

# EVALUERINGSNYT

NR. 29 MAJ 2013

I DETTE NUMMER:

1 FOREBYGGELSE AF  
UNGDOMSKRIMINALITET  
- AF MORTEN GREVE

2 REFLEKSIONER OM  
EVALUERINGSANVENDELSENS  
KONTEKST  
- AF STEVEN HØJLUND

3 HVORDAN FÅR MAN FORKLARINGER  
MED I PROCESORIENTEREDE  
EFFEKTEVALUERINGER?  
- AF MORTEN ERIKSEN

4 HVOR REALISTISK ER REALISTISK  
SYNTESE?  
- AF CHRISTINA STENTOFT HOXER  
& STEFFEN BOHNI NIELSEN

Dansk Evalueringsselskab  
C/O Institut for Statskundskab  
Aalborg Universitet Fibigerstræde 1  
9220 Aalborg Ø.

Telefon: 9940 8202

E-mail: [sekretariat@danskevalueringsselskab.dk](mailto:sekretariat@danskevalueringsselskab.dk)

[www.danskevalueringsselskab.dk](http://www.danskevalueringsselskab.dk)

**DANSK**  
**EVALUERINGSSELSKAB**



## FOREBYGGELSE AF UNGDOMSKRIMINALITET:

### TIDLIG OPSPORING SOM FAGLIG UDFORDRING

Af Chefkonsulent Morten F. Greve, CFK – Folkesundhed og Kvalitetsudvikling, Region Midtjylland

*Kriminalitet blandt børn og unge er et samfundsproblem forbundet med store direkte og afledte omkostninger, såvel menneskeligt som økonomisk. Både lands- og kommunalpolitisk er man optaget af at styrke den tidlige opsporing af kriminalitetstruede børn og unge, fx via anvendelse af risikobaserede screeningsværktøjer. Aktuell forskning viser dog, at opsporingsopgaven er kompleks, i den forstand at kriminalitetsadfærd ikke kan forudsiges på baggrund af en 'objektiv' risikoscore. Spørgsmålet er, hvordan man, i dette lys, kan understøtte opsporingsindsatsen evaluering- og udviklingsfagligt. Artiklen beskriver de udfordringer, som forskningen udpeger i forhold til det systematiske opsporingsarbejde, men den præsenterer også nogle eksempler på, hvordan disse udfordringer er søgt imødegået i forskellige sammenhænge.*

#### STYRKET TIDLIG OPSPORING

Omfanget af kriminalitet begået af unge har de senere år været faldende set i forhold til antallet af individer i befolkningsgruppen. En anden central udviklingstendens er imidlertid, at mens andelen af helt lovlydige unge er blevet større over tid, har den lille andel af ungegruppen med betydelig og alvorlig kriminalitetsadfærd været stort set uændret over en længere årrække (Balvig 2011; Justitsministeriet 2012). Kommunerne og samfundet som sådan står således med en forebyggelsesmæssig udfordring.

De senere år er der da også blevet sat en offensiv dagsorden med henblik på bl.a. at styrke den tidlige kriminalitetsopsporingsindsats. Siden 2010 er kommunerne blevet forpligtet til at udarbejde kriminalitetsforebyggelsesplaner i forlængelse af den enkelte kommunes sammenhængende børnepolitik – og her spiller en styrkelse af den tidlige opsporingsindsats typisk en central rolle. Som et andet eksempel har Socialstyrelsen siden 2011 i samarbejde med fem kommuner udviklet et screeningsværktøj med sigte på kommunernes tidlige opsporing af kriminalitetstruede børn og unge under overskriften "Projekt Tidlig Identifikation" (se [www.socialstyrelsen.dk](http://www.socialstyrelsen.dk)).

Dette er væsentlige og relevante initiativer, men det er vigtigt samtidig at tage bestik af nogle forbehold over for muligheden af at forudsige kriminalitetsadfærd, som påpeget af bl.a. registerforskningen de senere år.

## KAN KRIMINALITET BLANDT BØRN OG UNGE FORUDSIGES?

Det aktuelle arbejde med at styrke opsporingsindsatsen tager bl.a. afsæt i en forhåbning om, at tidlig risikobaseret identifikation kan gøre det muligt på et nogenlunde 'objektivt' grundlag at udpege og sætte ind over for en kriminalitetsrisiko, før der indtræffer egentlig kriminalitetsadfærd hos barnet. Samtidig er aktørerne på feltet naturligvis opmærksomme på faren for stigmatisering, så det har stor betydning, hvor gode udsigterne er til at opstille præcise prognoser.

SFI har for nylig gennemført et forskningsprojekt, der bl.a. sigtede på at teste, i hvor høj grad danske børn og unges faktiske kriminalitetsadfærd kan forudsiges ud fra den enkeltes samlede risikoprofil på et givet tidspunkt (Christoffersen m.fl. 2011). På baggrund af en systematisk litteratursøgning udviklede forskerholdet en prognostisk model, som bl.a. inkluderer de risikofaktorer, der ifølge forskningen udviser stærkest sammenhæng med efterfølgende kriminalitetsadfærd, herunder manglende skolegang, forbrug af rusmidler og forskellige forhold hos barnets forældre (bl.a. misbrug, vold og arbejdsløshed).

Registerdata vedrørende de ca. 54.500 individer i 1984-årgangen blev bearbejdet, således at det enkelte individs samlede risikobelastning kunne opgøres år for år frem til det 21. leveår. På dette grundlag blev det så beregnet, om risikobilledet kunne give anledning til en prognose om, at vedkommende ville begå førstegangskriminalitet med en vis sandsynlighed det efterfølgende år. Modellens prognoser viste sig dog at være ganske upræcise. I de tilfælde, hvor modellen udpegede en kriminalitetsrisiko, holdt prognosen i mange tilfælde ikke stik. Der var altså en stor andel "falsk positive" blandt de udpegede personer. Og blandt de unge, der ikke forudgående blev udpeget af modellen som kriminalitetstruede, befandt sig til gengæld langt hovedparten af de unge i undersøgelsesgruppen, som faktisk begik kriminalitet. Der var med andre ord en stor andel "falsk negative" blandt de unge, som begik kriminalitet et givet år.

Forskerne havde adgang til uddybende interviewdata vedrørende knapt 3.000 personer fra 1984-årgangen, hvor risikoprognoserne alt andet lige kunne nuanceres og styrkes. Denne specifikation af modellen var imidlertid ikke nævneværdigt mere velfungerende end den registerbaserede version. Den overordnede konklusion på denne del af SFI-studiet var således, at det er vanskeligt at opstille praktisk anvendelige prognoser vedrørende kriminalitetsadfærd på basis af opgørelser af den enkelte persons samlede objektive risikobelastning på et givet tidspunkt (Christoffersen m.fl. 2011: 78-79).

#### RISIKO, SÅRBARHED – OG MODSTANDSKRAFT

Det ovenfor beskrevne undersøgelsesresultat er, kan man hævde, ikke så overraskende i lyset af den udviklingspsykologiske forskning med fokus på de beskyttelses- og sårbarhedsmekanismer, som former det enkelte barns modstandskraft over for belastninger i opvæksten (Borge

2004; Rutter og Rutter 1997). Ifølge denne litteratur medieres risikofaktorer sammenhæng med udvikling af problemadfærd, herunder kriminalitet, af individets modstandskraft. Modstandskraft er ikke en statisk egenskab hos individet, men styrkes og svækkes undervejs i et livsforløb bl.a. i mødet med forskellige risikofaktorer (Socialstyrelsen 2012). Faktisk peger forskningen på, at udvikling af modstandskraft i et vist omfang *forudsætter* kontakt med risikofaktorer. Forældrenes skilsmisse kan således både være en kilde til sårbarhed og modstandskraft for et barn. Som et eksempel på de komplekse sammenhænge mellem risikofaktorer, modstandskraft og kriminalitetsadfærd kan det nævnes, at Christoffersen m.fl. fandt en sammenhæng mellem mobning under skoleopvæksten og *reduceret* sandsynlighed for efterfølgende kriminalitetsadfærd (Christoffersen m.fl. 2011: 90-91).

I lyset af det ovenstående må vellykket tidlig opsporing derfor også betragtes som en krævende faglig ydelse. Frem for alt forudsættes helhedsmæssig dømmekraft hos de lærere, pædagoger, socialrådgivere osv., der i praksis løbende foretager risikovurderinger som led i deres arbejde med børne- og ungegrupper. Afdækning af konkrete belastninger – og ressourcer – i det enkelte barns liv er et vigtigt led i opsporingsarbejdet, men det synes kun i ringe grad at være muligt at 'rutinisere' eller 'objektivere' bedømmelsen af, om der her og nu bør sættes forebyggende ind i forhold til det enkelte barn. Brug af screeningsværktøjer m.m. kan med andre ord ikke stå alene.

Spørgsmålet er, hvilke muligheder interne og eksterne evaluatore, udviklingskonsulenter m.m. har for at arbejde perspektivrigt med at styrke opsporingsarbejdet i landets kommuner?

### EVALUERINGS- OG UDVIKLINGSMÆSSIGE PERSPEKTIVER

Kommunerne kan naturligvis overordnet evaluere, hvor godt man lykkes med opsporings- og forebyggelsesarbejdet ved at monitorere på udviklingen i kriminalitetsomfanget i aldersgruppen såvel som på andre relevante indikatorer – over tid og i forhold til sammenlignelige kommuner. Men netop fordi arbejdet med at opspore kriminalitetstruede børn er så krævende, rent fagligt, er det også et oplagt behov at gå så tæt som muligt på denne fagpraksis med sigte på at understøtte, evaluere og udvikle opsporingsarbejdet. Jeg vil i det følgende kort beskrive to aktuelle eksempler på, hvordan danske kommuner har arbejdet med netop dette sigte. De to konkrete eksempler, som fremhæves nedenfor, er begge kendetegnet ved at indebære et samspil mellem eksterne konsulenter og en praksisnær faglig udviklingsdagsorden sat i hver enkelt kommune.

Det første eksempel stammer fra Københavns Kommune, hvor borgerrepræsentationen i begyndelsen af 2010 besluttede, at der skulle udvikles et screeningsredskab med henblik på identifikation af bekymringsvækkende børn og unge. I forlængelse af denne beslutning satte man et udviklingsarbejde i gang i samarbejde med konsulentfirmaet Oxford Research, som mundede ud i et dialogbaseret redskab til tværfaglig klassegennemgang (se [www.kk.dk/tvaerfagligklassegennemgang](http://www.kk.dk/tvaerfagligklassegennemgang)). Første skridt i udviklingsforløbet var en systematisk litteraturgennem-

gang gennemført af Oxford Research, som i samarbejde med Socialforvaltningen i Københavns Kommune blev omsat til et forskningsbaseret bekymringsredskab med tretten bekymringsfaktorer – en "opmærksomhedsliste". Faktorerne henter belæg i risikoforskningen med fokus på kriminalitet, misbrug og selvskadende adfærd, men er udvalgt efter, at de alle i princippet skal kunne identificeres og afhjælpes af fagfolk i skoleregiet.

På baggrund af et forældresamtykke benyttes denne liste til det første trin i klassegennemgangen, hvor fagpersonerne kort forholder sig til hvert enkelt barn ud fra de tretten bekymringsfaktorer. Formålet er at afgrænse den uddybende gennemgang til de børn, hvis situation kalder på en nærmere drøftelse. Deltagerne i klassegennemgangen er en teamlærer, skolens skolesocialrådgiver (som er mødeleder) og evt. andre lærere og fagpersoner. Den centrale pointe er, at klassegennemgangens efterfølgende trin tager afsæt i kommunens etablerede systemiske platform for fagpersoners helhedsmæssige og reflektive drøftelser af konkrete bekymringssager på børne- og ungeområdet, herunder især Signs of Safety-modellen. Opsporingsværktøjet er derfor ikke udformet med henblik på at beregne en risikoscore, som umiddelbart kan omsættes til en kriminalitetsprognose. I stedet sætter værktøjet en struktureret ramme omkring en drøftelse og vurdering af hvert enkelt barns adfærd, relationer og trivsel. Først uddybes og konkretiseres bekymringen (trin 2), dernæst konkretiseres ressourcer og *undtagelser* fra bekymringen (trin 3), og på dette grundlag tages der så stilling til, om der samlet set er behov for at opstille konkrete forebyggelsesmæssige mål og handlinger i forhold til det enkelte barn.

Socialforvaltningen har evalueret redskabet i praktisk brug på baggrund af en pilotafprøvning på tre skoler (Københavns Kommune 2012). Evalueringen viser, at redskabet er blevet modtaget positivt på to af de tre pilotskoler, mens man på den tredje skole vurderede, at skolen allerede havde velfungerende opsporingsprocedurer, som overflødiggjorde en permanent ibrugtagning af det nye redskab. På baggrund af pilotafprøvningen vedtog Børne- og ungdomsforvaltningen og Socialforvaltningen medio 2012, at alle skoler skulle gennemføre sociale tjek/klassegennemgang i tværfagligt regi med fokus på de enkelte elevers faglige såvel som sociale forhold. Kravet er, at man anvender en systematisk metode, men skolerne er ikke pålagt at bruge et bestemt redskab. Forventningen er dog, at de fleste skoler vil anvende det udviklede redskab.

