med innovationen. Derfor fører evaluability assessment ofte til en opfordring til organisationer om at opbygge letfodede, men relevante systemer til indsamling af data, så man uden særlig anstrengelse følger med i, hvordan det går.
- sidst, men absolut ikke mindst: Kan der peges på konkrete personer, hvis beslutninger vil blive påvirket af evalueringen og dens resultater? Når man har opbygget et innovativt sundhedshus ved at samle flere decentrale sundhedsfunktioner, er det ikke fornuftigt at stille et evalueringsspørgsmål, der hedder: "Var det en god idé at samle funktionerne i huset?" – hvis man alligevel ikke har tænkt sig at pille huset fra hinanden igen og spole processen tilbage. Der skal være en logisk sammenhæng mellem evalueringsresultater og beslutninger. Også derfor er det vigtigt at gøre sig den tilsigtede anvendelse klar.

Hvis svarene er negative på alle ovenstående spørgsmål, er det en advarsel om, at evalueringsarbejdet kan vise sig nytteløst. Det kan også være, at evalueringsarbejdet kan have til formål at afklare alle disse ting, eventuelt under inddragelse af mange interessenter, men hvis det er tilfældet, så skal man have ressourcerne hertil, og evaluator skal have et stort handlerum, når så lidt er afklaret fra starten. Derfor bør man reagere, når en evaluerbarhedsvurdering med et negativt udfald lyser som en rød lampe.

Måske kan man massere virkeligheden lidt, sådan at flere af svarene bliver mere positive, og udgangspunktet for evaluering bliver klarere og bedre. Det hedder "afklaring" og er en vigtig opgave for en evaluator. Hvis der er meget begrænsede ressourcer til evaluering, bør man ikke kaste sig over meget omfattende evalueringsopgaver uden først at bruge bare 45 minutter på en evaluerbarhedsvurdering.

Hvis betingelserne viser sig mere eller mindre ufrugtbare, så tænk over, om evalueringskræfterne bruges optimalt, hvor de sættes ind. Det er det, som en evaluerbarhedsvurdering kan gøre for dig. Har du to innovative tiltag, men kun ressourcer til et begrænset evalueringsarbejde, så sæt evalueringen ind over for det tiltag, hvor evaluerbarheden er størst.

Ved at bringe evaluering tidligt om bord kan man styrke evaluerbarheden af et innovativt tiltag. I nogle tilfælde må du lade fortiden fare og til gengæld sikre, at din næste projektbeskrivelse for et innovativt tiltag allerede dér rummer klare svar på de seks spørgsmål i den fremtidige evaluerbarhedsvurdering. Så gør du dit næste innovative tiltag mere klar til evaluering.

PROPORTIONALITET, SKALA OG KVALITET

Evaluerbarhedsvurderingen siger, at man tidligt skal vurdere fornuften ved at evaluere i den konkrete situation. Her tager vi hensyn til evalueringens mulighed for at sige noget fornuftigt, brugbart og præcist om evaluanden. Men vi kan også sætte ressourcerne til evaluering i relation til ressourcerne til evaluanden. Brug et proportionalitetsprincip: Man skal ikke skyde gråspurve med kanoner.

Hvis man i øvrigt har lavet en fornuftig evaluerbarhedsvurdering, kan man afbalancere ressourcerne til evaluering i forhold til samtlige omkostninger ved den innovative indsats. Praktiske tommelfingerregler vil måske sige 2 % til evaluering. Eller 8 %. Eller et sted derimellem. I mange tilfælde kan det være svært at skelne mellem ressourcer til udvikling af evaluering og ressourcer til evaluering, der fører til udvikling. Men alligevel er proportionalitetsprincippet godt at holde sig for øje. Evalueringens skala må afpasses.

Mange forveksler imidlertid skala med kvalitet. Lille betyder faktisk ikke dårlig. Man kan stille gode, klare og præcise spørgsmål i en overkommelig evaluering. Man kan finde de bedst mulige data, der besvarer spørgsmålene. For eksempel er dans for demente et indsatsområde, der måske ikke kræver meget omfattende evaluering. Men også til demens-dans kan man stille skarpe spørgsmål: Har vi tegn på, at de demente nyder at danse? Har vi tegn på, at de bliver frygtsomme, fordi de skal gøre uvante bevægelser, og fordi de er sammen med mange mennesker, de ikke genkender? Er der taget hånd om risikoen for faldulykker og hoftebrud?

Ofte er gode spørgsmål en bedre indgangsvinkel end mange data.

4

UDFORDRINGER VED EVALUERING AF DET INNOVATIVE TILTAG

De overordnede hensyn i evaluering til viden, værdier, anvendelse og evaluanden gælder generelt. I denne del går vi nærmere i clinch med selve det innovative tiltag. Der findes teoretiske definitioner på innovation (Hartley 2005; OECD 2005), som vi ikke vil diskutere her. I stedet ser vi en praktisk situation for os, hvor nogen beder om vores hjælp til at evaluere, hvad de benævner som "et innovativt tiltag". Vi starter derfor med at beskrive tiltaget, så vi ved, hvad vi har med at gøre.

KARAKTERISTIK AF DET INNOVATIVE TILTAG

Innovative tiltag kan byde på særlige problemer vedrørende målbarhed, programmerbarhed og kompleksitet. Derfor er det ofte en god idé forud for evaluering at karakterisere et innovativt tiltag i forhold til de tre dimensioner.

Målbarheden er høj, hvis målsætninger for det innovative tiltag er klare, velbeskrevne og operationaliserbare. Operationaliserbar betyder, at man ved, hvilke kvalitative eller kvantitative data, der vil tjene som tegn på, at målene er indfriet.

I mange tilfælde er der begrænsninger i målbarheden i innovative tiltag, for eksempel fordi man undervejs finder på nye uforudsete applikationer. Et eksempel er, at teflonpanden er et resultat af rumfartsindustriens arbejde med nye materialer til raketter. Det betyder imidlertid ikke, at al målopfyldelse er ligegyldig. Man kan stadig se på, om det lykkedes at udvikle selve raketterne og sende dem afsted sikkert.

Hvis man undervejs i en innovativ proces er nødt til at forlade oprindelige målsætninger, gør det stadig en forskel, om nye målsætninger sættes i samarbejde med den, der finansierer tiltaget, eller om vurderingen af effekter og bieffekter sker i en legitim kreds af interessenter, der foretager en legitim drøftelse, eller om vurderingerne er selvopfundne.

Også hvis der er et vist rum til fortolkning af målsætninger, kan der træffes aftaler om, hvordan det fortolkningsrum udnyttes. Man kan aftale, hvem der foretager den fortolkende vurdering, f.eks. at det er op til brugerne eller til en ekspertvurdering. Også en sådan aftale kan bruges evalueringsmæssigt til at indhegne en eventuel uenighed om graden af målbarhed.

Programmerbarheden er høj, når arbejdsprocesserne til at nå målene er klare, velbeskrevne og operationaliserbare. Programmerbarheden er naturligvis ofte begrænset i innovative processer, netop fordi man skal søge nye veje til opnåelse af målene. Det gælder både i forhold til teknik og i forhold til mennesker.

Der er imidlertid stor forskel på, hvor gode innovatorer er til at begrunde de veje, de vælger. Der er som regel et godt argument for at afprøve en bestemt vej til en løsning, også selv om man ikke er helt sikker på, det er den absolut bedste vej. Inden for eksempelvis biokemiske analysemetoder er der meget godt styr på, hvad den enkelte analysemetode kan, også selv om man måske har tænkt sig at afprøve den på et nyt materiale.

Inden for den pædagogiske verden kan der være usikkerhed eller diskussion om effekterne af bestemte pædagogiske metoder. Men det betyder ikke, at en metode blot vælges tilfældigt. Der er stadig både pædagogiske og etiske overvejelser i metodevalget. Der er eksempelvis enighed om, at man bør undgå pædagogiske metoder, der udsætter børn og unge for fare. Der bør derfor være en vis programmerbarhed i valget af metode, også selv om der er tale om en innovativ indsats.

Kombinationen af målbarhed og programmerbarhed giver en firfeltstabel, der kombinerer egenskaber ved innovative tiltag med de muligheder, der findes for kontrol og evaluering.

FIGUR 4 MÅLBARHED OG PROGRAMMERBARHED

Målbarhed og programmerbarhed (Andersen 2008: 16).

	Høj programmerbarhed	Lav programmerbarhed
Høj målbarhed	I	II
Lav målbarhed	III	IV

Hvis målbarheden er høj, kan man benytte resultatkontrol (I og II). Hvis programmerbarheden er høj, kan man benytte adfærdskontrol (I og III). Hvis hverken målbarheden eller programmerbarheden er høj (IV), må man dyrke klankontrol og socialisering. Man må udvælge og socialisere dygtige medarbejdere, som af egen drift vil søge at realisere målene, uden at man direkte kan kontrollere deres adfærd eller deres resultater.