Det andet eksempel på aktuelle forsøg på at styrke opsporingsarbejdet, som kort skal nævnes her, starter så at sige i den anden ende af kæden. Siden 2010 har CFK som VISO-specialister samarbejdet med en række kommuner om at lægge et videns- og evalueringsgrundlag for udarbejdelsen af de ovennævnte kriminalitetsforebyggelsesplaner. I flere af disse forløb har kommunerne valgt at interessere sig særligt for unge, som for nylig har modtaget domme for alvorlig kriminalitet.

Der er i den forbindelse udviklet en undersøgelsesmodel, som via tre gensidigt understøttende led fører frem mod en nuanceret, men også ganske koncis afdækning af sagsforløbene med sigte på de mere generelle læringsperspektiver. Modellen tager udgangspunkt i en *journalanalyse*, hvor de ofte meget omfattende skriftlige sagsmateriale bearbejdes i form af komprimerede, kronologisk ordnede notater med fokus på problemindikationer, sagsbehandling og udredning, samt

iværksatte foranstaltninger. På baggrund af journalanalyse gennemføres semistrukturerede *interview* med den unge samt en af eller begge den unges forældre, hvor der blandt andet er særligt fokus på at få interviewpersonernes oplevelse og vurdering af det tidlige sagsforløb. Det sidste led i modellen er en *faglig auditering* af et par af de belyste sagsforløb på basis af journalanalyser og interviewreferater. Et auditforløb baserer sig på tre grundelementer: Et tværfagligt auditpanel af udvalgte fagfolk (typisk 6-8 inviterede deltagere); paneldeltagernes forudgående faglige vurdering af sagsforløbet ud fra det skriftlige materiale og i henhold til fastsatte vurderingskriterier (opstillet i et vurderingsskema); og en struktureret mødeform, hvor der tages udgangspunkt i de individuelle vurderinger som afsæt for en fælles drøftelse. Den fælles drøftelse afsøger mulighederne for at formulere generelle anbefalinger med hensyn til god praksis på centrale områder (se Isager 2009).

I forhold til artiklens tema er det interessante, at denne del af VISO-forløbene – i form af afrapportering såvel som faglige drøftelser – har vist sig at have en værdi i retning af at understøtte faglig udvikling, ikke blot i forhold til den sene del af et sagsforløb, men i høj grad også i forhold til de første tegn på sociale og adfærdsmæssige problemer hos barnet eller den unge, og de faglige reaktioner på samme.

### AFRUNDING OG PERSPEKTIVERING

Denne artikel tematiserer nogle påtrængende udfordringer i forbindelse med kommunernes arbejde med tidlig forebyggelse af kriminalitetsadfærd hos børn og unge – og præsenterer to bud på, hvordan der i forskellige sammenhænge er gjort forsøg på at understøtte og kvalificere det helhedsmæssige opsporingsarbejde i kommunerne. Det er bestemt ikke hensigten at postulere, at de to ovenfor beskrevne eksempler i sig selv kan bringe os i mål i forhold til at imødegå udfordringerne. Hensigten er snarere at give et signalement af, hvordan en konstruktiv reaktion på udfordringerne *kan* se ud. Der er givetvis brug for at afsøge og afprøve mange forskellige stier mod målet: En forebyggelsespraksis som tager kompleksiteten i opsporingsopgaven alvorligt.

### LITTERATURHENVISNINGER

Balvig, F. (2011): "Lovlydig ungdom", Glostrup: Det Kriminalpræventive Råd.

Borge, A. (2004): "Resiliens. Risiko og sund udvikling", København: Hans Reitzel Forlag.

Christoffersen, M. m.fl. (2011): "Tidlig identifikation af kriminalitetstruede børn og unge", København: SFI – Det Nationale Forskningscenter for Velfærd, 11:34.

Isager, A. (2009): "Om audit. En metode til læring og kvalitetsudvikling i fagpraksis", Aarhus: CFK, Region Midtjylland.

Justitsministeriet (2012): "Udviklingen i omfang og art af ungdomskriminalitet 2002-2011", København: Justitsministeriets Forskningskontor.

Københavns Kommune (2012): "Evaluering af pilotprojektet: Systematisk tværfaglig klassegennemgang i 4. klasse, København: Socialforvaltningen.

Rutter, M. og M. Rutter (1997): "Den livslange udvikling. Forandring og kontinuitet", København: Hans Reitzels Forlag.

Socialstyrelsen (2012): "Forebyggelse af kriminalitet blandt børn og unge. En videns- og erfaringsopsamling", Odense: Socialstyrelsen og CFK, Region Midtjylland.

# REFLEKSIONER OM EVALUERINGSANVENDELSENS KONTEKST

Af Steven Højlund, Erhvervs-PhD-Stipendiat, Copenhagen Business School & COWI A/S

*Denne artikel er baseret på et oplæg afholdt på DES's årskonference 2012 i Kolding. Artiklen er teoretisk og stiller spørgsmål ved, om evalueringsslitteraturen har et tilfredsstillende svar på, hvorfor evaluering somme tider ikke anvendes eller kun anvendes symbolsk. Desuden forsøger artiklen at opstille en model, som tager organisatoriske kontekstuelle faktorer med i betragtning, når evalueringsanvendelse forklares. Formålet med artiklen er at tilskynde til refleksion angående evalueringens anvendelse og ikke-anvendelse i organisationer.*

## INTRODUKTION

Evalueringsslitteraturen har ved mange lejligheder påvist at evalueringer sjældent har indflydelse på udformningen af politik (Patton, 1997; Pawson og Tilley, 1997: 2-3; Shulha og Cousins, 1997). Dette udgør et paradoks, eftersom selve formålet med evaluering er at forbedre politik fremadrettet. Denne artikel tager udgangspunkt i dette paradoks. For det første, hævdes det at ovennævnte paradoks er uløseligt forbundet med evalueringens væsen. For det andet hævder artiklen at kunne afhjælpe paradokset ved at forklare evalueringsanvendelse med faktorer, der relaterer sig til den evaluerende organisations kontekst i stedet for evalueringens kontekst. Der er med andre ord tale om, at man tager de organisatoriske faktorer i betragtning, som har indflydelse på, hvordan en organisation anvender viden.

## EVALUERINGENS RATIONELLE OG KAUSALE VÆSEN

Evaluering blev født i en tid, der troede på en bedre verden gennem rationelle politiske indgreb og social ingeniørkunst (Vedung, 2010). Evaluering opfattes traditionelt som et forvaltningsredskab, der leverer feedback til politikere og embedsmænd om resultaterne af offentlige interventioner. De fleste evalueringsmodeller er baseret på ideer om årsags-virkningssammenhænge almindeligt forekommende i klassisk økonomisk tænkning og rationel-choice teori, hvor aktøren anses for at være rationel og nyttemaksimerende (se f.eks. Sanderson, 2000). Disse antagelser indeholder en underliggende positiv og evolutionær antagelse om fremskridt og forbedring ('social betterment') (Henry og Mark, 2004; Dahler-Larsen, 2012).

Evalueringens rationelle og kausale væsen indebærer to logiske problemer, som er relateret til det ovennævnte paradoks.

### PROBLEM NR. 1

Det første problem opstår, da formålet med evaluering er tæt knyttet til ideer om kausalitet og positiv evolutionær udvikling. Dette er tydeligt i standarddefinitioner på evaluering: Evaluering defineres almindeligvis som en "*retrospektiv bedømmning av fattade beslut, deras genomförande (implementering), slutprestationer eller utfall i offentlig verksamhet, vilken avses spela en roll i praktiska beslutssituationer.*" (Vedung, 1997: 3; se også f.eks. Cousins et al, 2004: 105). Det fremgår klart af ovenstående, at anvendelsen af en evaluering er en iboende del af selve evalueringens væsen. Men dette indeholder også et logisk problem, idet en uanvendt evaluering i så fald ikke er en evaluering jævnfør definitionen. Ikke alle definitioner af evaluering omfatter et formål. Ikke desto mindre formuleres formålet med evaluering stort set altid i tråd med ovenstående (f.eks. at forbedre de politiske indgreb, læring og/eller vurdere værdien af interventioner) (se f.eks. Henry et al., 2000).

### PROBLEM NR. 2

Litteraturen vedrørende evalueringsanvendelse fandt tidligt frem til fire typiske anvendelser af evalueringer: *Instrumental-, begrebsmæssig/konceptuel-, proces- og symbolsk anvendelse* (Alkin og Taut, 2003; Kirkhart, 2000) (inddelingen er blevet kritiseret en del for at være forsimplet, men den diskussion ligger uden for denne artikels fokusområde). Den grundlæggende antagelse om evalueringsanvendelse er, at evaluering anvendes instrumentelt til at forbedre politik gennem evalueringens resultater (konklusioner og anbefalinger). Begrebsmæssig/konceptuel- og procesanvendelse regnes ligeledes for anvendelsestyper, der kan forbedre politik gennem henholdsvis begrebsmæssig/konceptuel læring og tilegnelse af viden om evaluering under evalueringprocessen. Problemet opstår med den symbolske anvendelsesform.

Symbolsk evalueringsanvendelse sker, når evaluering bliver anvendt til at legitimere den evaluerende organisation eller organisationens aktiviteter overfor omverdenen. Litteraturen skelner mellem symbolsk- og legitimerende anvendelse (Alkin and Taut, 2003). Disse to typer af anvendelser blev formuleret af Pelz (1978) og er siden blevet dokumenteret til overflod. Problemet opstår, fordi symbolsk- og legitimerende evalueringsanvendelse ikke bidrager til politikudvikling eller samfundsmæssige forbedring på længere sigt, som det ellers *a priori* forudsættes, at evaluering gør. F.eks. bliver vandkvaliteten ikke forbedret, hvis en hjælpeorganisation kun anvender sin evaluering til at legitimere sig selv overfor sin donor.

### LØSNING SØGES

Paradokset, der udgøres af de to problemer forklaret ovenfor, er velbeskrevet i evalueringsslitteraturen (se f.eks. Kirkhart, 2000). Ikke desto mindre er der ikke fundet en løsning, og det må siges at være utilfreds-

stillende. Hidtil har størstedelen af evalueringsslitteraturen om evalueringsanvendelse fokuseret på identificering og diskussion af anvendelsestyper (som de fire præsenteret tidligere) og betingelser for evalueringsanvendelse. Med andre ord har litteraturen fokuseret på evalueringen og evalueringens kontekst, men i mindre grad på den organisatoriske kontekst, som evalueringen foregår i.

At dette er tilfældet bliver tydeligt, hvis man læser de oversigtsartikler (reviews), der er lavet af den litteratur, som omhandler evalueringsanvendelse. Et eksempel er tre af litteraturens nok væsentligste oversigtsartikler, som tilsammen dækker perioden fra 1971 til 2005 (Leviton og Hughes, 1981; Cousins og Leithwood, 1986; Johnson et al.; 2009). De giver et overblik over de væsentligste artikler skrevet om de faktorer, som ligger til grund for evalueringsanvendelse. De tre oversigtsartikler arbejder på det, man kan kalde et 'mikro-teoretisk' forklaringsniveau, hvor der spørges til, 'hvilke omstændigheder er til stede, når evaluering anvendes?' F.eks. konkluderes det, at evalueringskvalitet og kommunikationen af evalueringsresultater er væsentlige forudsætninger for, at evalueringer bliver anvendt. Men disse og andre oversigtsartikler forklarer ikke, hvorfor evalueringer ikke bliver anvendt, og omtaler ikke de organisatoriske rammer i relation til evalueringsanvendelse. Det er måske grunden til, at der er kommet mere fokus på det kontekstuelle perspektiv i det senere år. F.eks. konkluderer Ledermann (2011: 160): "...det er på tide at opgive ambitionen om at finde 'den vigtige' karakteristika for anvendelse og i stedet fokusere på kontekst-bundne mekanismer." Peter Dahler-Larsen (2012), Sanderson (2000) og Van der Knaap (1995) har også argumenteret i retning af større fokus på evalueringens bredere kontekst.

Argumentet i denne artikel er, at organisatoriske faktorer må tages i betragtning for at forklare symbolsk- og legitimerende evalueringsanvendelse samt den til tider manglende anvendelse. Hvis man søger at forklare evalueringsanvendelse udelukkende med et fokus på evalueringen, dens gennemførelse, dens resultater osv., ignorerer man den organisatoriske kontekst, hvor andre faktorer kan være på spil. Disse overordnede faktorer kan – hævder jeg – være mere væsentlige end faktorerne relateret specifikt til evalueringen.

Hvis man søger svar i den akademiske litteratur på ovenstående problemstilling, så er manglende anvendelse og symbolsk- og legitimerende anvendelse af evaluering i god tråd med centrale argumenter i organisatorisk institutionel teori. Selznick (1949) og et stort antal senere empiriske undersøgelser, viser at organisationer under visse omstændigheder handle i strid med deres formål og den rationelle nyttemaksimerende adfærd, der forventes af dem. I stedet handler organisationer efter normer og værdier i deres miljø med henblik på at legitimere sig selv (Zald og Denton, 1963). En interessant konklusion i disse tidlige undersøgelser er, at organisationer tilpasser sig den praksis, der er legitimeret af de organisatoriske omgivelser (Meyer og Rowan, 1977). Dette er især tilfældet, hvis organisationen opfatter omgivelserne som ustabile og usikre i forhold til f.eks. fordelingen af ressourcer mellem sig selv og rivaliserende organisationer (DiMaggio og Powell, 1991: 30-1). DiMaggio og Powell (1991: 190) hævder, at organisationer primært agerer for at stabilisere sociale relationer og dermed ikke primært agerer for at maksimere magt eller nytte som oftes antages.

Mærkeligt nok, er disse pointer fra organisatorisk institutionel teori næsten ikke reflekteret i evalueringsslitteraturen i relation til evalueringsanvendelse. De allerede nævnte oversigtsartikler samt Cousins et al. (2004) and Amo og Cousins (2007) behandler ikke institutionel teori. To systematiske søgninger<sup>1</sup> i samtlige udgivelser af tidsskrifterne 'Evaluation', 'American Journal of Evaluation', 'Evaluation Review' og 'Canadian Journal of Programme Evaluation' resulterede i 24 resultater, hvoraf kun en artikel forsøgte at forklare evalueringsanvendelse med antagelser fra organisatorisk institutionalisme (Eckerd and Moulton, 2011).

## DE ORGANISATORISKE FAKTORER

Man forbinder typisk en række begreber, metoder, praksisser, principper, normer, værdier og artefakter med evaluering. Med andre ord kan man godt tale om en 'evalueringsinstitution'; en praksis eller en norm, man tager for givet, og som er relativt stabil. For organisatorisk institutionalisme er fokus på institutioner helt centralt. En institution bør ikke forveksles med en organisation. Organisationen er fysisk og udgøres af et kollektiv af individer, mens en institution sætter rammerne for dette kollektivs handlinger.

En organisation forandrer sig ved at tilpasse sine institutioner over tid. Det samme sker med 'evalueringsinstitutionen'. For ikke så mange år siden, var evaluering ukendt i mange offentlige organisationer. Men langsomt forandrede det sig, og offentlige organisationer blev ramt af evalueringsbølgen. Argumentet i denne artikel er, at årsagene til evalueringens institutionalisering også kan forklare, hvorfor en given organisation anvender (eller ikke anvender) sine evalueringer.

## EKSTERNT PRES FRA OMGIVELSERNE

Organisatorisk institutionalisme er antager, at organisationer sikrer deres legitimitet ved at opfylde omgivelsernes forventninger til deres institutioner (struktur, praksis, retorik eller aktiviteter). Organisationer gør dette for at mindske usikkerhed og stabilisere sociale relationer (DiMaggio og Powell, 1991: 19). DiMaggio og Powell (1983: 147-160) identificerer tre hovedprocesser for institutionel tilpasning:

- 1) *Tvangsmæssig tilpasning* sker, når en organisation er begrænset af generelle kulturelle forventninger (f.eks. manifesteret i love) og må tilpasse sig dem for at opretholde sin legitimitet og undgå sanktioner;
- 2) *Efterlignende tilpasning* sker, når en organisation opererer i usikre omgivelser (f.eks. ustabil politisk støtte, kamp om knappe ressourcer). Som følge deraf vil organisationen forsøge at begrænse disse usikkerheder ved at efterligne vellykkede organisationers praksis;

<sup>1</sup> Søgeordene 'institutionalism' og 'neo-institutional\*' blev anvendt i to søgninger i alle kategorier (f.eks. 'references', 'abstract', 'title', 'text' etc.)

3) *Normativ tilpasning* sker, når en organisation påvirkes af normative forventninger i relation til professionalisering og uddannelse af personale. Tilpasningen i organisationen kan ske, hvis organisationen anvender eksterne konsulenter eller ansætter nye ledere.

### INTERN EVALUERINGSSTILBØJELIGHED

En organisations tilpasning afhænger ikke blot presset fra de organisatoriske omgivelser, men også af interne faktorer i organisationen. Nils Brunsson skelner mellem 'handlingsorganisationen' og den 'politiske organisation' som to yderpunkter på et kontinuum (Brunsson, 2002). En handlingsorganisation er fokuseret på at øge sin effektivitet, fordi den får sin legitimitet fra omgivelserne ved at producere konkrete produkter eller services. Denne type organisation er derfor tilbøjelig til at bruge viden instrumentelt til at forbedre sin effektivitet og dermed øge sin legitimitet. I modsætning til handlingsorganisationen, får den 'politiske organisation' sin legitimitet fra politiske beslutninger og handlinger.

Denne type organisation har en lav tilbøjelighed til at evaluere, fordi de politiske resultater og virkninger ofte ikke er målbare og er sværere at evaluere.

Politiske organisationer har tendens til at bruge viden - såsom evalueringer - symbolsk snarere end instrumentelt for at legitimere sig selv. To empiriske undersøgelser med relevans for evaluering bekræfter Brunssons antagelser (Boswell, 2008; Eckerd og Moulton, 2011).

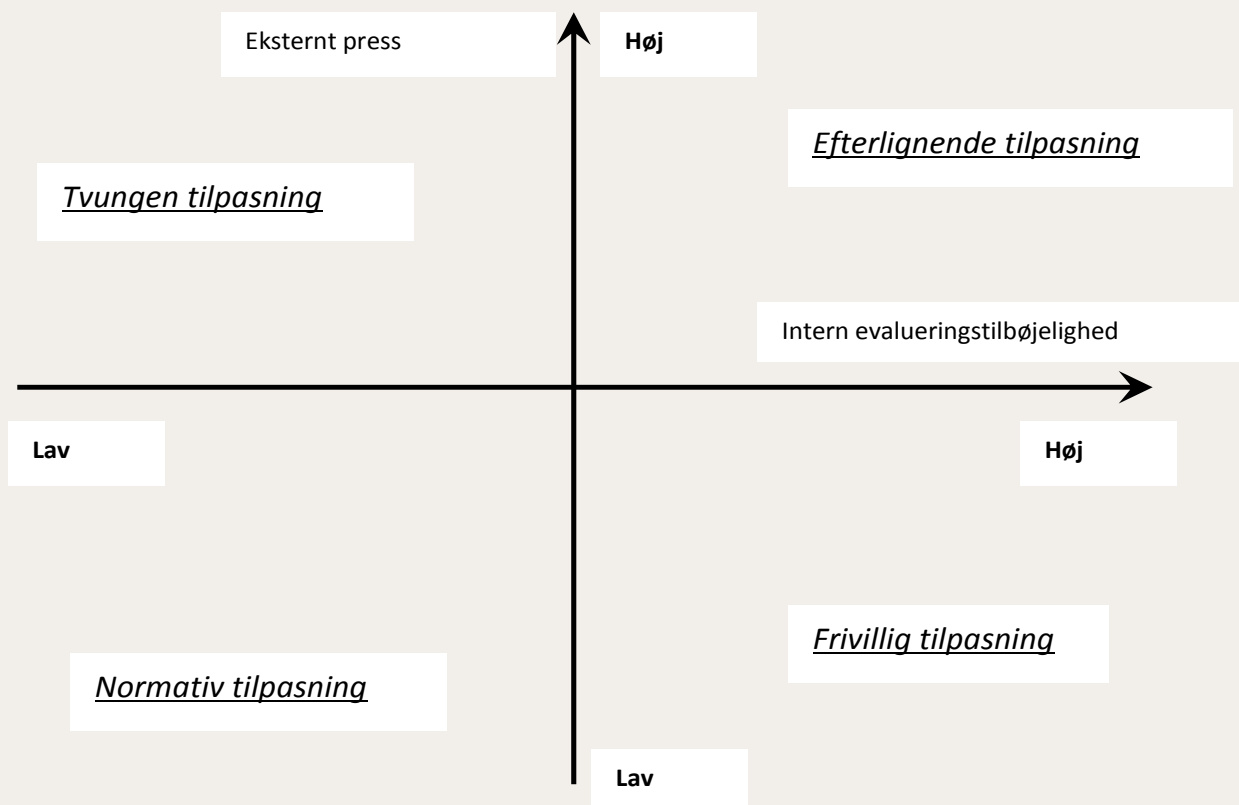
I det følgende afsnit præsenteres en model med eksternt pres og intern tilbøjelighed til at evaluere som to forklarende variable. Artiklens pointe er, at disse to faktorer er vigtige i forhold til; 1) hvordan en organisation tilpasser sig til sine omgivelser (institutionalisering af evaluering); og 2) på hvilken måde organisationen anvender evalueringer efterfølgende.

### MODELLEN

På basis af de to kontinua beskrevet ovenfor, kan man opstille en model med fire 'ekstreme positioner'. Disse er: *Tvungen tilpasning*, *efterlignende tilpasning*, *normativ tilpasning* og *frivillig tilpasning*. De første tre er inspireret af organisatorisk institutionel teori. Sidstnævnte repræsenterer en tilpasningsform, der er i overensstemmelse med den generelle antagelse om en rationel organisation. Pointen er, at hver idealtypisk tilpasningsform også vil anvende evaluering forskelligt fra de andre tilpasningsformer.

Figur 1 nedenfor viser de to kontinua på de to akser og placerer de fire idealtypiske tilpasningsformer i hvert sit hjørne.

Figur 2: De fire idealtypiske tilpasningsformer



I de følgende afsnit eksemplificeres de fire idealtypiske tilpasningsformer.

### TVUNGEN TILPASNING

Tvungen tilpasning af evalueringspraksis finder sted i en organisation, der; 1) er under pres fra sine organisatoriske omgivelser; og 2) som har en lav evalueringstilbøjelighed. Formelle evalueringspraksisser vil være på plads for at tilfredsstille f.eks. donorer, men paradoksalt nok søger organisationen ikke instrumentelt at anvende sine evalueringer. Hvis organisationen anvender sine evalueringer, vil det højst sandsynligt være symbolsk for at legitimere sig selv. Dette forklarer den adfærd, som almindeligvis betragtes som et paradoks i litteraturen om evalueringsanvendelse, og som organisationslitteraturen refererer til som 'afkobling' af en struktur eller praksis (f.eks. evalueringspraksis) fra kernefunktionerne i organisationen (de funktioner, som er forbundet med organisationens output).

Et eksempel på en sådan adfærd kunne være en mindre NGO, der arbejder med udviklingsbistand. Organisationen er helt afhængig af donorer og arbejder i et ustabil og politiseret felt, hvor organisationens finansiering er usikker og ustabil. Desuden er organisationen lille og overbelastet og afhænger langt hen ad vejen af de ansattes idealisme. Personalet vil højst sandsynligt finde donorenes krav om evaluering kontraproduktive ift. NGO'ens praksis og mål.

### EFTERLIGNENDE TILPASNING

Efterlignende tilpasning af evalueringspraksis foregår i organisationer, hvor evalueringstilbøjeligheden er stor, men hvor presset for at evaluere fra omgivelserne også er stort. Fordi organisationen er meget modtagelig over for evaluering, institutionaliserer den nemt en evalueringspraksis. Den vil efterligne praksis fra de bedste i feltet for at lære at blive mere effektiv. Men organisationen arbejder også i usikre omgivelser med kamp om ressourcerne, og organisationen søger derfor også at øge sin legitimitet ved at vise omgivelserne, at dens aktiviteter er legitime.

Resultatet er, at organisationen ikke vil afkoble (som er normalt i den tvungne tilpasningsform), men snarere bruger evalueringer aktivt for at legitimere sine programmer eller politikker (i modsætning til at legitimere selve organisationen, som det var tilfældet i den tvungne tilpasning). Et eksempel på kopierende tilpasning kunne være en skole eller et universitet. Uddannelsesinstitutioner opererer i et ustabil felt. Undervisningsmetoder og læseplaner diskuteres næsten dagligt i offentligheden og myndighederne er tvunget til at træffe foranstaltninger i overensstemmelse hermed eller som reaktion på international benchmarking osv. Samtidig, har uddannelsesorganisationer en målekultur (f.eks. eksamener). Evalueringer forventes derfor at blive brugt både til at tilfredsstille offentligheden (legitimerende ift. aktiviteter) samt instrumentalt for at øge effektiviteten.

### NORMATIV TILPASNING

Normativ tilpasning finder sted via spredning af ideer i faglige netværk. Ledere skifter stillinger, konsulenter spreder modeller i organisationer osv. En organisation som tilpasser normativt er under lavt pres udefra og har lav tilbøjelighed til at evaluere. Det er derfor ikke sandsynligt, at organisationen i udgangspunktet ville adoptere evaluering. Når organisationen evaluerer, er det således mere tilfældigt og knyttet til en bestemt leders idé, end det er som følge af legitimitetssøgende adfærd. Evaluering i en sådan organisation vil være svagt institutionaliseret, og når evalueringer udføres, er den mest almindelige anvendelse den begrebsmæssige/konceptuelle anvendelse, hvor organisationen lærer nye metoder, eller ingen anvendelse. Evalueringsforløb og -resultater vil variere meget i kvalitet og indhold på grund af manglende organisatorisk erfaring med evaluering. Et eksempel på normativ tilpasning kunne være en efterretningstjeneste, der er under relativt lille eksternt pres (pga. sit monopol på samfundsopretholdende sikkerhedsydelse), og som ikke har en intern tilbøjelighed til at evaluere (se f.eks. undersøgelse af intern evaluering i FBI i Sonnichsen, 1999).

### FRIVILLIG TILPASNING

Frivillig vedtagelse af evalueringspraksis foregår i organisationer med et svagt eksternt pres og en stærk evalueringstilbøjelighed. Evalueringens praksis i organisationen vil sandsynligvis være uformel og ad hoc snarere end meget regelbaseret, da formelle regler kun anvendes i det omfang, de giver mening for det overordnede formål, som er at øge effektiviteten. Organisationen anvender evalueringer instrumentelt for at øge effektiviteten. Frivillig tilpasning er den idealtypiske tilpasning, der normalt antages i evalueringens litteraturen.