Evalueringsmæssigt må man i felt IV søge at imitere klankontrol og socialisering gennem en tæt og vedvarende dialog, der udvikler sig i takt med indsatsen. Evalueringen må være dialogisk og dynamisk og i tæt samspil med den indsats, der evalueres. Det er den resterende mulighed, hvis man ikke kan styre via resultater eller via adfærdsregulering.

I mange tilfælde vil tilhængere af et innovativt tiltag placere sig selv i felt IV. I så fald er det en god idé at afklare, om også tiltagets politiske ledere og sponsorer mener, at hverken adfærdskontrol eller resultatkontrol er på sin plads. Og, hvis ja, er det en god idé at tydeliggøre, hvilke principper man i øvrigt mener, der skal karakterisere den gode innovative indsats. Eller hvilke interessenter, der skal have lov til at lave den vurdering. I felt IV må man eksplicitere, hvilke værdier man har tænkt sig at leve op til og "kunstigt" skabe nogle tilbagemeldingsmekanismer, der giver mulighed for løbende at pejle indsatsen i forhold til disse værdier. Hvordan det skal ske, kan man med fordel afklare, når man formulerer, hvad den innovative indsats går ud på. I modsat fald kan det blive svært at forklare, hvad man står til ansvar for, og hvad grundlaget for evaluering skal være.

En lav grad af målbarhed og programmerbarhed (felt IV) kan ikke mindst skyldes kompleksitet.

KOMPLEKSITET

Man kan skelne mellem to slags komplicerethed, teknisk og social (Patton 2012), der tilsammen fører til kompleksitet.

Teknisk komplicerethed handler om graden af sikkerhed i årsags-virkningsforhold. Den tekniske dimension omfatter i bredeste forstand viden om, hvordan teknologier (forstået som virkemidler) virker. Den sociale dimension omhandler samspillet mellem parter med forskellige synspunkter, interesser og målsætninger.

FIGUR 5
KOMPLEKSE INNOVATIVE TILTAG



De to slags komplicerethed håndteres typisk på to forskellige måder. Den tekniske komplicerethed tackles ved at eksperimenterer og ved at inddrage ekspertise. Den sociale komplicerethed håndteres ved dialog, involvering, forhandling, eventuelt inden for rammerne af organisatoriske og politiske strukturer og processer.

De to dimensioner spiller sammen. Et teknisk kompliceret tiltag involverer faggrupper og aktører placeret forskellige steder i organisatoriske landskaber. Teknisk komplicerethed fører organisatorisk komplicerethed med sig. Et allerede eksisterende stykke teknologi kan benyttes til formål, som grundlæggende ændrer sociale relationer. Et eksempel er mobning via sociale medier, som må reguleres i skolen. Et andet eksempel er ændring af forholdet mellem pædagoger, forældre og børn, når forældre kan videoovervåge børns leg i daginstitutionen.

På den anden side kan social komplicerethed også påvirke den tekniske side af en innovation. Organisatoriske roller og funktioner mobiliserer typisk også forskellig faglig indsigt og ekspertise såvel som særinteresser. De to slags komplicerethed forstærker ofte hinanden.

Tiltag findes i skemaet omkring tallene 1, 2, 3 og 4. Innovative tiltag kan være af mange slags, men de mest spændende og udfordrende af dem befinder sig omkring firtallet.

Et eksempel er nye patientforløb i samarbejde mellem hospital, praktiserende læge og kommune. Det er ikke nok at se på det tekniske og medicinsk-faglige indhold i manualer for det gode patientforløb, når det virkeligt svære er at sikre et godt samarbejde mellem de forskellige parter, som har forskellige vaner, fagligheder, interesser og værdier. Netop fordi det innovative tiltag ofte kræver koordinering af mange parter, er selve organiseringen af innovationsprocessen ofte afgørende for tiltagets skæbne og derfor et evalueringstema i sig selv.

Som Patton (2012) bruger terminologien, så fører de to slags komplicerethed tilsammen til kompleksitet. Komplexitet og dynamik hører sammen. Hos ham er det dynamikken, der er det spændende og udfordrende ved område 4.

Nogle kendetegn ved et komplekst tiltag:

- formålet beskrives ved ord, der lyder positivt, men er åbne og omstridte, når de skal omsættes, såsom kvalitet, sikkerhed, bæredygtighed, ligestilling, kreativitet, og, ja, innovation.
- den organisatoriske struktur i tiltaget, strukturen i det problem, der skal løses og strukturen i den måde, man holder folk ansvarlige på, passer ikke sammen på forhånd.
- der er et stærkt indslag af organisation, koordinering, tværfaglighed og/eller samarbejde på tværs af gængse grænser og niveauer.
- der er mange sammenhænge og indbyrdes afhængigheder i tiltaget, og da det tillige er svært at adskille fra sin kontekst, er det præget af uigennemsigtighed (Mikkelsen and Riis 2008: 16).
- det innovative tiltag udvikler sig, så man ikke kan forbinde dets effekt med en stabil årsag.
- tiltagets meningsfuldhed for forskellige aktører betyder meget, fordi deres engagement i projektet er afgørende for projektets udfald.

I et innovativt tiltag kan der således være mere eller mindre kompleksitet. Evalueringsmæssigt kan man forholde sig til kompleksitet på to måder. Den ene er at udvikle programteori, der beskriver hovedindholdet i indsatsen (se virkningsevaluering i del 5) og bruge denne som tilnærmning til at besvare effektspørgsmål, mens andre effekttilgange, som ikke er egnede under omstændighederne, tages ud af evalueringsarbejdet (husk at undgå rygsæksyndromet). En anden er, at man helt opgiver at teste programteori og at placere ansvaret for, hvorvidt indsatsen virker. Så kan man i stedet overveje systemisk evalueringsarbejde (del 5).

UDFORDRINGER VED AT FINDE EVALUERINGSKRITERIER

Som vi så ovenstående, er evaluering en bedømmelse, der forudsætter værdier. Det er ofte svært at artikulere præcist, hvilke evalueringskriterier et innovativt tiltag skal leve op til. Sprogligt udtrykker vi det i dag ved, at en innovativ indsats skal give "mening". Det er fint, men evalueringsmæssigt meget uspecifikt. Hvilken mening? For hvem? Vi taler også om, at innovation skal skabe værdi. Hvilken værdi? Værdien kan for eksempel komme til udtryk på en eller flere af fire offentlige bundlinjer: Administrativt, som henviser til øget effektivitet internt eller eksternt. Service, som henviser til bedre serviceoplevelser for borgere eller virksomheder. Policy, som er den adfældsændring, som er skabt i samfundet som en konsekvens af innovationen, og slutteligt kan innovationen skabe værdi ift. øget demokratisk deltagelse, retssikkerhed eller gennemsigtighed (Carstensen 2012-13).

Der kan være konflikt mellem værdier. Hvordan skal de så vægtes? Under overskriften værdier kan nye interessenter også bringe nye emner op. Det kan være ligestilling, bæredygtighed, sikkerhed eller forskellige strategiske visioner. Men hver gang værdier nævnes, kræves en yderligere specifikation, før man har et evalueringskriterium.

Brugen af brede ord som mening og værdi udtrykker, at vi i vores samfund for tiden har svært ved at præcisere og operationalisere, hvad vi mener er det positive og konkrete udfald af vore indsatser. Det kan ikke undre, for mange af vore indsatser har allerede i selve deres definition indbygget i sig store, brede og fortolkningsbare begreber som ligestilling, kvalitet, sikkerhed, bæredygtighed og – ja – innovation.

Evaluators forsøger som regel at finde frem til evalueringskriterier, der er officielt gældende, autoritative, vedtagne, meldt ud, lovede, eller som der kan skabes en eller anden form for enighed om blandt centrale interessenter. Valget af evalueringskriterier kan være mere eller mindre velbegrunderet ud fra en demokratisk styreform eller ud fra en etisk teori. Men der findes intet evalueringsværktøj, der kan fremtrylle enighed om værdier og kriterier. I praksis kan man begynde en afklaring af evalueringskriterier for et innovativt tiltag med følgende spørgsmål:

- har man i aftaler eller projektbeskrivelser udstedt forpligtende løfter om det innovative tiltags karakter og formål?
- er formålet med det innovative tiltag klart artikulert? Er der tale om en specifik form for værdi udtrykt i konkrete målsætninger?
- hvis der er et ønske om at tilgodese flere værdier ved samme innovative tiltag, hvordan vægter man så hensynet imellem dem?

Hvis der ikke kan etableres en enighed om bestemte evalueringskriterier, kan der så aftales en proces, hvor man finder frem til sådanne? Kan man aftale, hvordan man vurderer gode og dårlige bieffekter, når de opstår og opdages undervejs?