### KONKLUSION

Denne artikel omhandler det paradoks, at evalueringer sjældent bidrager til den politiske beslutningsproces på trods af, at det er evalueringens formål. Paradokset forklares med evalueringens rationelle, kausale og evolutionære væsen, som er uforenelige med symbolske evalueringens anvendelsestyper og 'ikke-anvendelse'. Artiklen finder ingen løsninger på dette problem i litteraturen. Derfor benytter artiklen organisatorisk institutionel teori til at opstille en model, som kan forklare symbolsk og legitimerende anvendelse samt 'ikke-anvendelse'.

Artiklens konklusion er, at forklaringer på evalueringens anvendelse i højere grad bør medtage de organisatoriske rammer i den evaluerende organisation frem for kun at se på evalueringens muliggørende faktorer, som det i høj grad er tilfældet på nuværende tidspunkt. Det er artiklens påstand, at vi ved at komplementere 'mikro-niveaet' med et mere overordnet organisatorisk analyseniveau kan forklare symbolske evalueringens anvendelser og ikke-anvendelse bedre.



## REFERENCER

- Alkin, M. and Taut, S. (2003) 'Unbundling Evaluation Use', *Studies in Educational Evaluation* 29(1): 1-12.
- Amo, C. and Cousins, J. B. (2007) 'New Going Through the Process: An Examination of the Operationalization of Process Use in Empirical Research on Evaluation', *New Directions for Evaluation* 116(1): 5-26.
- Brunsson, N. (2002) *The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations*. Copenhagen: Abstrakt & Liber.
- Cousins, J. B. and Leithwood, K. A. (1986) 'Current empirical research on evaluation utilization', *Review of Educational Research* 56(3): 331-364.
- Cousins et al. (2004) 'Integrating evaluative inquiry into the organisational culture: A review and synthesis of the knowledge base', *The Canadian Journal of Program Evaluation* 19(2): 99-141.
- Dahler-Larsen, P. (2012) *The Evaluation Society*. Palo Alto: Stanford Press.
- DiMaggio, P. J. and Powell, W. W. (1983) 'The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields', *American Sociological Review* 48(2): 147-160.
- DiMaggio, P. J. and Powell, W. (1991) *Introduction* in Powell, W. W. and DiMaggio, P. J. (ed.) *The new institutionalism in organizational analysis*, pp. 1-38. Chicago: University of Chicago Press.
- Eckerd, A. and Moulton, S. (2011) 'Heterogeneous Roles and Heterogeneous Practices: Understanding the Adoption and Uses of Nonprofit Performance Evaluations', *American Journal of Evaluation* 32(1): 98-117.
- Henry, G. T. and Mark, M. M. (2003) 'Beyond Use: Understanding Evaluation's Influence on Attitudes and Actions', *American Journal of Evaluation* 24(3): 293.
- Johnson, Kelli et al. (2009) 'Research on Evaluation Use - A review of the Empirical Literature From 1986 to 2005', *American Journal of Evaluation* 30(3): 377-410.
- Ledermann, S. (2011) 'Exploring the necessary conditions for evaluation use in program change', *American Journal of evaluation* 33(2): 159-78.
- Leviton, L. C., and Hughes, E. F. X. (1981) 'Research on the utilization of evaluations: A review and synthesis', *Evaluation Review* 5(4): 525-548.
- Mark, M. M.; Henry, G. T. and Julnes, G. (2000) *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco: Jossey Bass.
- Mark, M. M. and Henry, G. T. (2004) 'The Mechanisms and Outcomes of Evaluation Influence', *Evaluation* 10(1): 35-57.
- Meyer, J. W. and Rowan, B. (1977) 'Institutionalized organizations: Formal structure as Myth and Ceremony'. *American Sociological Review* 83(2): 340-63.
- Meyer, J. W.; Rowan, B. (1991) 'Institutionalized Organizations: Formal Structure as Myth and Ceremony in Powell, W. W.; DiMaggio, P. J. (1991) *The New Institutionalism in Organizational Analysis*, pp. 41-62. The University Chicago Press.
- Kirkhart, K. E. (2000) 'Reconceptualizing Evaluation Use: An Integrated Theory of Influence', *New Directions for Evaluation* 88: 5-23.
- Patton, M. Q. (1997) *Utilization-focused evaluation: The new century text (3rd ed.)*. Thousand Oaks, CA: Sage.
- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: SAGE Publication Ltd.
- Pelz, D. C. (1978) 'Some expanded perspectives on use of social science in public policy' in Yinger, J. M. and Cutler, S. J. (ed.) *Major Social Issues: A Multidisciplinary View*, pp. 346-357. New York: Macmillan.
- Sanderson, I. (2000) 'Evaluation in Complex Policy Systems', *Evaluation* 2000 6(4): 433-454.
- Selznik, P. (1949) *TVA and the Grass Roots*. Berkeley: University of California Press.
- Shulha, L. M. and Cousins, B. (1997) 'Evaluation Use: Theory, Research and Practice Since 1986', *Evaluation Practice*, 18(3): 195-208.
- Sonnichsen, R. C. (1999) *High Impact Internal Evaluation: A Practitioner's Guide to Evaluating and Consulting Inside Organizations*. UK: Sage Publications.
- Van der Knaap, P. (1995) 'Policy Evaluation and Learning: Feedback, Enlightenment or Argumentation?', *Evaluation* 1(2): 189-216.
- Vedung, E. (1997) *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Vedung, E. (2010) 'Four Waves of Evaluation Diffusion', *Evaluation* 16(3): 263-277.
- Zald, M. N. and Denton, P. (1963). 'From Evangelism to General Services: the Transformation of the YMCA', *Administrative Science Quarterly* 8: 214-234.

# HVORDAN FÅR MAN FORKLARINGER MED I PROCESORIENTEREDE EFFEKTEVALUERINGER?

Af Morten Eriksen, seniorprojektleder i KORA<sup>1</sup>

*Mekanismer er centrale til at forklare resultater i procesorienterede effektevalueringer, men evaluatører er ikke enige om, hvad mekanismer er. Så hvordan får vi indkredset og operationaliseret mekanismer i teoribaserede evalueringer?*

## HVORFOR MEKANISMER I EVALUERING?

Når vi designer og effektevaluerer offentlige indsatser, er det centralt at få indkredset mekanismerne for indsatsen. Hvorfor bidrager indsatsen til resultater? Hvad er mekanismene mellem aktiviteter og problemløsninger? Det er vigtigt at gøre disse antagelser tydelige, så man undgår at iværksætte aktiviteter, der ikke bidrager til resultater – såkaldte aktivitetsfælder. Det vil være økonomisk og fagligt uansvarligt at investere ressourcer i indsatser, hvor der ikke er en begrundet forestilling om, at de kan gøre en forskel. Man vil kunne dokumentere, at man gennemfører de planlagte aktiviteter, men man vil ikke kunne dokumentere, at man gør det rigtige.

Evalueringsteorien betoner da også i den procesorienterede effektevaluering, at det er vigtigt, at mekanismer bliver tydeliggjort og testet – holder antagelse om hvorfor indsatsen giver resultater for bestemte målgrupper i bestemte sammenhænge (Pawson og Tilley 1997, Pawson 2007, Pawson 2013)?

Centralt i den procesorienterede effektevaluering er det forklarende aspekt, at koble årsag og virkning, og evalueringsteorien nævner i den sammenhæng "the black box problem", hvor evalueringer ikke har det forklarende aspekt med i deres undersøgelser (Mahoney 2003, s. 1, Pawson og Tilley 1997, s. 30, Mark m.fl. 2011, s. 13). Dette problem kan man opfatte som en manglende analyse af implementeringen af indsatser: aktiviteter og ydelser, men problemet kan også anskues som en manglende analyse af de mekanismer, der leder til resultater. Det er ikke selve indsatsen, der fører til resultater, men de mekanismer, som indsatsen er påtænkt at påvirke, som fører til resultater.

Så det er vigtigt at indkredse mekanismer, men der er dog ikke enighed om, hvad mekanismer er (Gerring 2010, s. 1500-1502, 1510, Bunge 1997, s. 414, Hedström 2009, Mahoney 2001, s. 578-580, Gerring 2007, s. 165, Little 2012). Den australske evalueringsforsker Brad Astbury og den hollandske evalueringsforsker Frans Leeuw har i en vigtig artikel gjort os opmærksomme på, at mekanismer er et tvetydigt begreb, som bruges forskelligt af evaluatører og forskere (Astbury og Leeuw 2010).

Som evaluator har jeg set mekanismer som de variable, der medierer indsatsen frem til outcome eller problemløsning (KREVI 2008). Men ifølge førnævnte forskere kan man ikke betragte mekanismer som medierende variable, da mekanismer som oftest er ikke-observerbare enheder, der ikke kan anvendes på samme måde som variable i statistiske analyser.

Spørgsmålet er så, hvad det er, vi måler på i teoribaserede evalueringer? Får vi fat i mekanismerne i vores procesorienterede effektevalu-

eringer, eller måler vi på noget andet? Kan vi dermed stille spørgsmålstegn ved de teoribaserede evalueringers validitet – får vi testet og belyst kausalsammenhænge? Får vi det forklarende aspekt med? Eller bliver der i højere grad målt på implementeringen af indsatser? Artiklen forsøger først at svare på, hvad mekanismer er i procesorienterede effektevalueringer, og dernæst forsøger den at svare på, hvordan man får indkredset mekanismer i procesorienterede effektevalueringer, så man kan teste disse og give troværdig viden om, hvorvidt indsatsen bidrager til resultater.

## HVAD ER MEKANISMER I PROCESORIENTEREDE EFFEKTEVALUERINGER?

Der er identificeret i alt 24 forskellige definitioner af mekanismer, men jeg vil i det nedenstående afgrænse det til de tre forståelser, der går igen i litteraturen (Astbury og Leeuw 2010, s. 367, Mahoney 2003, s. 3-6).

Den første forståelse ser indsatser/programmer/projekter som mekanismer, der medfører resultater:

*Figur 1 - Mekanismen kan betragtes som den uafhængige variabel, der påvirker den afhængige variabel*



Den uafhængige variabel X kan være en offentlig indsats (politik, program, projekt m.v.), som påvirker den afhængige variabel Y: problemløsningen eller outcome. I denne forståelse forårsager indsatsen X outcome Y (Pawson 2007, s. 3, Astbury og Leeuw 2010, s. 367, Bunge 1997, s. 427-428, 436, Gerring 2010, s. 1500).

Indsatsen X kan fx være et kommunalt aktiveringsprojekt, og resultatet Y kan være, at de aktiverede kommer i beskæftigelse. Det er altså selve aktiveringsprojektet, der medfører, at de aktiverede kommer i beskæftigelse.

Astbury og Leeuw problematiserer denne forståelse, da det ikke er projekter eller indsatser i sig selv, der fører til outcome. Derimod kan indsatser påvirke mekanismer i bestemte sammenhænge, som så kan føre til outcome:

*"A common mistake is for evaluators to conflate the term mechanism with program activity. As noted earlier, we are not alone in pushing for a distinction between underlying mechanisms of change and a to do list of program activities. Mechanisms appear too frequently as unexplained causal arrows that seem to flourish so well in the present climate of enthusiasm with visual logic models. This does not seem to be what theory-driven evaluators had in mind when they introduced the concept of mechanism to the evaluation community" (Astbury og Leeuw 2010, s. 367).*

Den engelske evalueringsteoretiker Ray Pawson skriver:

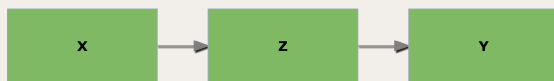
*"The most palpable and the most irritating usage of the shorthand is in evaluation research, in centring on the question, "does the programme work?" This is a causal arrow proposition – did the intervention bring*

about the intended outcome? But it is not "programmes" that work; that is merely a lazy linguistic expression" (Pawson 2007, s. 25).

Indsætter har ifølge Pawson ikke i sig selv den kausale kraft til at generere outcome, men de kan være med til at give ressourcer til mekanismer, som kan generere outcome.

Den anden forståelse ser mekanismer, som de variable, der medierer eller intervererer mellem indsats og outcome:

Figur 2 - Mekanismen kan betragtes som den medierende variabel

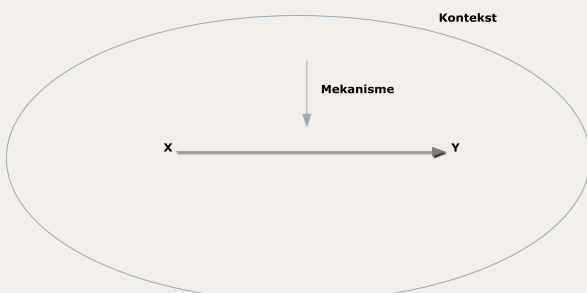


Den uafhængige variabel X påvirker den medierende variabel Z, som så igen påvirker den afhængige variabel Y. Det vil sige, at der er en variabel, der medierer mellem indsats og outcome. Det er denne forståelse, man opererer med, når man opstiller logiske modeller, hvor kortsigtede og mellemlange virkninger er medierende variable, der fører frem til outcome (Bunge 1997, s. 427-429, Mahoney 2001, s. 578, Gerring 2005, s. 179, 189, KREVI 2008).

I det førnævnte eksempel vil det kommunale aktiveringsprojekt, X, medføre, at de aktiverede søger arbejde, Z, som så igen medfører, at de kommer i beskæftigelse, Y.

Medierende variable er en del af en resultat kæde eller logisk model, hvor man kan spørge, hvorfor sammenhængen mellem variable fører til resultater. De medierende variable har underliggende forklaringer på, hvorfor de giver resultater. Hvorfor forholder målgruppen sig til nye informationer? Hvorfor ræsonnerer målgruppen? Hvorfor får målgruppen nye opfattelser? Hvorfor ændrer målgruppen adfærd? Disse spørgsmål leder til, at man skal identificere underliggende mekanismer, som påvirker forandringer i de medierende variable. Mekanismer drejer sig om hvorfor (Anderson m.fl. 2006, s. 103, Hedström og Swedberg 1998). De er tandhjulene (Weber 2006, s. 120, Pawson 2013, s. 6) til at forklare varians i bl.a. de medierende variable. Derfor er mekanismer ikke medierende variable, som sidstnævnte anvendes i statistiske analyser (Steel 2004, s. 57-59, Hedström 2009, Hedström 2005, s. 23, Mayntz 2004, s. 244, Hernes 1998, s. 88, Bunge 1997, s. 455). Dette er den tredje forståelse af mekanismer, som ser dem som underliggende forklaringer på, hvorfor indsatsen giver resultater (Mark m.fl. 2000, s. 154):

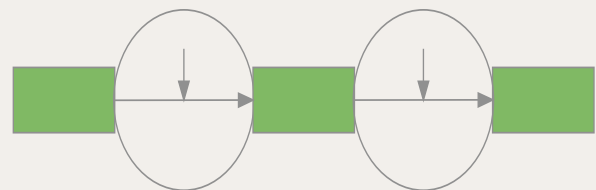
Figur 3 - Mekanismen kan betragtes som en ikke-observerbar enhed i en kontekst, der genererer resultater



Det er de engelske evalueringsteoretikere Ray Pawson og Nick Tilley, som står for ovennævnte velkendte model, som udtrykker: Hvad virker for hvem under hvilke omstændigheder? Resultatkæden eller den logiske model i form af den uafhængige variabel X og den afhængige variabel Y er de observerbare variable, den sociale regularitet, som bliver påvirket af den usynlige mekanisme og af kontekst.

Pawsons og Tilley's model i figur 3 er ikke umiddelbart selvforklarende. Men hvis man sammenligner modellen med de logiske modeller i figurene 1 og 2, er det pilene mellem variablene, der udgør kausaliteten. Modellen placeres mellem variablene på pilene, hvor mekanisme og kontekst synliggøres, som vist i nedenstående figur 4:

Figur 4 - Pawsons og Tilley's model i den logiske model



Figur 4 kan være med til at tydeliggøre, at det ifølge Pawson og Tilley ikke er variablene i sig selv, der udgør den kausale kraft, men at det er mekanismer i en kontekst, der udgør denne kraft.

Det tidligere nævnte kommunale aktiveringsprojekt og dets resultat skal ses i relation til kontekst og mekanisme. Mekanismen kan fx være "afskrækkelse", som forårsager, at aktiverede i matchgruppe 1 søger arbejde i en kontekst med gunstige økonomiske konjunkturer og kommer i beskæftigelse. Truslen om aktivering afskrækker målgruppen til at søge arbejde. Man kan antage, at den samme indsats for matchgruppe 2 i samme gunstige økonomiske konjunkturer ikke medfører, at de aktiverede kommer i beskæftigelse, fordi mekanismen "afskrækkelse" ikke påvirker denne målgruppe. Eller at matchgruppe 1 med samme indsats ikke kommer i beskæftigelse, fordi konjunkturerne ikke er gunstige. Så indsætter kan virke forskelligt for forskellige målgrupper i forskellige sammenhænge.

Mekanismen "afskrækkelse" i ovennævnte eksempel kan forklare, hvorfor de aktiverede søger arbejde, og den er ikke umiddelbart observerbar. Ray Pawson (2007) skriver:

*"...we often explain by going beneath the surface of objects to explain their "inner workings". We cannot understand the working of a clock by examining its face and hands, rather we need to know about clockworks (balanced springs or, nowadays oscillating caesium atoms)."*

Af andre mekanismer kan fx nævnes rygestop ved hjælp af nikotinplaster. Her er det rygerens "motivation", der er mekanismen, som medfører rygestop. Patienter bliver i et eksperiment behandlet med kalkpiller og oplever forbedring. Her er "placebo-effekten" mekanismen, som medfører, at patienterne får det bedre. I en skoleklasse får elever med en udsat social baggrund lavere karakterer end elever med en mindre udsat social baggrund. En forklaring på dette kan være, at de udsatte elever ikke er blevet stimuleret hjemmefra af deres forældre til skolearbejde. En anden forklaring kan være, at læreren på forhånd har større negative forventninger til de udsatte elever, og at eleverne på grund af dette præsterer dårligere. Mekanismen er her lærerens "forventninger" eller den "selvopfyldende profeti". Mekanismerne motivation, placebo-effekt og forventninger er ikke-observerbare, men er alligevel vigtige til at forklare mulige resultater.

Spørgsmålet er, om vi er kommet nærmere en forståelse af, hvad mekanismer er? Den amerikanske evalueringsteoretiker Carol Weiss giver en

god indfaldsvinkel til begrebet. Weiss ser mekanismer som menneskers viden, forståelse, tænkning, adfærd, tro, færdigheder, tillid, motivation, følelser, handling, værdier, læring m.v. – psykosociale faktorer (Weiss 1997, s. 73, se også Little 1991 og Little 2011 samt Elster 2007). Dermed taler vi også om faktorer, der ikke altid er observerbare, da de kan foregå i menneskers hoveder – konstruktioner. De menneskelige konstruktioner skal klarlægges for at kunne sige noget om kausalitet. Ray Pawson (2013) skriver i sin nye bog "The Science of Evaluation – A Realist Manifesto":

*"Mechanisms are embodied in the subjects' reasoning and they are best investigated therein (s. 20)".*

Det betyder, at mennesker får en central placering i de procesorienterede effektevalueringer. Det er mennesker og ikke programmer/indsatser, der får sociale programmer/indsatser til at virke:

*"...people do not react to programs like billiard balls that are hit; rather programs only work if people choose to make them work and are placed in the right conditions to enable them to do so" (Astbury og Leeuw 2010, s. 370).*

Sociale indsatser indgår dog som en vigtig faktor for, at den menneskelige vilje træder i kraft, men det er ikke indsatserne i sig selv, der er triggeren for outcome: Mekanismen er triggeren (Pawson og Tilley 1997, s. 35-40) Det er mennesker eller aktører, som har kausal kraft i form af, de agerer i bestemte situationer, og indsatser kan være med til at stimulere aktørers ageren (Hedström 2005, s. 26):

*"A person is a causal unit because it thinks and feels and the thoughts and feelings can dominate a line of action" (Stinchcombe 1991, s. 372).*

Det betyder dog ikke, at der er tale om fri menneskelig vilje eller valg. Vi taler ikke om rendyrket mikroadfærd eller rendyrket metodisk individualisme (Little 1996, s. 35-36, Little 2011, s. 274, 278-280). Vores handlingsmønstre indgår i en kompleks sammenhæng mellem strukturinstitutioner-aktører, hvilket vil sige, at vi handler inden for nogle rammer, som er med til at bestemme vores handlingsmønstre:

*"People make choices but don't choose the choices open to them" (Pawson 2007, s. 38).*

Den realistiske forståelse eller evalueringsretning ser mekanismer som forklaringsenheder, der kan give forklaringer på problemer og løsninger på tværs af indsatser: det sociale område, beskæftigelsesområdet, børn og unge, sundhed, uddannelse m.v.

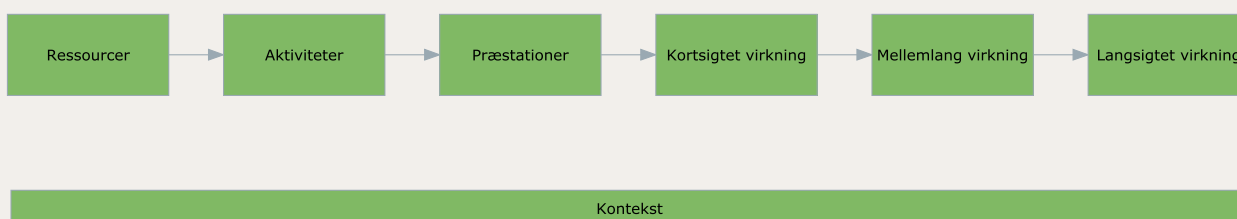
På tværs af forskellige indsatsområder vil der være mekanismer, som går igen, og dermed kan man finde samme forklaringstyper på tværs af områderne (Pawson 2013, s. 127-137, Little 2011, s. 284).

## HVORDAN FÅR MAN INDKREDSET MEKANISMER I PROCESORIENTEREDE EFFEKTEVALUERINGER, SÅ MAN KAN TESTE DISSE OG GIVE TROVÆRDIG VIDEN OM, HVORVIDT INDSATSEN BIDRAGER TIL RESULTATER?

I procesorienteret effektevaluering er det centralt, at der bliver opstillet en programteori for indsatsen, som så efterfølgende bliver testet for, om dens antagelser om kausalitet holder. Mekanismer indgår i denne programteori eller mere klargørende på dansk "indsatsteori". Ideelt set vil denne indsatssteori være opstillet af programmagerne i selve designet af indsatsen. Det kan dog være vanskeligt for programmagerne at synliggøre ofte tavs viden om, hvorfor indsatsen giver resultater, og det er langt fra altid, at indsatsen kan understøttes af forskningsbaseret viden, men i højere grad må trække på mere erfaringsbaseret viden.

Derfor skal evaluatoren "rekonstruere" programmagerens indsatssteori. Denne rekonstruktion kan tage udgangspunkt i de eksisterende beskrivelser af indsatsen: projektbeskrivelser, problemformuleringer, lovmateriale m.v. Beskrivelsen kan sammenfattes i en logisk model, som vist i nedenstående (se KREVI 2008 for vejledning til at udarbejde den logiske model):

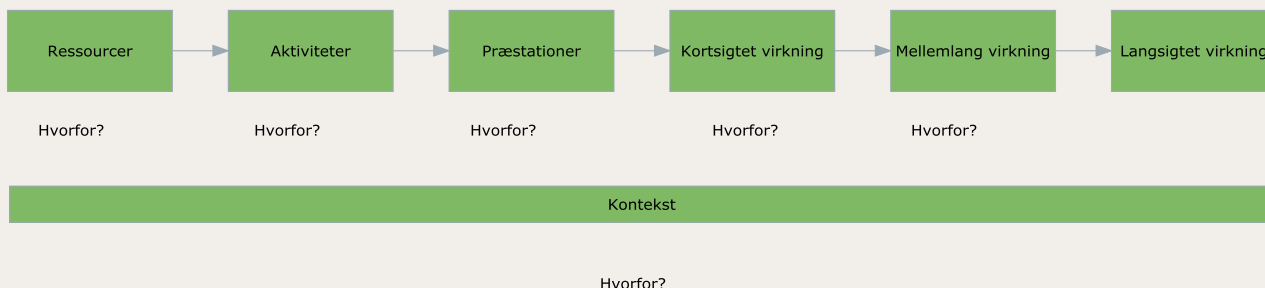
Figur 5 - Logisk model for en indsats



Ressourcer, aktiviteter og præstationer udgør indsatsen i modellen – den uafhængige variabel. De kortsigtede og mellemlange virkninger udgør de medierende variable, og den langsigtede virkning udgør den afhængige variabel. Pilene mellem variable udgør den sociale regularitet – den horisontale pil i figur 3. Kontekst er de politiske, økonomiske, geografiske, sociale, demografiske m.v. forhold, der kan påvirke den sociale regularitet – cirklen i figur 3.

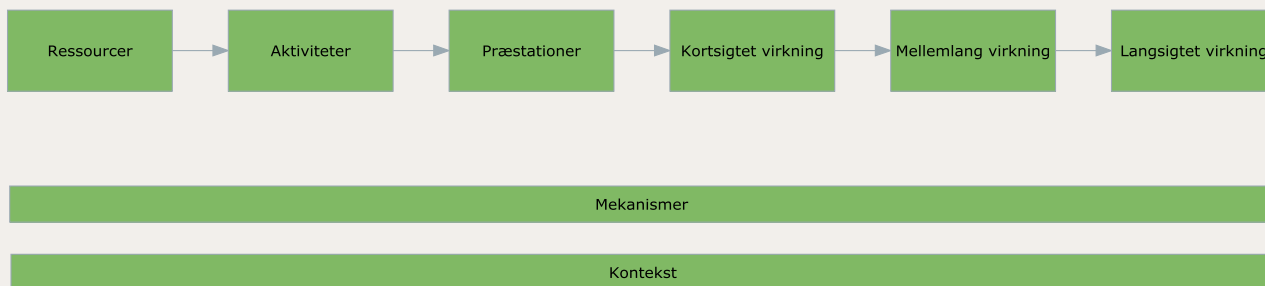
Den logiske model i figur 5 kan være med til at give et overblik over indsatsen, og den kan være med til at indkredse indsatsens mekanismer, da den kan være udgangspunktet for de spørgsmål, man vil stille programmagerne, som vist i figur 6. Da mekanismer kan være usynlige og have fokus på menneskelig tænkning, skal man bruge metoder, der har adgang til menneskelige konstruktioner. Dertil kan fx interviews eller surveys være en velegnet metode.