Hvis nej, er det så aftalt, hvordan man kommer en indkredsning heraf nærmere, eventuelt hvem der har ret til at foretage en sådan indkredsning? Hvis dynamik og uforudsigelighed er forventelig i det foreliggende innovative tiltag, er der så aftalt et grundlag eller nogle principper ud fra hvilke, man kan vurdere håndteringen af dynamik og uforudsigelighed? Til ethvert innovativt tiltag er der et eller andet råderum inden for et politisk, ledelsesmæssigt og organisatorisk system. Er det råderum udnyttet på en god måde? Det kan f.eks. være, at tiltaget skal leve op til bestemte principper om god faglig skik, god ledelsespraksis, lytten til brugere eller andet. Er det innovative tiltag godt koncipert i forhold til sin kontekst?

Sjældent bruger offentlig sektor skatte kroner, uden at der er aftalt hverken mål, processer eller principper. Dermed har man som regel et evalueringsgrundlag. COI's innovationsbarometerdata viser, at der i 84 % af tilfældene på forhånd er opstillet et formål med innovationen (Center for Offentlig Innovation 2015a: 41). Selv om der kan være yderligere evalueringskriterier i tilgift, så er i hvert fald nogle evalueringskriterier inden for rækkevidde.

Hvis der ikke foreligger noget sådant grundlag, når en evaluator går i gang, hvilket handlerum har en evaluator så til selv at afklare evalueringsgrundlaget? Hvis der ikke i evaluering af det innovative tiltag etableres et nogenlunde klart evalueringsgrundlag, så løber man en stor risiko for, at der spildes ressourcer på unyttig evaluering. Selv om det er vanskeligt at finde frem til gode evalueringskriterier, går problemet ikke væk, for når et innovativt tiltag finder sted for offentlige midler, er det ikke tilstrækkeligt at nogen blot vurderer det ud fra helt egne subjektive kriterier. Der er brug for en offentlig begrundelse, dialog og argumentation for at etablere gældende evalueringskriterier.

Heldigvis er der mange evalueringsmæssige svar på udfordringen om manglende eller bredt formulerede evalueringskriterier. Få evaluering ind tidligt. Evaluering bør nævnes som en integreret del af en projektbeskrivelse. Vær sikker på, at det er det innovative tiltags hovedidé, som er grundlag for evalueringsarbejdet. Det bør lede til gode evalueringskriterier. Pejle efter, hvilken evalueringsmodel med tilhørende evaluerings spørgsmål, det vil være relevant at stille (se del 6). Alt dette er god medicin mod fortsatte vage og uspecifikke evalueringskriterier.

NÅR EFFEKTVURDERING BLIVER UMULIG: EVALUERINGERNES BERMUDATREKANT

Med ordet effekt henviser vi til et årsags-virkningsforhold mellem indsats og resultat. Effekten er en relation. Det er ofte svært at fastlægge effekten af innovative indsatser.

Se for dig et innovativt tiltag, hvor man etablerer nye patientforløb. Man vil gerne overbevise om, at indsatsen har haft en effekt. Lægerne kan godt lide randomiserede eksperimenter.

I det konkrete forløb har man ikke en kontrolgruppe, for alle patienter i regionen er med i det nye forløb. Der findes ikke personaler, der kun arbejder med det gamle forløb, så det er svært at skille det gamle fra det nye. Selv om det nye forløb for patienter med en bestemt diagnose er lovende, er virkeligheden plaget af, at patienter ofte har mere end én diagnose samtidig, så der må træffes en række praktiske afgørelser, der ikke er beskrevet i standardiserede manualer.

Det nye forløb virker meget uens, fordi samarbejdspartnerne (kommuner og praktiserende læger) arbejder meget forskelligt med det nye koncept. Det nye forløb er ikke standardiseret, for man lærer og justerer undervejs. Det er jo en indsats med stor komplicerethed, både teknisk og socialt. Man tager et sygdomsområde ad gangen og prøver at lære fra område til område.

Er det nogenlunde standardiseret, hvad forløbskoordinatorerne foretager sig? Nej. De er en uens flok behjertede ildsjæle, det skal man være som forløbskoordinator, men netop derfor har de meget forskellig arbejdsstil. De har forskellige erfaringer og kender forskellige mennesker i sundhedssystemets store netværk. Derfor griber de også opgaverne an på forskellig måde.

Det samlede overblik har ingen. Den store udfordring er at få kommunerne med på vognen. Er der nogen repræsentanter for kommunerne i styregruppe og evalueringsgruppe? Nej, nu er det jo regionen og dermed sygehusene, der har søgt og fået de ekstraordinære midler til at udvikle det nye patientforløb.

Mens man drøfter alle disse emner, cirkler man om samme spørgsmål. Det handler om det optimale design. Man vil gerne have et randomiseret eksperiment, der kan overbevise. Men nu er man i gang, og dette innovative tiltag ligner et randomiseret eksperiment mindre og mindre, for hver dag, der går. Og så vil man have dokumenteret effekten – i bestemt form, ental!

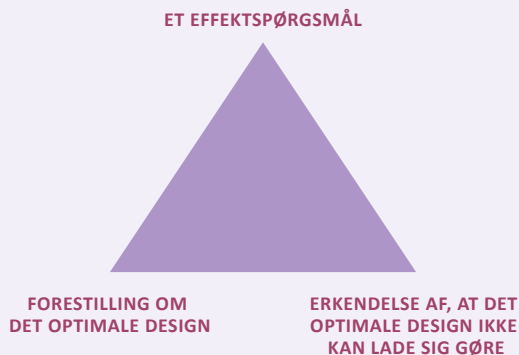
Opgaven er, som den er formuleret her, umulig at løse. Man er havnet i evalueringernes Bermudatrekant.

I evalueringens Bermudatrekant risikerer evalueringer at blive til ingenting, ligesom fartøjer forsvinder i den rigtige Bermudatrekant.

Mange innovative tiltag er baseret på store løfter og udsigter til fantastiske langsigtede effekter. I evalueringens øjeblik må man konstatere, at metodeproblemer forhindrer, at man kan påvise disse. De fleste ønskværdige effekter ligger ude i fremtiden. Evalueringens resultat er, at man fornemmer, at en god proces er sat i gang. Evalueringsmæssigt er en sådan iagttagelse tyndbenet.

Evalueringsmæssigt bør der være en logisk sammenhæng mellem, hvad man lover, og hvad man har tænkt sig at påvise. Bermudatrekanten viser, at der er tre mulige veje ud af problemet. Enten må man revidere sine forestillinger om det optimale design, således at man vælger det bedste design blandt dem, som er realistiske. Det kan være en virk-

FIGUR 6
EVALUERINGS BERMUDATREKANT



ningsevaluering. Eller også indretter man fra starten sin indsats sådan, at den bliver mere evaluerbar ud fra forventningerne. Det kræver f.eks. kontrolgruppe og standardisering. Eller også opgiver man at stille et effektspørgsmål. I stedet kan man f.eks. spørge om forskellige interessenters vurdering af tiltaget.

Metodestyling kan være overstyring. Man kan ikke i en vilkårlig situation blot forlange, at effektspørgsmålet besvares optimalt, hvis man ikke har indrettet det innovative tiltag derefter, og hvis man ikke er parat til at revidere sine forestillinger om, hvordan effekten bedst påvises.

Erkend, at effekt er et resultat af et kompromis. Hvis en klassisk variansbaseret effekt-evaluering ikke kan lade sig gøre, bør man overveje en god virkningsevaluering (beskrevet i del 5). Hvis man ønsker andre spørgsmål end effektspørgsmål besvaret, kan man overveje en brugerorienteret evaluering, en partcipatorisk evaluering med forskellige interessenter, en ekspertbaseret evaluering, eller en systemisk evaluering (del 5).

EVALUERINGSRESISTENS SOM UDFORDRING

Man kan risikere at udnævne det nye til at være så unikt, at det bliver resistent over for evaluering. Flyvbjerg (2014) har peget på, at meget store projekter tit er forbundne med forestillinger om det sublime i flere henseender: politisk, økonomisk, æstetisk, teknologisk. Risikoen herved er, at man har så store forventninger og foretager så store investeringer, at man kommer til at stille sig i vejen for en realistisk afprøvning af ideerne. Når man er på vej til det sublime, er der heller ingen grund til at danne sig et nøgternt billede af, hvad man kan lære af tidligere erfaringer. Mange af de samme forestillinger kan godt genkendes vedrørende innovative tiltag.

Ved et stort offentligt byggeri var ambitionen at skabe helt nye fysiske rammer for fremtidens videns- og forskningsinstitution. Der blev oprettet en brugergruppe, der skulle give input til bygningschef og arkitekt. Det var en interessant mulighed for at præge det innovative tiltag. En deltager bad om lov til at gå en kort tur med arkitekten i det eksisterende byggeri for at lære ham om de svagheder, som kunne forbedres fremover. Der blev givet det svar, at fortidens byggeri var irrelevant, fordi nybyggeriet var for det næste årtusind.

Derfor endte det nye byggeri – trods dets mange kvaliteter – med at gentage nogle af de fejl, der findes i det gamle.