Figur 6 – Logisk model som interviewguide



Figur 6 kan fungere som interviewguide, hvor man spørger programmagerne, hvad der fx er baggrunden for at vælge medarbejdere med en bestemt uddannelse til indsatsen, hvorfor man har valgt bestemte metoder, hvorfor bestemte ydelser eller præstationer fører til resultater, hvorfor de kortsigtede og mellemlange virkninger fører til problemløsning, hvorfor påvirker den beskrevne kontekst indsats og resultater? Denne øvelse skal sikre, at evaluator får de ofte udtalte antagelser eller mekanismer synliggjort, så de kan medtages i indsats teorien, som vist i figur 7:

Figur 7 – Indsatsteori for indsatsen



Figur 7 udgør den samlede indsats teori med indsatsens komponenter, dens mekanismer og den kontekst, der kan tænkes at påvirke indsatsen og dens resultater. Evaluator vil derefter undersøge, om indsatsen er blevet implementeret som planlagt og vil teste, om dens mekanismer eller antagelser er holdbare i den virkelige verden, hvor indsatsen er blevet implementeret. Den procesorienterede effektevaluering skal svare på, hvorvidt den offentlige indsats virker for hvem og under hvilke omstændigheder. I denne evalueringsform er det ikke nok at undersøge, om indsatsens planlagte aktiviteter er blevet gennemført som planlagt, og om der er opnået resultater i indsatsen. Man skal også undersøge, hvorfor resultaterne er opnået. Sidstnævnte gøres ved at teste mekanismerne empirisk.

## REFERENCER

Anderson, Peter J.J. m.fl. (2006), "Understanding Mechanisms in Organizational Research: Reflections from a Collective Journey", i: Journal of Management Inquiry, 15, s. 102-113.

Astbury, Brad og Frans L. Leeuw (2010), Unpacking black boxes: Mechanisms and Theory Building in Evaluation", i: American Journal of Evaluation, 31 (3), s. 363-381.

Bunge, Mario (1997), "Mechanism and Explanation", i: Philosophy of the Social Sciences, 27, 4, s. 410-465.

Elster, Jon (2007): Explaining Social Behavior – more nuts and bolts for the social sciences. Cambridge University Press. Kapitlerne 1 og 2.

Gerring, John (2005), "Causation: A Unified Framework for the Social Sciences", i: Journal of Theoretical Politics, 17 (2), s. 163-198.

Gerring, John (2007), "Review Article: the Mechanistic Worldview: Thinking Inside the Box", i: British Journal of Political Science, 38, s. 161-179.

Gerring, John (2010), "Causal Mechanisms: Yes, but..." , i: Comparative Political Studies, 43 (11), s. 1499-1526.

Hedström, Peter og Richard Swedberg (1998), "Social mechanisms: An introductory essay", i: Hedström og Swedberg (ed.): Social Mechanisms - an Analytical Approach to Social Theory. Cambridge University Press.

Hedström, Peter (2005): Dissecting the Social – On the Principles of Analytical Sociology. Cambridge University Press.

Hedström, Peter (2009), "Studying mechanisms to strengthen causal inferences in quantitative research", i: Box-Steffensmeier, Brady og Collier (red.): The Oxford Handbook of Political Methodology. Oxford University Press.

Hernes, Gudmund (1998), "Real virtuality", i: Hedström og Swedberg (ed.): Social Mechanisms - an Analytical Approach to Social Theory. Cambridge University Press.

KREVI 2008. Den logiske model – et værktøj til at planlægge, gennemføre og evaluere sociale indsatser. Notat.

Little, Daniel (1991): Varieties of Social Explanation – An Introduction to the Philosophy of Social Science. Westview Press, s. 1-38.

Little, Daniel (1996), "Causal Explanation in the Social Sciences", i: Southern Journal of Philosophy, 34: Supplement, s. 31-56.

Little, Daniel (2011), "Causal mechanisms in the social realm", i: Williamson m.fl. (red.): Causality in the Sciences. Oxford University Press. S. 273-295.

Little, Daniel (2012), "Social Mechanisms and Scientific Realism Discussion of "Mechanistic Explanation in Social Contexts" by Johannes Persson", i: Social Epistemology Review and Reply Collective, 1 (3), s. 1-5.

Mahoney, James (2001), "Beyond Correlational Analysis: Recent Innovations in Theory and Method", i: Sociological Forum, 16, 3, s. 575-593.

Mahoney, James (2003): Tentative Answers to Questions about Causal Mechanisms. Paper præsenteret på det årlige møde i American Political Science Association, Philadelphia, 28. august.

Mark, Melvin M., Gary T. Henry og George Julnes (2000): Evaluation. Jossey-Bass.

Mark, Melvin M., Stewart I. Donaldson og Bernadette Campbell (2011), "The Past, the Present, and Possible Futures of Social Psychology and Evaluation", i: Mark, Donaldson og Campbell (ed.): Social Psychology and Evaluation. The Guildford Press.

Mayntz, Renate (2004), "Mechanisms in the analysis of Social Macro-Phenomena", i: Philosophy of the Social Sciences, 34, s. 237-259.

Pawson, Ray og Nick Tilley (1997): Realistic Evaluation. Sage Publications.

Pawson, Ray (2007): Causality for Beginners. Upubliceret manuskript.

Pawson, Ray (2013): The Science of Evaluation – A Realist Manifesto. Sage.

Steel, Daniel (2004), "Social Mechanisms and Causal Inference", i: Philosophy of the Social Sciences, 34, 1, s. 55-78.

Stinchcombe, Arthur L. (1991), "The Conditions of Fruitfulness of Theorizing About Mechanisms in Social Science", i: Philosophy of the Social Sciences, 21, 3, s. 367-388.

Weber, Klaus (2006), "From Nuts and Bolts to Toolkits – Theorizing With Mechanisms", i: Journal of Management Inquiry, 15, 2, s. 119-123.

Weiss, Carol H. (1997), "Theory-based evaluation: past, present and future", i: New Directions for Evaluation, nr. 76. \_\_\_

# HVOR REALISTISK ER REALISTISK SYNTSE? REFLEKSIONER FRA EVALUERINGSPRAKSIS<sup>iii</sup>

Af Christina Stentoft Hoxer<sup>2</sup>, Steffen Bohni Nielsen<sup>3</sup>

## INTRODUKTION

I den seneste snes år har der været en international tendens i offentlig forvaltning, hvor der har været et øget fokus på at anvende eksisterende forskning i politikformulering. Denne tendens har bl.a. fundet en accentuering som evidensbaseret politik og praksis (EBPP), hvor forvaltere og politikere, især i de angelsaksiske lande, argumenterede for, at politik i øget grad skulle baseres på viden om "hvad der virker". Denne tendens fandtes på tværs af policy områder og blev ofte kombineret med resultatbaseret styring (Bhatti, Foss Hansen & Rieper, 2006),

Én konsekvens var en stigende efterspørgsel af meta-evalueringer, der opsummerer, hvad evalueringer har vist om effektive interventioner. Produktionen af sådanne meta-evalueringer<sup>iv</sup> har haft to særlige kendetegn: For det første en institutionalisering af såkaldte "clearinghouses" og for det andet etableringen af metodologiske standarder for, hvad der accepteres som gyldig viden (evidens).

I forhold til institutionalisering er de klareste eksempler på denne tendens herhjemme etableringen af clearinghouses på sundheds-, social- og uddannelsesområdet i form af Cochrane Centret, Nordisk Campbell Center (Nu SFI Campbell) og Clearinghouse for Uddannelsesforskning på DPU, Aarhus Universitet.

En anden konsekvens er en øget metodologisk systematisering af, hvordan meta-evalueringer bør gennemføres. Inden for evidensbevægelsen har meta-evalueringer således fået en bestemt form: Som systematiske forskningsoversigter (UK: systematic reviews). Kendetegnende for denne tilgang er (i) en systematisk afsøgning af evidens om effektive indsatser, som ikke alene stiller krav til meta-evalueringens metodologi, men (ii) også stiller krav om, at gyldig viden skal være frembragt pba. velgennemførte eksperimentelle designs (randomiserede kontrollerede forsøg, RCTer). Det betyder med andre ord, at viden genereret vha. andre evalueringdesigns ekskluderes. (iii) Endelig er et karakteristika ved systematiske reviews den form for statistisk meta-analyse som gennemføres, hvor fundene fra enkelte RCTs syntetiseres til

et enkelt kvantitativt resultat (med større n, og mindre konfidensinterval).

Der har været skrevet mangt og meget herom, oftest med "evidensdebatten" som en semantisk magnet, både herhjemme og internationalt (fx Bhatti, Foss Hansen & Rieper 2006; Krogstrup 2011; Rieper & Foss Hansen, 2007).

Vi vil her koncentrere om ét af de alternativer til meta-analyse som blev foreslået af den engelske evalueringforsker, Ray Pawson. For et årti siden publicerede han to artikler, hvor han på den ene side, gik i rette med den metodologi som kendes i systematiske reviews (2002a) og efterfølgende pegede på realistisk syntese (RS), som et modsvar til meta-analyser i systematiske reviews (Pawson, 2002b).

Hans kritik er (i) videnskabsteoretisk, (ii) metodologisk og (iii) ud fra et anvendelsessigte.

For det første anfægter han det videnskabsteoretiske fundament, positivismen, som systematiske reviews beror på og som ser kausalitet som lovmæssigheder.

For det andet anfører han, at med et ensidigt fokus på RCTer, og derved eksklusion af andre forskningsdesigns, risikerer forskere og beslutningstagere at "smide guld på gaden" og negligerer samtidig muligheden for, at der kan være elementer i en ineffektiv indsats, der rent faktisk har virket.

For det tredje kan sådanne guldklumper, ifølge Pawson findes, hvis man tager udgangspunktet i indsatsers programteori som det analytiske genstandsfelt. Denne viden vil have stor relevans for såvel beslutningstagere som for en systematisk akkumulering af viden.

Til trods for Pawsons bidrag har fået "ny-klassisk" status i evalueringforskningen har der været forbausende få eksempler på praktisk anvendelse af RS indenfor evaluering. Det er først de sidste par år, at flere artikler med denne tilgang er blevet offentliggjort (Fx Kane, Gerretsen, Scherpbier, Dal Poz, & Dieleman, 2010; Rycroft-Malone et al, 2012; Wong, Pawson & Owen, 2011).

Formålet med denne artikel er at præsentere vores afprøvning af metoden RS i praksis. Der er gjort ifm. med gennemførelsen af en meta-evaluering af kommunale projekter vedrørende børns overvægt (Sundhedsstyrelsen, 2012). På denne baggrund vil vi vurdere RS' styrker og svagheder.

I det følgende præsenteres RS' videnskabsteoretiske grundlag og den metodologiske tilgang som Ray Pawson igennem flere artikler har uddybet.

Herefter gives en kort beskrivelse af den praktiske fremgangsmåde med brug af RS i form af en specifik meta-evaluering, hvor tilgangen blev anvendt. Slutteligt diskuteres RS' styrker og svagheder pba. dens praktiske applicering.

<sup>2</sup> Rambøll Management Consulting (pt. på orlov og studerer MSc. International Health Policy på London School of Economics and Political Science)

<sup>3</sup> Socialstyrelsen. Steffen var ansat i Rambøll Management Consulting da evalueringen, som artiklen omhandler, blev gennemført.



### REALISTISK SYNTSE

Som nævnt ovenfor er RS udviklet af den britiske evalueringsforsker, Ray Pawson, som en alternativ tilgang til meta-analyse (Pawson, 2002b). Realistisk syntese, bygger videre på en række præmisser, der kan genfindes indenfor teori-baseret evaluering i almindelighed og i hans klassiske værk "Realistic Evaluation" (Pawson & Tilley, 1997) i særdeleshed.

Centralt i denne tilgang ligger en fundamentalt anderledes ontologisk og videnskabsteoretisk forståelse, som er forankret i realismen. Der betyder bl.a. at Pawsons forståelse af kausalitet er en anden end i positivismen.

I modsætning til positivismen anser Pawson det ikke muligt, at en bestemt indsats kan *forårsage* bestemte udfald (outcomes). Pawson hævder, at *aktiviteterne* i en indsats kan igangsætte (UK: trigger) en bestemt mekanisme, der, under givne betingelser, afføder en reaktion hos en eller flere aktører i en bestemt kontekst, som bevirker et udfald. Der er tale om en generativ opfattelse af kausalitet.

Der er således en tre nøglebegreber i denne tilgang kontekst (UK: context), mekanismer (UK: mechanisms) og udfald (UK: outcomes). Alle er irreducerbare entiteter i studiet af social handlen. Dette betyder, at alle dele skal medtages i studiet heraf. Pawson taler om således om Context-Mechanism-Outcome (CMO) konfigurationer (ibid).

Denne type konfigurationer anerkender mere komplekse og systemiske sammenhænge mellem input og udfald (O). For at udlede et kausalt udfald (O) mellem to hændelser (X og Y) er man nødt til at forstå den underliggende generative mekanisme (M), der forbinder dem og konteksten (C), hvor forholdet forekommer (Pawson, Greenhalgh, Harvey & Walshe 2004).

Realismen anerkender den kompleksitet indsatsen forekommer i, og at succes af en indsats ikke blot et spørgsmål om, hvorvidt de underliggende teorier er korrekte (er indsatsen effektiv eller ej), men derimod også, at det afhænger af de involverede personer, interpersonelle relationer, institutioner og infrastrukturer som indsatsen leveres i.