Man bør i stedet modvirke evalueringsresistens ved at få evaluering meget tidligt ombord i et innovativt tiltag.

HVORDAN EVALUERING KAN MEDVIRKE TIL SPREDNING

Der er penge at spare ved at bruge gode ideer på tværs af tid og rum. Spredning af innovation kan ske ved hjælp af grænseobjekter (COI 2015b). Evaluering kan fungere som et sådant grænseobjekt, men det afhænger af modvindsfaktorer og medvindsfaktorer. En modvindsfaktor er, hvis man taler om et innovativt tiltag på en sådan måde, at man helt overdriver det enestående og unikke ved netop dette tiltag på netop den nuværende lokalisering. Scenesættelsen af det innovative tiltag har med andre ord betydning for evaluators evne til at medvirke til spredning.

En vigtig medvindsfaktor er omvendt, at man er god til at forklare, hvordan et innovativt tiltag løser et problem, som mange deler. Tiltaget er ikke mere unikt, end at det falder i en klasse af tiltag, der sigter mod at løse samme problem. Også derfor er det klogt at være meget klar på, hvad hovedideen er i det innovative tiltag. En evaluering, der dokumenterer, at et givet tiltag har virket godt til at løse et veldefineret problem, som andre også har, kan være et utroligt hjælpsomt middel i forhold til spredningen. Det kan typisk være en varians-baseret effektevaluering eller en virkningsevaluering. Evaluering faciliterer spredning ved at vise, at et tiltag baseret på en klar idé reelt fungerer som løsning på et typisk problem, der findes flere steder.

LÆR AF ERFARINGER FRA ANDRE OMRÅDER

Hvad der gælder for innovative tiltag, gælder også for evaluering af innovative tiltag: Man kan lære af andre. Man har nemlig på andre områder gjort erfaringer med at styrke tilsvarende læreprocesser, som dem vi søger at fremme, uanset om man lige brugte ordet ”innovative tiltag.”

Det gælder f.eks.:

- ved opsamling af erfaringer af store IT-projekter har man f.eks. fundet ud af, at kravspecifikationer ikke bør formuleres langt væk fra de konkrete brugere. Det er bedre at tage udgangspunkt i brugernes behov. Når udviklingen går stærkt, er det også en god idé at præsentere prototyper og pilotprogrammer for brugerne med henblik på løbende udvikling og korrektion. Der skal ikke være for langt i tid og rum mellem behov, kravspecifikation, leverance og afprøvning i praksis.
- ved opsamling af erfaringer med kvalitetsarbejde i sundhedsvæsenet har man fundet ud af, at de metoder, man normalt bruger til effektvurdering af medicinske tiltag, har begrænset duelighed, når det gælder store organisatoriske kvalitetstiltag. Til gengæld er det vigtigt at begrebsliggøre og modellere indsatsen meget klart og sikre en flerhed af løbende organisatoriske tilbagemeldinger.
- fra kvalitetsarbejde i organisationer har vi lært om værdien af at gå på tværs af den gængse funktionelt opdelt organisation for til gengæld at sætte sig sammen og tale konkret om problemer og løsningsforslag.

- fra studier af teknologi og teknologivurdering har vi lært, at modeller, der inddrager forskellige interessenter, fører til forskellige typer af vurdering. På den ene side har man ekspertbaserede og eliteprægede vurderinger. På den anden side har man former for teknologivurdering, der samtidig er demokratiske eksperimenter, f.eks. fremtidsværksteder og folkehøringer. Her har Danmark en gang været førende.
- fra designområdet har vi lært betydningen af at lave mock-ups, modeller og prøveopstillinger i mindre skala for at sætte en tidlig læreproces i gang.
- fra erfaringer med projektledelse har vi lært, at det er for snævert at evaluere ud fra, om et produkt, lever op til på forhånd opstillede kriterier (budget, deadline, produkt-specifikationer). Man har lært også at tage i betragtning, hvordan brugere tilegner sig et nyt produkt og om de kan anvende det i hverdagen, ligesom man kan problematisere, om der er i det hele taget er tale om det rigtige produkt. Alle faser i projektet kan være genstand for evaluering, og problemet kan ligge i kravspecifikationerne lige såvel som i ideen bag produktet.
- fra kvalitative undersøgelser af brugeres oplevelser af service ved vi, at de ikke altid kan artikulere, hvad det senere viser sig, at de bliver glade for. Vi ved også, at brugere kan have meget forskellige behov og ønsker i forhold til samme tiltag.
- fra governance-studier og studier af boligsociale indsatser ved vi, at kvaliteten af netværk og samarbejde ofte er helt afgørende for udfaldet af en innovativ indsats.

Den bedste fortolkning af innovative tiltag får man ikke ved at gøre innovation som generelt begreb helt forskelligt fra alt andet, men ved at inddrage alle relevante erfaringer i det innovative. Man skal ikke i offentlig sektor opfinde den dybe tallerken hver gang. Læring består i at knytte ny viden sammen med gammel.

5

HVILKE SPØRGSMÅL KAN BESVARES MED HVILKE EVALUERINGSMODELLER?

Ikke mindst af overvejelserne om Bermudatrekanten kan vi lære, at man må være omhyggelig i evaluering med at stille et spørgsmål, som faktisk kan besvares. Forskellige evalueringsmodeller er egnede til at besvare forskellige spørgsmål.

Vi har tidligere set, at vi i dag er i en situation, hvor det er svært at artikulere tydelige evalueringskriterier. Derfor må du tidligt afklare evalueringsgrundlaget. Hvis du kan skabe

FIGUR 7
EVALUERINGSMODELLER

	Målopfyldelses evaluering	Varians-baseret effekt-evaluering	Virknings-evaluering	Bruger-orienteret evaluering	Interessent-baseret evaluering	Ekspert-evaluering	Syste-misk evaluering
Er officielle, klare mål for det innovative tiltag opfyldt?	×						
Hvilken effekt har det innovative tiltag haft under kontrollerede betingelser?		×					
Hvilken effekt har det innovative tiltag haft, hvis man systematisk sammenligner det med en anden situation uden tiltaget?		×					
Kan man sandsynliggøre en effekt af den innovative indsats via processtudier baseret på programteori?			×				
Er det innovative tiltag relevant i forhold til brugernes ønsker, forventninger og behov?				×			
Hvordan opleves det innovative tiltag af forskellige interessenter?					×		

enighed om valget af en evalueringsmodel i samarbejde med din opdragsgiver eller dine vigtigste interessenter, kan du indkredse evalueringskriterierne.

Hvis du ønsker at besvare flere spørgsmål og kombinere flere modeller, kan du så rumme dette inden for de ressourcer i tid og penge, som er sat af til evalueringsarbejdet? Hvis resultaterne er forskellige baseret på flere evalueringsmodeller, på hvilken måde vil du så sammenfatte resultaterne? Hvis evalueringsarbejdet ikke skal overbelastes, må man prioritere. Hvis du ønsker at stille et spørgsmål, der ligger meget langt fra et af de foreslåede spørgsmål, er det en god idé at tænke over, om dit spørgsmål overhovedet kan besvares.

Figur 7 viser eksempler på spørgsmål, man kan stille til det innovative tiltag. Oversigten viser, hvilke(n) evalueringsmodel(ler), der vil være et fornuftigt valg givet det enkelte spørgsmål. Hver model står for bestemte evalueringskriterier eller måder at finde evalueringskriterier på. Hver model stiller krav til evalueringsarbejdet. Hver model har fordele og ulemper.

	Mål- fyldelses evaluering	Varians- baseret effekt- evaluering	Virknings- evaluering	Bruger- orienteret evalue- ring	Interessent- baseret evaluering	Ekspert- evalue- ring	Systemisk evalue- ring
Er der enighed om det innovative tiltags mål og mening, herunder betydningen af de vigtigste begreber?				×	×		
Kan man fremme udvikling af tiltaget ved at inddrage nye synspunkter?				×	×		
Kan man øge anvendelsen af evaluering ved at inddrage flere interessenter?					×		
Hvordan vurderer en ekspert det innovative tiltag i sin helhed?						×	
Har det innovative tiltag bivirkninger?			×	×	×	×	×
Hvordan organiseres foreløbig feedback i lyset af det innovative tiltags kompleksitet og dynamik?							×

MÅLOPFYLDELSESEVALUERING

I målopfyldelsesevaluering tager man udgangspunkt i formelt opstillede mål. Det kan være mål i lovgivningen, i politiske beslutninger eller i løfter afgivet i projektbeskrivelser, i ansøgninger eller i aftaler med en opdragsgiver. Målopfyldelsesevaluering er nemmest i situationer, hvor målene er klare og operationelle.

I tilfælde af uklare mål, flere mål eller mål med et vist rum for fortolkning er det stadig en god idé at gå disse på klingen. Hvis mål ændrer sig eller fortolkes på ny undervejs, kan der være mere eller mindre gode argumenter for at gøre det. Og en forskydning af mål kan være mere eller mindre i overensstemmelse med ønsker og forventninger hos den, der finansierer en innovation. Det er derfor en god idé at komme en målaflarung så tæt som muligt og være i løbende dialog i tilfælde af udvikling af mål for derved at give evaluatoren det bedste mulige grundlag for at vurdere graden af målopfyldelse.

Det kan være uhensigtsmæssigt både for et innovativt tiltag og for evaluering heraf, hvis man holder store og væsentlige målsætninger skjult. Et typisk eksempel er et innovativt tiltag, der skal føre til en rationaliseringsgevinst. Hvis det forbliver uklart, om denne gevinst skal komme brugerne, personale eller kommunekassen til gode, kan det skabe kontroverser undervejs, og man giver ikke evalueringen en chance for at forholde sig til tiltagets egentlige målsætninger. Hvis evaluering skal påpege reelle resultater, skal de vigtigste målsætninger være kendt.

En af svaghederne ved målopfyldelsesevaluering er, at den ikke svarer på, hvorfor man eventuelt opfyldte eller ikke opfyldte målet. I nogle tilfælde behøver man heller ikke vide det. Hvis målopfyldelsen er meget ringe, kan det være nok at konstatere, at det pågældende innovative tiltag ikke lykkedes.

En anden svaghed er, at bieffekter af det innovative tiltag ikke belyses i målopfyldelsesevaluering.

VARIANSBASERET EFFEKTEVALUERING

Formålet er her at bestemme effekten af en innovativ indsats via systematisk sammenligning med en situation, hvor indsatsen ikke har fundet sted. Mange mener, den fornemste løsning på denne opgave er at etablere to ens situationer, hvor man så kun foretager det innovative tiltag i den ene af disse. Et kontrolleret lodtrækningsforsøg realiserer denne tanke, men er i mange tilfælde svært at gennemføre. Man har ikke altid en velegnet kontrolgruppe. Man har ikke altid mulighed for at sikre, at kontrolgruppen ikke har adgang til det innovative tiltag eller noget, der ligner. Man har ikke altid udviklet det innovative præcist nok til, at det er stabilt gennem en forsøgsperiode. Og så ved man ikke bagefter, hvilken udgave der eventuelt virkede.

Men man skal ikke opgive for tidligt. Min kollega Kasper Møller Hansen med flere (Bhatti et al. 2015) har gennemført kontrollerede forsøg i situationer, hvor andre ikke troede det muligt. For eksempel har han sammen med Indenrigsministeriet sendt grundloven og en opfordring til at stemme ud til 18-årige på differentierede tidspunkter (før og efter valget), så man kan se effekten på valgdeltagelsen i et konkret valg. Mange havde ikke set den mulighed for sig.

Uden et lodtrækningsforsøg kan man forsøge at løse sammenligningsproblemet via statistisk sammenligning med andre steder uden samme tiltag eller man kan sammenligne situationen samme sted før og efter det innovative tiltag. Men så kan effekten af tiltaget ikke observeres rent.

En før-og-eftermåling kan være vildledende, hvis store eksterne kræfter påvirker udfaldet i tilgift til det innovative tiltag. Et eksempel er, at mange innovative tiltag til reduktion af antal sygedage tilsyneladende har været succesrige de senere år, i hvert fald hvis man kun ser på før-og-efter-målingen. Men da antallet af sygedage generelt har været stærkt faldende i Danmark siden krisen indtrådte i 2008, og hvor mange nu frygter fyringer og omstillinger, kan man komme til at tilskrive de særlige innovative tiltag til reduktion af sygedage større succes, end de faktisk fortjener.

VIRKNINGSEVALUERING

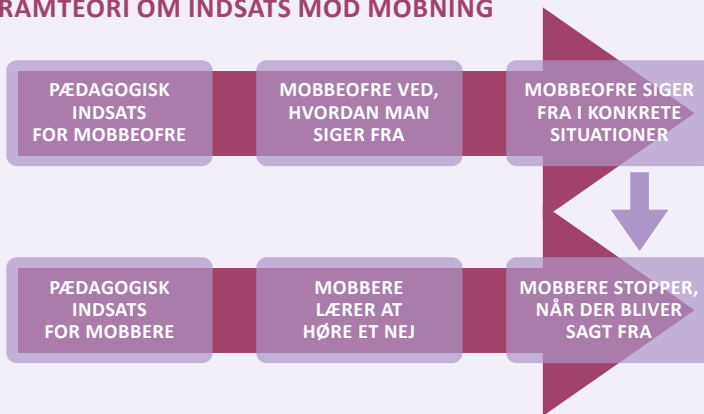
I en virkningsevaluering opstilles en programteori, dvs. en oversigt over, hvorfor og hvordan et innovativt tiltag virker. Det kan gøres sprogligt eller ved en grafisk skitse. Evalueringen går ud på at holde hver fase i programteorien op mod data for derved at sandsynliggøre, om effekten er indtrådt eller ej. Virkningsevaluering starter med en konkret beskrivelse af indsatsen og forventningerne til, hvordan den virker.

Et innovativt tiltag, der handler om "at styrke en yderligere målretning af indsatsen" eller "et kvarterløft", er ikke tæt nok beskrevet. Prøv med konkrete personer og verber. Hvem skal gøre hvad? Man kan udtrykke en innovativ indsats som en kæde af aktør-verbum-ting. Ting-verbum-aktør. Det er verbet, der er forbindelsesleddet. Når man skal forstå en innovativ indsats i forbindelse med f.eks. patientforløb kan man forestille sig, at hospitalet sender en epikrise til den praktiserende læge. Hvad skal lægen gøre? Læs epikrisen. Handle derefter. Det er ikke blot "styrket kommunikation". Det er et konkret tiltag, når man kan se aktører og verber.

Fordelene ved virkningsevaluering er mange. Man kan sandsynliggøre, om der har været en effekt, også uden en kontrolgruppe. Man kan inddrage fagfolk i artikuleringen af, hvorfor denne indsats antages at virke på denne måde. Man kan specificere nogle gode succeskriterier, når man er teoretisk klar over, præcis hvad man vil opnå og hvordan. I en innovativ indsats for at inddrage rygpatienter i et behandlingsforløb skal man måle på muskelstyrke og daglig funktionsevne. Man skal ikke måle på smerter, for man skal nemlig træne, selv om man har smerter, siger filosofien i denne indsats. Succeskriterierne følger heraf.

Et andet eksempel: En gruppe pædagoger ønsker at udvikle en ny indsats mod mobning. De er optaget af, at man skal lære at respektere hinandens etiske urørlighedszone. Det er en god idé helt alment, men ikke operationelt nok til at artikulere en innovativ indsats og heller ikke specifikt nok til at blive grundlag for evaluering. Via samtale finder man ud af, at det må handle om at lære mobbeofrene at sige tydeligt nej og at lære mobberne at høre og respektere dette nej. Det giver følgende programteori:

FIGUR 8
PROGRAMTEORI OM INDSATS MOD MOBNING



Der resterer noget arbejde for at finde ud af, præcis hvilke data der skaffes for at belyse hver af boksene. Men det kan gøres f.eks. via interviews og observation. Med data om hver boks kan man sandsynliggøre, hvorvidt indsatsen har virket eller ej. Hvis en innovation er original, må man skræddersy en programteori dertil. Men mange innovative tiltag bygger videre på andres gode ideer. Så kan man tage udgangspunkt i en arketype (Funnell og Rogers 2011: kap. 12). En arketype beskriver kendte træk ved en hel gruppe af indsatser baseret på samme idé. En programteoretisk arketype kan f.eks. beskrive god case-management ved hjælp af handleplaner, brugen af nudging, eller hvordan man via kampagner formidler viden, der ændrer holdninger og adfærd.

Ligeledes kan erfaringer fra innovative tiltag efter evaluering opsummeres i form af arketyper, som andre kan bruge. For eksempel kan konsulentvirksomheder og videncentre bruge programteoretiske arketyper til at opsummere erfaringer fra mange projekter rundt i landet inden for samme indsatsområde, f.eks. en satspulje. Hver arketype repræsenterer en gruppe af innovative tiltag baseret på samme grundide.

I komplicerede organisatoriske sammenhæng kan et innovativt tiltag omfatte mange aktører, mange organisatoriske lag og flere virkemidler. Det kan medføre en mere kompliceret programteori. Alligevel er det en god idé at artikulere den. Også hvis programteorien via evaluering viser sig ikke at holde, har man fremmet en organisatorisk læreproces, netop fordi udgangspunktet for læring var meget eksplicit.

Virkningsevaluatorer ønsker at finde ud af, om en evaluering svækker eller styrker tiltroen til programteorien i det innovative tiltag. Til det brug er det gavnligt at sondre mellem implementeringsfejl og teorifejl.

IMPLEMENTERINGSFEJL ELLER TEORIFEJL?