RS handler således om at konstruere et billede af, hvordan forskellige kombinationer af sådanne sammenhænge og omstændigheder kan styrke eller hæmme virkningen af en indsats "programteori" (Pawson et al., 2004). Ifølge RS er det således program-mekanismer som frembringer forandringer frem for programmer i sig selv.

For Pawson, er det ved at udlede en indsats' programteori (og ikke blot implementeringsteorien i form af en logisk model), at studiet af CMO konfigurationer kan lade sig gøre. En bestemt indsats kan bestå af flere CMO konfigurationer. Én konfiguration kan fx omhandle rekruttering af overvægtige til at indgå i en indsats. En anden kan omhandle selve behandlingsdelen og en tredje vedrørende efterværn og opfølgning.

Heri ligger også Pawsons kritik af meta-evalueringer i form af systematiske reviews, der ekskluderer evalueringer udelukkende pba. deres evalueringdesign. Dette er ensbetydende med, at elementer af en programteori, der har været effektiv ligeledes ekskluderes. Herudover har Pawson en yderst hård kritik af statistisk meta-analyse, og den gyldighed, der kan tillægges disse resultater.

Pawsons forestilling er, at der kan opbygges en kumulativ viden om, hvordan mekanismer virker under forskellige betingelser, og hvilke udfald de kan bevirke (2002b; 2003)

En række af forskelle som kan findes med systematiske reviews udformet i en positivistisk tilgang og realistisk syntese er opsummeret i tabellen nedenfor.

TABEL 1: SAMMENLIGNING AF REALISME- OG POSITIVISME-BASEREDE REVIEWS

Variable	Realist review	Positivistisk review
<b>Ontologi</b>	Sondrer mellem underliggende mekanismer, handlinger og udfald	Flad, begrænsede teoretiske antagelser
<b>Kausalitet</b>	Generativ  Ser effekten af en given intervention som subjektiv gennem virkningerne af diverse kontekstuelle mekanismer men sammenhængende med objektive programkarakteristika	Successionistisk  Forudsætter at kontekstuelle faktorer kan kontrolleres og giver ingen forklaring på variation i effektivitet indenfor og imellem programmer. Ingen anerkendelse af at konteksten spiller ind
<b>Antagelse</b>	Programmets aktiviteter, "trigger" mekanismer. Fokus på variation i udfald	Programmet forårsager givne udfald. Fokus på hårde udfald
<b>Aktør- Struktur</b>	Struktur begrænsninger, aktører kan initiere begivenheder, men er altid "embedded"	Begrænset fokus på aktør. Programmet skaber udfald. Kontekstfaktorer kan der kontrolleres for
<b>Styrker</b>	Genererer viden om "hvad der virker, for hvem, under hvilke omstændigheder"	Generere viden om "hvad der virker, og hvad der ikke virker"
<b>Svagheder</b>	Hvad er mekanismer egentlig?  Kontekst kan variere ad infinitum	Kan ikke forklare variationer i effekt
<b>Kilde: Tilpasset fra Clark, MacIntyre &amp; Cruickshank, 2007.</b>		

## REALISTISK SYNTSE I PRAKSIS

I dette afsnit vil vi beskrive, hvordan realistisk syntese er blevet anvendt i praksis. Vi beskriver indledningsvist den case, hvor RS fandt anvendelse.

I 2012 færdiggjorde Rambøll Management Consulting den første evaluering herhjemme, som har anvendt realistisk syntese.

Der var tale om en indsats udmøntet af Sundhedsstyrelsen (SST) og finansieret af satspuljemidler. Som en del af satspuljeforliget for 2005-2008 blev der afsat 73 mio. kr. til kommunale indsatser med henblik på at forebygge og behandle overvægt blandt børn og unge.

Udmøntningen af midlerne skete i tre tempi. I alt gennemførtes 30 kommunale projekter, hvoraf 20 havde to spor- en generelt forebyggende og en specifik indsats - som skulle evalueres. De øvrige 10 kommunale indsatser havde udelukkende en specifik indsats. Samlet set blev 42 forskellige indsatser afprøvet.

Indsatsen havde til formål at afprøve og udvikle flere forskellige metoder og måder at organisere sig på i forhold til forebyggelse af overvægt blandt børn og unge i en kommunal kontekst

SST havde i puljeudmøntningen lagt stor vægt på, at de kommunale indsatser var udviklingsprojekter, som havde fået høj grad af metodefrihed. SST havde primært stillet krav til at designet skulle være flerstrengt og således fokusere på både psykosociale aspekter ved overvægten samt fysisk aktivitet og kost.

I forhold til gennemførelse af evaluering stillede SST krav til de deltagende kommuner om at udfylde baseline-, midtvejs- og afsluttende evalueringsskabeloner (som var konstrueret før forfatterens involvering) samt, at kommunerne lokalt skulle gennemføre eksterne evalueringer for hvert projekt. Der blev ikke stillet krav om, at specifikke evalueringsspørgsmål skulle besvares. Ligeledes blev der ikke stillet krav til evalueringdesign. Endelig var det valgfrit, hvilke indikatorer blev anvendt til at måle udfaldet af indsatserne. Meta-evalueringen af den samlede indsats blev igangsat, da projekterne fra første udmøntningsrunde var tæt på deres afslutning og var således ikke en integreret del af designet fra begyndelsen.

Meta-evalueringen havde fokus på følgende seks temaer:

- 1) Forankring og organisering
- 2) Metoder og tilrettelæggelse
- 3) Aktører og opgaver
- 4) Kompetenceudvikling
- 5) Rekruttering, fastholdelse og opfølgning
- 6) Mål og resultater.

Evalueringdesign udvælges i et samspil mellem evalueringens spørgsmål, tilsigtede anvendelse, mulige designs, indsatsens karakteristika og de tilgængelige ressourcer.

For at gennemføre meta-evalueringen valgte evaluator RS. Pawson har løbende udviklet og uddybet RS' metodologi (Pawson, 2006; Pawson et al., 2004; 2005). Her peger han på, at RS består af fem trin, som er illustreret i figuren nedenfor. Disse vil strukturere vores gennemgang af RS i praksis.

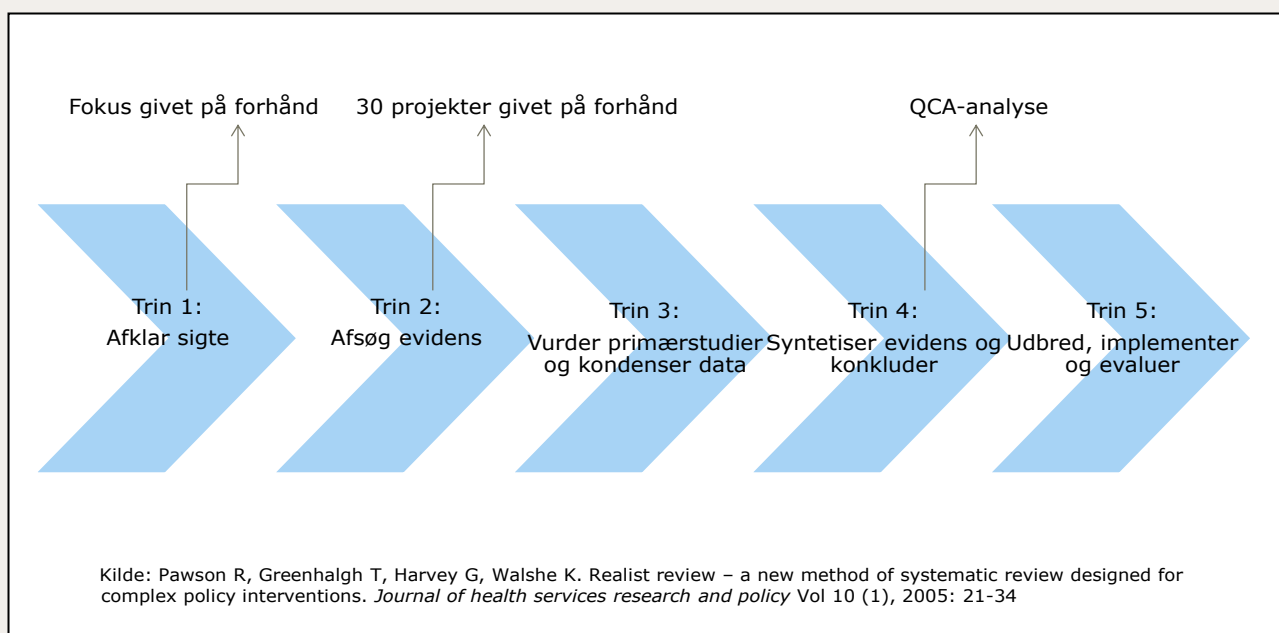
I vores anvendelse af realistisk syntese var opgaven snarere at afklare, hvilke evalueringsspørgsmål tillagdes særlig betydning af SST og samtidig forstå det empiriske genstandsfelt i puljen.

Idet fokus for evalueringen og det empiriske genstandsfelt allerede var defineret, afveg vi fra Pawsons metodiske forskrifter på dette trin. I stedet anvendte vi en tilgang til RS, som blev anvendt af et skotsk evalueringsteam (McCormack et al., 2006).

Her gennemførte forskerne indledningsvis en forskningsoversigt, hvor de søgte efter viden om effektive indsatser og sammenholdt disse med den eksisterende praksis, der var genstand for deres evaluering.

Formålet var at identificere kontekster, mekanismer og udfald (CMO konfigurationer) som positivt samvarierede i effektive indsatser og herefter identificere om tilsvarende CMO konfigurationer fandtes i de konkrete cases, som undersøgtes.

**FIGUR 1 – TRIN I REALISTISK SYNTSE**



## DE ENKELTE TRIN I REALISTISK SYNTSE

### Første trin

Ifølge Pawson m.fl. (2005) er første trin at afklare, hvilke undersøgelses-spørgsmål der ønskes svar på, samt at forstå den indsats, der skal undersøges, karakteristika, tilsigtede resultater og konteksten den implementeres i. Endelig skal de teoretiske antagelser der er underliggende indsatsens forskellige dele udledes. Dette kan gøres, ved at trække på eksisterende teori og empirisk forskning om genstandsfeltet eller indsats typer.

Tilsvarende afdækkede vi programteorien i 19 forskellige primærstudier af forskellige effektive indsatser overvægt blandt børn og unge, som var identificeret i to nylige forskningsoversigter (Gittelsohn, Kumar 2007; Summerbell et al. 2005).

I gennemgangen af den identificerede litteratur fokuserede vi *både* på, hvorvidt interventionerne havde vist sig effektive men også på, hvilke mekanismer, som var i spil i interventionen og i hvilken kontekst (operationaliseret som arenaer som skole, dagtilbud, hjem, fritidstilbud, mv.). På baggrund heraf lavede vi en "long list" af programteorier baseret på teorier som havde vist sig effektive med samme udfald og målgruppe (overvægtige børn og unge) andre steder.

Herudfra syntetiserede vi data fra programteoriene og identificerede seks "master programteorier" som de effektive indsatser var variationer af (med fokus på dimensionerne målgruppe, udfald, mekanismer, anvendt metode, arena etc.)

### **Andet trin**

I andet trin af review-processen gennemføres, ifølge Pawson, selve litteratursøgningen efter en eksplicit søgestrategi. Litteraturen fremskaffes, hvorefter litteraturen screenes for, hvorvidt den siger noget om den teori, der undersøges. Endelig vurderes forskningens kvalitet og i hvilket omfang konklusioner understøttes af data.

Det var især i forhold til denne fase, at den aktuelle evaluering adskilte sig fra RS' metodiske forskrifter, da data var afgrænset af det genstandsfelt, puljen til overvægt blandt børn og unge, der blev evalueret.

Således søgte evalueringsteamet i højere grad at skabe en "bro" mellem den eksisterende forskning og de faktiske indsatser. Dette blev gjort ved at sammenligne de kommunale indsatser programteori med de "master programteorier", som forskningsoversigterne havde peget på – og hvorvidt disse kunne grupperes inden for samme "klynge".

Derfor (re)konstruerede evalueringsteamet programteorien for hver af kommunale indsatser. Dette blev gjort på baggrund af på baggrund af skriftligt materiale fra projekterne (fx projektbeskrivelser, statusrapporter og evalueringsrapporter).

Endelig operationaliserede evalueringsteamet evalueringens genstandsfelt som et sæt af hypoteser vedrørende forskellige dele af indsatsen vedrørende fx (i) opsporing, rekruttering og visitation, (ii) metoder, og (iii) opfølgning.

### **Tredje trin**

Tredje trin i RS er at etablere "data extraction forms", hvorved relevante data fra studierne kondenseres og samles. Det er her at Pawsons forestilling om CMO konfigurationer især kommer til anvendelse, da relevante kontekstfaktorer ligeledes skal registreres i studiernes data extraction forms.