Implementeringsfejl findes, hvis indsatsen ikke er udført som planlagt. Der kan mangle personale, information, vilje, ressourcer og kunnen. Teorifejl findes, hvis indsatsen er godt implementeret, men resultaterne alligevel udebliver. Kun hvis der ikke er fundet implementeringsfejl, kan man fastslå en eventuel teorifejl.

FIGUR 9
IMPLEMENTERINGSFEJL ELLER TEORIFEJL?

	Resultat indtraf	Resultat indtraf ikke
Implementering var OK	Tiltro til programteori styrket	Teorifejl
Implementering var ikke OK	Andre forklaringer end programteorien gælder	Implementeringsfejl Muligvis teorifejl

Den skematiske måde at konkludere på forudsætter, at både indsatsen og resultatet er veldefinerede og beskrevne som enten-eller. I et komplekst samarbejde omkring et innovativt tiltag er det somme tider svært at afgøre, om samarbejdspartnere er "derude", eller om de kan betragtes som "led i implementeringen." Tilsvarende kan resultatet være "mere eller mindre indtruffet." Alligevel er det fornuftigt at bruge skemaet til at afgøre, om der er flest gode argumenter for, at der findes en implementeringsfejl eller en teorifejl. For konsekvenserne er forskellige. Implementeringsfejl kan rettes op. Hvis der er teorifejl, bør man skifte grundideen ud.

Man kan kvalificere evalueringsarbejdet ved at se efter, om tegnene i data er nødvendige og/eller specifikke i forhold til programteorien. Et nødvendigt tegn er en forudsætning for, at tiltaget virker. I en række af tegn er hvert enkelt tegn nødvendigt, hvis der ikke er nogen veje uden om. Det er nødvendigt, at mobberne lærer at høre et nej, for der er ingen veje uden om i denne udgave af programteorien.

Et specifikt tegn viser hen til en bestemt form for kilde, ligesom DNA eller fingeraftryk gør i en retssag. I programteorien ovenfor vil det være et specifikt tegn, hvis mobbeofrene siger fra på præcis den måde, som de har lært i den pædagogiske indsats. Så øges sandsynligheden for, at det er indsatsens skyld, at de har lært at sige fra. Man kan med andre ord foretage logiske skridt for at teste programteorien så indgående som muligt.

I et arbejdsmiljøforsøg arbejdede man ud fra den idé, at en individuel og en organisatorisk indsats måtte kombineres for at opnå en effektiv strategi til bekæmpelse af stress. Man havde gode teorier om, hvordan hver type kunne bidrage til mindsket stress. I udgangspunktet en god idé, idet tidligere indsatser for ofte kun har fokuseret på det ene af disse niveauer.

I evalueringen fandt man ud af, at der manglede tilstrækkelig involvering af de aktører (ledelse, tillidsfolk og sikkerhedsrepræsentanter), der skulle implementere indsatsen. Og man fandt ud af, at indsatserne på de to niveauer ikke var flettet godt nok sammen. For når en enkelt medarbejder bad om arbejdsmodifikation, blev vedkommendes kolleger blot yderligere belastet. Det skabte et modsætningsforhold mellem de stressede og de andre medarbejdere. Det var det modsatte af, hvad man ønskede at opnå.

Programteoretisk tænkning hjalp til at identificere netop de punkter, der skulle arbejdes mere med i næste fase.

Når indsatsen bliver kompleks kan man gå to veje. Den ene er at gøre programteorien mere kompliceret, men stadig med udgangspunkt i, at man vil påvise, om den holder eller ej. Man kan tillige lade programteorien udvikle sig, så den følger med det innovative tiltags udvikling:

2014: Programteori i tidlig fase trykprøves

2015: Programteori i projektbeskrivelse

2016: Programteori efter analyse af implementering

2017: Programteori efter midtvejsevaluering

2018: Programteori efter evaluering til brug for andre og for fremtiden

Funnell og Rogers (2011) er eksponenter for denne tankegang. Den anden tanke er ikke at holde bog med programteorien, men blot bruge den som læringsramme om et systemisk evalueringsarbejde (se side 41).

BRUGERORIENTERET EVALUERING

I brugerorienteret evaluering tager evalueringskriterierne udgangspunkt i brugernes ønsker og behov. Det er ikke enhver inddragelse af brugere, der er brugerorienteret i den forstand. Krogstrup (2001) skelner således mellem top-down og bottom-up orienteret brugerevaluering alt efter, hvem der bestemmer, hvilke emner og kriterier evalueringen handler om. Man kan godt bede brugere om deres vurdering af emner og kriterier, som er opstillet af en organisations ledelse, altså top-down. Men genuin brugerorienteret evaluering foregår med udgangspunkt i brugernes egne ønsker og behov, altså bottom-up. Så kan temaer og kriterier ikke være sat på forhånd af andre.

Inddragelse af brugere kan medvirke til udvikling af bedre og mere effektive innovative produkter og services. Forbedring er her set fra udbyderens side. Service eller produktet kan også tilpasses brugerens behov. Her er synspunktet forbedring for brugeren. Krogstrup (2001) har beskrevet, hvordan brugere kan inddrages i kvalitetsudvikling via fokusgrupper.

Brugerinddragelse rejser nogle principdiskussioner. Hvordan afgrænses brugere? Er det kun faktiske brugere, eller bør potentielle brugere tælles med? Måske når et produkt eller en service kun frem til en snæver gruppe, mens andre står på venteliste eller slet ikke har hørt noget. Hvis faktiske og potentielle brugere har meget forskellig vurdering, er det af stor betydning for det samlede evalueringsresultat, hvordan bruger kategorien afgrænses. Pårørende er berørte og har sommetider stærke meninger, også selv om de ikke er primære brugere. At afgrænse en brugergruppe er vanskeligt og værdiløst.

Brugere kan være inkonsistente. Det betyder, at de kan ønske sig noget meget intenst uden dog at blive tilfredse, når ønsket opfyldes. Der kan være uoverensstemmelser mellem ønsker og behov. Den innovative bilproducent Henry Ford sagde, at hvis han virkelig skulle give kunderne, hvad de ønskede sig, så havde han givet dem hurtigere heste.

Forholdet mellem ønsker og behov betyder, at man kan blive glad for noget, man ikke vidste, man ønskede eller havde behov for, og man kan være utilfreds med noget, man selv har ønsket. Før en togrejse ønsker de fleste passagerer et roligt sæde, der er så privat som muligt. Alligevel rummer deres mest mindeværdige togrejser et møde med spændende ukendte fremmede.

Ikke mindst i vanskelige læreprocesser kan man som bruger opleve en slags overskridelse, hvor der kommer et nyt perspektiv gennem erfaringen af bemestring, hvilket gør

tidligere tiders besværligheder mere tålelige, end da man stod midt i dem. De fleste danskere var modstandere af Storebæltsbroen i udgangspunktet. Da den stod klar til brug, var de fleste danskere begejstrede.

Brugere kan være uenige. Deres ønsker kan være i konflikt med hinanden, hvilket tydeliggøres, hvis de skal deles om organisatoriske løsninger. På nogle biblioteker går man innovative veje for at skabe et aktivt center for informationssøgning og kultur. Nogle brugere søger imidlertid biblioteket for at finde ro som i en kirke eller et museum. Plejehjemspatienter ønsker ofte individuelle løsninger, men henvises til at bo og spise sammen med mennesker, de ikke selv har valgt.

Brugerevaluering bør rumme modsætninger, hvis uenigheder og inkonsistenser er en del af virkeligheden.

Man selvstændiggør i brugerorienteret evaluering et enkelt aspekt af mennesket kaldet "brugeren". Det har en høj pris. En studerende har for eksempel et medansvar for sin læreproces, hvor brugeren som oftest anlægger et ensidigt evalueringsperspektiv. Tosi-dige, interaktive processer beskrives fra brugerens side som om, de blot er varer, der leveres. Man reducerer en rigere rolle, f.eks. som studerende eller som borger, til en fattigere rolle som bruger. I innovative processer, f.eks. co-production eller samskabelse, har man netop brug for en gensidig omdefinering af rollerne mellem den offentlige sektor og brugeren. Ønsker man en sådan omkalfatring, må evalueringsperspektivet ikke medskabe en bruger, som mere ligner en forbruger. I stedet må ønsket om nye relationer afspejles i evaluering, der på innovativ måde bidrager til en omdefinering af brugerbegrebet.

Og man skal naturligvis ikke spørge brugere om noget, hvis man ikke har en reel interesse i at høre, hvad de har at sige.

PARTICIPATORISK EVALUERING

Participatorisk evaluering omfatter flere forskellige tilgange og modeller. De har det tilfælles,

1. at forskellige interessenter deltager i evalueringen,
2. at deres perspektiver, situationer, erfaringer eller synspunkter indgår med vægt i evalueringen og præger dens indhold eller form,
3. at der trods forskelligheden tilsigtes en eller anden form for samarbejde om evalueringen, dens proces, temaer og/eller anvendelse.

Participatorisk evaluering og brugerorienteret evaluering adskiller sig grundlæggende fra målopfølgelsesevaluering og effektevaluering derved, at evalueringens kriterier, temaer og tilrettelæggelse udledes som et led i selve evalueringsprocessen. Nu er mål og effektspørgsmål ikke fastlagt, før evalueringen går i gang. Dette er skellet mellem "preordinate" og "responsive" evaluering (Stake 2004).

Inden for den participatoriske ramme findes forskellige udformninger af evalueringen. Det skyldes, at begrundelserne for at arbejde participatorisk er forskellige.

I pragmatisk participatorisk evaluering er evalueringens anvendelse det primære sigte. Centrale interessenter er dem, der står med nøglen til at få evalueringens resultater brugt

FIGUR 10
PRAGMATISK OG TRANSFORMATIV EVALUERING

	Pragmatisk evaluering	Transformativ evaluering
Mængde af interessenter	Lille eller mellem	Stor
Interessenternes forskellighed	Lille eller mellem	Stor
Magtdifference mellem interessenterne	Lille eller mellem	Stor
Dybde i involvering	Mellem	Stor
Ressourcetræk til deltagelse	Mellem	Stor
Uforudsigelighed i proces	Lille eller mellem	Stor
Konfliktpotentiale	Lille eller mellem	Stor

i forhold til et innovativt tiltag. En enkelt person (f.eks. en projektleder) eller en lille gruppe af interessenter (f.eks. en styregruppe) bliver i den henseende til primære tilsigtede brugere af evalueringen i henhold til en primær tilsigtet brug (f.eks. læring og udvikling). Som regel knytter evalueringen sig til en officielt gældende organisations- og ledelsesstruktur.

I transformativ evaluering ses evalueringen i sig selv som en innovativ social forandringsproces. Med vilje inviteres interessenter, der ikke plejer at blive hørt eller ikke plejer at tale sammen. Med vilje sættes uvante rammer for samtalen. Idealet er en demokratisk dialog, hvor en grænseoverskridende og transformativ social forandring sættes i scene. Hvor man i pragmatisk partcipatorisk evaluering er kvalificeret til deltagelse, hvis man har indflydelse på evalueringens anvendelse, er man i transformativ evaluering kvalificeret, hvis ens liv er berørt af det innovative tiltag.

I pragmatisk evaluering er mængden af interessenter begrænset. Man kan fokusere på en enkelt primær tilsigtet bruger eller et mindre antal, der i fællesskab har ansvar for ledelse, udvikling eller udførelse af et projekt. I transformativ evaluering er alle berørte interessenter i udgangspunktet inviteret til at deltage (selv om der må ske en praktisk afgrænsning). Graden af forskellighed og magtdifference følger af dette valg.

Dybden i involveringen (hvor omfattende inddragelsen er i beslutning vedrørende alle evalueringens elementer) er typisk på mellemniveau i pragmatisk evaluering. For eksempel kan man drøfte evalueringens fokus og tilsigtede anvendelse i dybden, mens metode-spørgsmål overlades til evaluatør. Det er nemlig et ømt punkt i partcipatorisk evaluering, hvis ikke metodiske og analytiske opgaver varetages af specialiseret ekspertise (Cousins and Earl 1995).

I transformativ evaluering er der typisk stor dybde i involveringen, fordi man ser samtlige aspekter af evalueringsarbejdet som værdiladede. Derfor er ressourcetrækket blandt interessenterne til deltagelse i evalueringen også stort.

Når man kombinerer dybde i involvering med interessenternes antal, forskellighed, magtdifferentiale og ressourcetræk, bliver transformativ evaluering mere uforudsigelig og rummer et større konfliktpotentiale end pragmatisk evaluering.

Pragmatiske og transformativt participatoriske strategier kan være veje frem i evaluering af innovative tiltag, hvor målopfyldelsesspørgsmål og effektspørgsmål er umulige at besvare eller irrelevante på grund af det innovative tiltags dynamik og kompleksitet. Afgørende for, hvor langt man går i forhold til participatorisk eller måske endda transformativt evalueringsarbejde er ikke mindst, i hvilket omfang det formelle politiske og ledelsesmæssige system overlader innovationsansvar og evalueringsansvar til de involverede interessenter.

EKSPERTBASERET EVALUERING

I ekspertbaseret evaluering delegerer man vurderingen af et innovativt tiltag til en ekspert, som har indsigt i typen af innovation, innovationens fagområde, innovationsprocessen og/eller andre tilsvarende innovationsprocesser, hvorfra man kan overføre erfaringer.

Fordelen ved en ekspertvurdering er, at man kan få en helhedsvurdering af en indsats uden at være afhængig af særlige data om særlige enkeltaspekter, som kan være svære at skaffe. En ekspert kan lave en vurdering på baggrund af en helhedsbetragtning.

Ekspertvurderinger har også svagheder. Ekspertbegrebet er relativt til, hvilket aspekt af en innovation man ønsker at fokusere på. Hvert faglige aspekt eller hver synsvinkel kan være præget af blinde vinkler, fagligt snæversyn eller modeluner inden for feltet. Nogle eksperter synes, de bør være de eneste til at vurdere en sag. Også eksperter har personlige værdier og kæpheste. Derfor kan der være grunde til at sætte et tværfagligt team sammen med flere slags ekspertise, der måske afbalancerer hinanden.

SYSTEMISK EVALUERING

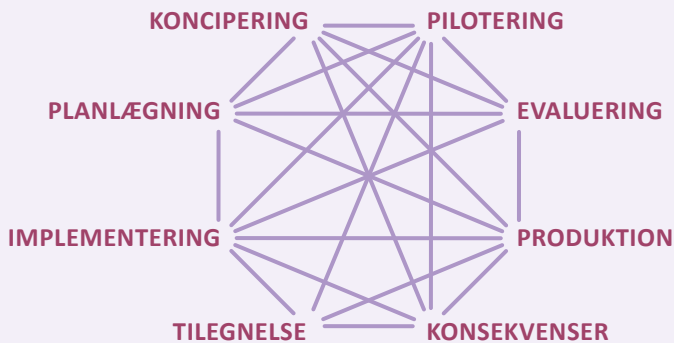
I de senere år er der brudt evalueringstendenser frem inspireret af systemisk tankegang. Hovedideen er her, at hvis en indsats er kompleks og dynamisk, bliver lineær planlægning mindre relevant. Når en indsats bliver uforudsigelig, består den gode håndtering ikke i planlægning forud for indsatsen, men i stedet i nærvær, samspil og dialog under hele processen.

Den dertil svarende evalueringstankegang er, at man ikke kan forhåndstopstille mål, evalueringskriterier og evalueringsspørgsmål. I stedet må man som evaluator være i løbende samspil med det innovative tiltag. Det samme gælder, hvis evalueringen bliver forældet i løbet af evalueringsprocessen, fordi det innovative tiltag udvikler sig meget hurtigt.

Vi kan i øjeblikket kun se konturerne af en sådan evalueringstankegang. Hos Patton er det sådan, at når en indsats bliver kompleks, kan man ikke længere afkode, hvordan den virker og placere ansvaret herfor (Patton 2015). Evaluator må så indtage en helt anden rolle. Man må være i løbende dialog med interessenter involveret i tiltaget. Man må – baseret på kompleksitetsteori og systemteori – medvirke til løbende at korrigere og justere. Man kan fortsat arbejde med en programteori, men den skifter rolle fra at være et sæt af hypoteser, man vil have testet eksakt, til i stedet blot at være en foreløbig læringsramme om interessenternes fælles arbejde med udvikling af tiltaget.

I systemisk evalueringsarbejde opgiver man en lineær opdeling af faser. Ideen er, at man først lærer det komplekse system at kende ved at engagere sig i dets udvikling. Et innovativt tiltag bliver derfor en proces, der sker i samspil mellem momenterne beskrevet i figur 11.

FIGUR 11
MOMENTER I SYSTEMISK EVALUERING



Men eftersom hver af disse faser involverer forskellige interessenter, der går i løbende samspil, må gængse forestillinger om effektmåling og ansvarsplacering opgives. Og det må man sige direkte til opdragsgivere, selv om de forventer såvel effektmåling som ansvarsplacering.

Man kan lære af Pattons markering, at det gælder om at stille spørgsmål til den innovative indsats for at finde ud af, om den er egnet til, at man fortsat står fast på ideer som effektmåling og ansvarsplacering, eller at man skal kaste sig ud i en fundamentalt anden evalueringdagsorden. Dilemmaet for evaluering af innovative tiltag kan ofte være, at man er bedt om at sikre effektivitet og ansvarsplacering samtidig med, at selve indsatsens dynamiske og uforudsigelige udformning ikke tillader at få netop det gjort.

En systemisk evalueringsspraksis indebærer, at evalueringsspørgsmålene sker i en løbende udvikling i samspil mellem interessenter og evaluator. Det udgør en besværlighed i forhold til et politisk styret system. For enten er politikere aktive deltagere i en sådan løbende dialog- og udviklingsproces, eller også delegerer de ansvaret for at formulere evalueringsspørgsmål til de interessenter plus evaluator, der i en given situation tager fælles ansvar for det innovative tiltag. Denne evalueringstankegang er defineret ved, hvordan den er organiseret, ikke ved at den stiller ganske bestemte typer af spørgsmål. Ved valget af denne evalueringstankegang mistes derfor retten til at fastholde, hvilke evalueringsspørgsmål, der skal stilles. Derfor kan det være farligt for en evaluator at indgå i systemisk evalueringssarbejde, medmindre opdragsgiver for evalueringen også selv deltager.

Fordelen ved et systemisk evalueringssarbejde kan være en intens og løbende læreproces, der kan forbedre det innovative tiltag i samspil med de omgivelser, det skal virke i. Men der er også ulemper. Det kan være svært for en opdragsgiver at se data som troværdige, når de er hurtigt indsamlede. Evaluators rolle bliver svær at beskrive og afgrænse, ligesom det kan være svært for en evaluator at se distanceret på tiltagets hidtidige udvikling, hvis vedkommende selv har været en toneangivende del af samme. Forskellen mellem udvikling og evaluering udviskes. Andre vil efterfølgende kunne sige, at evaluator har mistet sin uafhængighed og dermed troværdighed.

Vi har endnu ikke mange gode og dokumenterede erfaringer med systemisk evalueringsarbejde i Danmark.

Et interessant, men ikke uløseligt problem i systemisk evalueringsarbejde er spørgsmålet om, hvordan man sikrer legitime evalueringskriterier, hvis evalueringskriterierne udvikles undervejs.

Er der politisk og organisatorisk spillerum inden for et formelt, repræsentativt demokrati til at udvikle tiltaget løbende og til, at interessenter og evaluator selv formulerer evaluerings-spørgsmål og evalueringskriterier undervejs? Et bud på løsningen af dette problem finder man under overskriften "responsible innovation", på dansk bedst oversat til "ansvarlig innovation" (Stilgoe, Owen og Macnaghten 2013). Ideen er dels at sensitivisere alle, der arbejder med videnskabelige og teknologiske innovationer for et bredere ansvar end det rent tekniske, dels at skabe sociale og politiske fora, hvor dette ansvar artikuleres og bearbejdes.

"Responsible innovation" opererer med fire principper kaldet anticipation, reflexivity, inclusion og responsiveness. Anticipation (foregribelse) er tidligt at stille spørgsmål til det tænkelige og det mulige, som det innovative tiltag lukker op for. Refleksivitet indebærer, at innovatører og deres institutioner stiller spørgsmål til sig selv i lyset af de reaktioner, de får fra andre og i lyset af innovationens potentielle bivirkninger. Inklusion indebærer skabelsen af "mini-offentligheder", hvor innovationen drøftes åbent. Det implicerer ikke blot en invitation af forskellige grupper ind i evalueringsarbejdet, men også en åbning med de dagsordner og bekymringer, de bringer med. Responsivitet består i, at der faktisk lyttes til feedback, og at det innovative tiltag faktisk udviser "korrigerbarhed" som følge heraf.

Responsible evaluation er et bud på, hvordan man undgår, at evaluering af innovation finder sted i et lille ekkokammer mellem de nærmest involverede. Responsible innovation sætter evaluering af det innovative tiltag i et større demokratisk perspektiv. Det perspektiv har et slående sammenfald med det større billede af innovation, som blev antydnet i indledningen til denne vejledning.

6

HVOR KAN MAN LÆRE MERE OM EVALUERING

- Meld dig ind i Dansk Evalueringselskab og deltag i den årlige konference
- Følg kurser, f.eks. Sommeruniversitet på Københavns Universitet
- Lån tidsskrifter på biblioteket såsom "Evaluation" eller "American Journal of Evaluation"
- Læs bøger, f.eks.:

Peter Dahler-Larsen (2013): Evaluering af Projekter – og andre ting, som ikke er ting. Syddansk Universitetsforlag.

Michael Q. Patton (2011): Developmental Evaluation. Guilford Press.

Sue C. Funnell og Patricia Rogers (2011): Purposeful Program Theory. Jossey-Bass.

LITTERATURLISTE

- Andersen, E.S. 2008. *Rethinking Project Management: An Organisational Perspective*, Harlow, Pearson Education Gate.
- Bhatti, Y., Dahlgaard, J.O. Hansen, J.H., & Hansen, K.M. 2015. Getting Out the Vote With Evaluative Thinking. *American Journal of Evaluation*, 36, (3) 389-400.
- Carstensen, Helle Vibeke (2012-13). Effektmåling og evaluering af innovation i SKAT. Økonomistyring & informatik, Årg. 28, nr. 5, 499-521.
- Center for Offentlig Innovation (COI) 2015a. Innovationsbarometeret. Foreløbig baggrundsrapport april 2015.
- Center for Offentlig Innovation (COI). 2015b. Spredning af offentlig innovation. Hvad kan vi lære af forskningen?
- Cousins, J.B. & Earl, L.M. 1995. *Participatory Evaluation in Education: Studies in Evaluation Use and Organizational Learning*, London, Routledge-Falmer.
- Flyvbjerg, B. 2014. What you Should Know About Megaprojects and Why: An Overview. *Project Management Journal*, 45, (2) 6-19.
- Funnell, S.C. & Rogers, P.J. 2011. *Purposeful program theory: effective use of theories of change and logic models*, San Francisco, Jossey-Bass.
- Heartley, J. 2005. Innovation in Governance and Public Services: Past and Presence. *Public Money and Management*, (January) 27-34.
- Ika, L.A. 2009. Project Success as a Topic in Project Management Journals. *Project Management Journal*, 40 (4), 6-19.
- Jugdev, K. & Müller, R. 2005. A Retrospective Look at our Evolving Understanding of Project Success. *Project Management Journal*, 36 (4), 19-31.
- Krogstrup, H.K. 2001. "Brugerinddragelse i evaluering – top-down og bottom-up perspektiver". *Tendenser i evaluering*, P. Dahler-Larsen & H. K. Krogstrup (red.), Odense: Odense Universitetsforlag, 195-210.
- Ledermann, S. 2012. Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33 (2), 159-178.
- McLeod, L., Doolin, B. & MacDonell, S.G. 2012. A Perspective-Based Understanding of Project Success. *Project Management Journal*, 43 (5), 68-86.
- Mikkelsen, H. & Riis, J.O. 2008. *Adræt projektledelse*, Rungsted Kyst, Provedo.
- OECD 2005. Oslo Manual. *Guidelines for Collecting and Interpreting Innovation Data*, 3. udg.
- Patton, M.Q. 2012. A utilization-focused approach to contribution analysis. *Evaluation*, 18(3), 364-377.

- Patton, M.Q. 2011. *Developmental Evaluation*, Guilford Press.
- Patton, M.Q. 2015. Book Review: Evaluating the complex: Attribution, Contribution and Beyond. *American Journal of Evaluation*, 18, (March).
- Smith, M.F. 2005. "Evaluability Assessment". *Encyclopedia of Evaluation*, S. Mathison (red.), Sage.
- Stake, R.E. 2004. *Standard-Based and Responsive Evaluation*, Thousand Oaks, Sage.
- Stilgoe, J. Owen, R. & Macnaghten, P. (2013): Developing a framework for responsible innovation. *Research Policy* 42: 1568-1580.
- Vedung, E. 2009. *Utvärdering i politik och förvaltning*, Lund, Studentlitteratur.



Vejledning om at evaluere innovative tiltag giver medarbejdere og ledere i det offentlige en hjælpende hånd til at evaluere deres arbejde med innovation. Vejledningen henvender sig til ledere, innovationskonsulenter, evalueringskonsulenter, udviklingskonsulenter og projektledere.

Innovation kan udfordre evalueringer i forhold til programmerbarhed, målbarhed og kompleksitet.

Peter Dahler-Larsen fører læseren igennem en række relevante evalueringsmodeller og greb, og giver indsigt i dilemmaer, som kan opstå i mødet mellem evaluering og innovation. *Vejledning om at evaluere innovative tiltag* giver frem for alt handlekompetence til at forsøge at overkomme barrierer og udfordringer i evaluering af innovation.

Få indsigt i:

- hvordan man skaber plads til at støtte det gode samspil mellem evaluering og innovation.
- hvordan man opnår proportionalitet mellem indsats, ressourcer og evaluering.
- hvordan man stiller sig, hvis det innovative tiltag forandrer sig meget under evalueringsprocessen.
- hvor vigtigt det er at tage evaluering ombord tidligt, idet det bidrager til at indsnævre og fokusere hovedidéen.
- hvordan gode succeskriterier opstilles, så man kan styre efter dem og justere innovationen hen ad vejen.
- hvordan man opbygger en organisations evalueringskapacitet.

Med denne udgivelse er arbejdet med at evaluere innovation i den offentlige sektor et skridt videre.