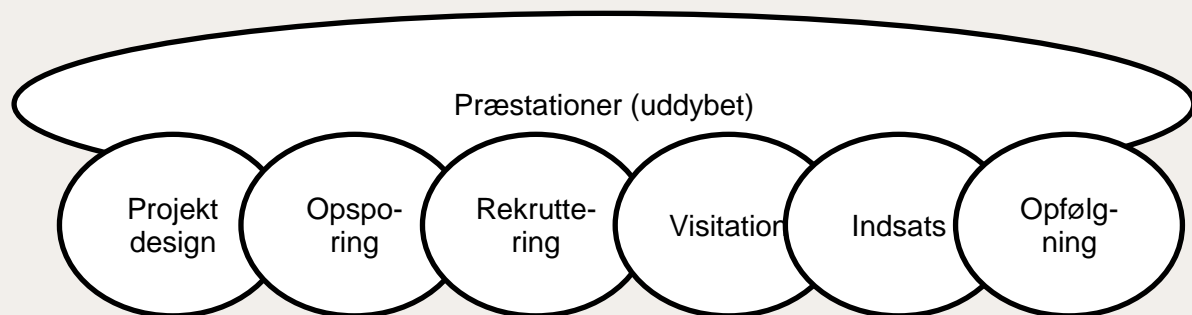
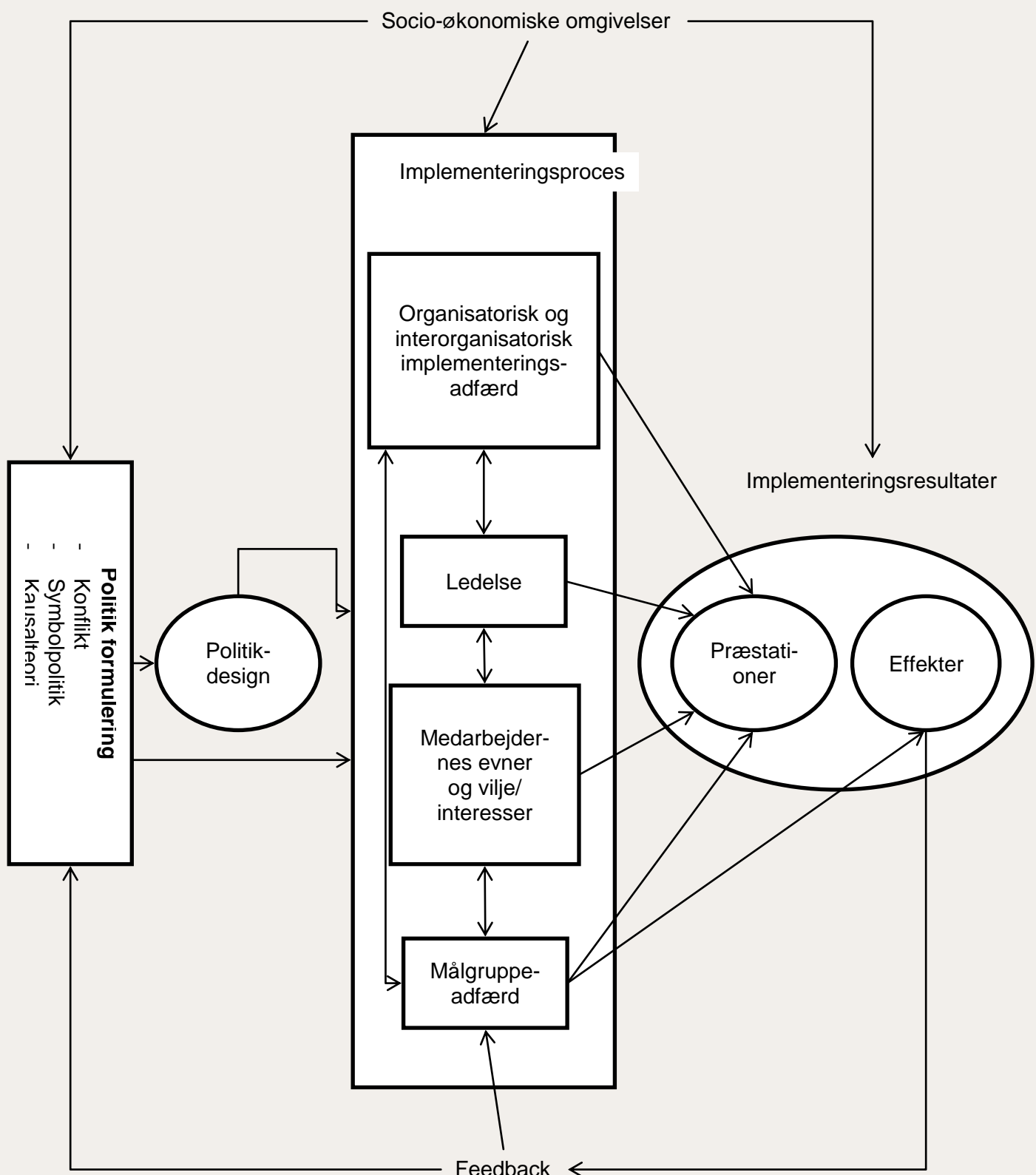
I evalueringen af puljen foretog vi en sammenligning mellem de faktiske indsatser og de, i forskningslitteraturen, bevist effektive indsatser. Her kunne vi konstatere, at de kommunale indsatser i forhold til deres design, underliggende antagelser og tilsigtede udfald havde tilsvarende CMO konfigurationer som fandtes i forskningen af effektive indsatser vedrørende overvægt blandt børn.

Denne sammenlignende analyse gjorde det således tydeligt, at de kommunale indsatser havde forholdsvis ens programteorier og fordelte sig inden for få af de identificerede master programteorier.

Vores overordnede analyse blev ændret fra at analysere på tværs af evidensen inden for forskellige masterteorier til i stedet at sætte fokus på CMO-konfigurationer og tendenser fra de danske projekter inden for få *specifikke master programteorier*, der havde samme CMO konfiguration.

Dette blev gjort ved at systematisere en model, som indsatserne faldt indenfor (se figur 1 nedenfor). Modellen er inspireret af Søren Winter og Vibeke Lehmann Andersens (2008), som peger på en række faktorer i en indsats' kontekst, der har betydning for udfaldet af dens implementering.

Fokus i meta-evalueringen var som omtalt også forhold i den kommunale implementeringsproces, herunder samarbejde og implementering mellem organisatoriske parter. Ligeledes peger modellen på, at der er forhold i selve den implementerende organisation som har betydning for udfaldet. Modellen blev anvendt til at identificere mulige konfigurationer af udfald og kontekst som kunne være "bærer" af mekanismer.



### *Fjerde trin*

Endelig er det ifølge Pawson en opgave at skabe en syntese af data, hvor evaluator søger efter at be- eller afkræfte de mønstre, som programteorien tilsiger, og derved raffinere program-teorien ved empirisk at teste den i forhold til kontekstfaktorer og udfald.

En udfordring i nærværende meta-evaluering var at antallet af cases krævede en systematisk analyse på tværs af relativt mange cases (42 specifikke indsats), men samtidig var for lille til avancerede statistiske analyser, der samtidig ville reducere kompleksiteten i de enkelte cases. Nyere evalueringsforskning har peget på en række ontologiske og videnskabsteoretiske ligheder mellem Qualitative Comparative Analysis (QCA) som analyseteknik og RS (Befani, Ledermann & Sager, 2007), ligesom QCAs praktiske anvendelighed er blevet demonstreret ifm. RS (Sager & Andereggen, 2011).

Essensen i QCA er, at kvalitative data fra de enkelte cases systematisk kodes som kvantitative værdier på en ordinal eller binær skala i forhold til en række variable, som forventes at have en relation til et (del-) resultat af indsatsen. QCA beregner så, hvilke kombinationer af kontekstbetingelser der korrelerer med udfaldet vha. boolsk algebra (Ragin, 1987; Ragin, 2008; Rihoux & Ragin, 2009).

Ca. 45 variable blev defineret og kodet for hver indsats. Disse blev efterfølgende "klynget" sammen i konfigurationer i forhold til indsatsmodellens faser, hvor de opstillede hypoteser om fx sammenhænge vedrørende opsporing kunne analyseres systematisk. Endvidere anvendes softwareprogrammet TOSMANA (et program til gennemførelse af data med få observationer ("small-n" analyser).

### *Femte trin*

Endelig skal evaluator, ifølge Pawson, drage konklusioner og afprøve disse samt anbefalinger med opdragsgiver og interessenter.

I nærværende evaluering blev RS blev afsluttet med afrapportering og præsentation af evalueringens resultater til evalueringens interessenter (Sundhedsstyrelsen og kommuner). Der var en løbende dialog med opdragsgiver om evalueringens metode, fortolkninger, foreløbige konklusioner og anbefalinger, mens der ikke processuelt var en høring af interessenter i fx kommuner, KL eller faglige organisationer.

## STYRKER OG SVAGHEDER

Ovenstående case tjener i denne sammenhæng som illustration for den praktiske gennemførelse af RS og danner grundlag for at vurdere RS' relative styrker og svagheder som en tilgang til meta-analyse. I det følgende diskuterer vi derfor disse i et teoretisk, metodologisk og anvendelsesorienteret perspektiv.

## TEORETISKE STYRKER OG SVAGHEDER

Den teoretiske styrke i RS ligger primært i dens fokus på CMO-konfigurationer og dermed anerkendelsen af, at kontekst og andre implementeringsforhold har betydning som en forklaringsramme. Realismen tilbyder således en alternativ, generativ, kausalitetsforståelse til en successionistisk og kontrafaktisk forståelse (Se Stern et al., 2012 for en uddybende diskussion).

Der er imidlertid også en del begrænsninger, som gør den teoretiske applicering problematisk. For det første er nøglebegrebet "mekanisme" uklart. Selvom bl.a. Pawson (2008) og Astbury & Leeuw (2010) ud fra en realistisk videnskabsteori har søgt at konkretisere "mekanismers" beskaffenhed, så er det ikke operationelt nok til praktisk anvendelse (Dybdal, Nielsen & Lemire, 2010; Lemire, Nielsen og Dybdal, 2012) og empirisk analyse på tværs af en række cases, som netop er RS' ambition (Pawson 2002b).

Endelig betyder denne mangel på præcisering af, hvad mekanismer "egentlig er", at det bliver såvel teoretisk som metodologisk problematisk.

For det første betyder det, at "mekanismer" bliver beskrevet så generelt og upræcist, at de passer til alting og ingenting – og derved ikke sub-stantierer noget som helst.

For det andet indebærer det, at aspirationen om at kunne analysere, hvordan mekanismer virker på tværs af policy områder og derved forskellige kontekster (Pawson, 2003), bliver operationelt besværligt. Det forekommer som en teoretisk abstraktion, som dårligt lader sig realisere i empirisk forskning.

Den samme knibe kommer man i, når man tager livtag med begrebet kontekst. Pawson m.fl. (2004) har defineret flere lag af kontekst, der kan være relevante i CMO konfigurationer i form af individuelle, interpersonelle, institutionelle og infrastrukturelle forhold. Imidlertid, kan disse blive så specifikke eller generelle, at den teoretiske (og empiriske) udsagnskraft reduceres til almindeligheder. Løsningen i nærværende evaluering var at afgrænse kontekstfaktorer i form af en række hypoteser, som blev undersøgt.

## METODOLOGISKE STYRKER OG SVAGHEDER

Der er flere metodiske styrker værd at fremhæve i anvendelsen af RS. Især forekommer der at være en øget systematik i tilgangen til at analysere betydningen af forskellige implementeringsfaktorer, når RS anvendes. Især kortlægningen af programteoriens bestanddele, fokus på mekanismer og deres interaktion med kontekstfaktorer fremmer denne forståelse. Dette giver mulighed for at identificere effektive dele af en indsats, som i sin helhed ikke har fungeret.

Samtidig er det en styrke ved RS, at der eksplicit bygges videre på et fundament af teori og empirisk forskning, som yderligere efterprøves og føjer til en akkumulering af viden på feltet. Endvidere bidrager anvendelsen af RS med et fokus på kontekstuelle forhold, hvilket udbygger og tilføjer yderligere indsigt ift. den eksisterende vidensbase.

De metodiske svagheder i RS er dog også tydelige.

Der mangler fortsat en operationel metodisk forskrift for, hvordan selve syntesen som en analytisk øvelse kan foretages. Der er som nævnt flere evaluatore (Befani et al., 2007; Sager & Andereggen, 2011), der har peget på QCA som en metode hertil, mens også mere pragmatiske metode- og datatrianglering har fundet anvendelse (Wong et al., 2011), mens andre mere systematisk har testet hypoteser (McCormack et al., 2006; Rycroft-Malone, Fontenla, Bigg, & Seers, 2010). Som nævnt anvendte vi i nærværende evaluering QCA.

Imidlertid, fandt vi, at QCA som metode lover mere end det kan holde. På den ene side er styrken ved QCA, at metoden kan identificere flere, *entydige* sammenhænge mellem faktorer, der kombineret har indflydelse på et givent resultat (dvs. hvilke kontekstfaktorer der skal være til stede for at rekruttering af overvægtige børn til en indsats lykkes).

Svagheden er imidlertid, at QCA har sværere ved at håndtere sammenhænge, der ikke er entydige, men hvor der alligevel er klare mønstre i data, fx at trejerdedele af indsatserne viser en sammenhæng mellem to faktorer og et bestemt udfald. Det er indlysende, at jo større antal observationer (cases) og jo flere variable (kontekstfaktorer, der tages højde for) jo større er sandsynligheden også for variationer mellem cases og derved, at der ikke kan findes entydige sammenhænge. Dette gjorde sig gældende i vores tilfælde.

Vi oplevede en meget "rodet" virkelighed på trods af, at kontekstuelle forhold blev holdt på et forholdsvis overordnet niveau (og således ikke synes at være forårsaget af et for højt detaljeringniveau). Vi fandt således meget få entydige sammenhænge. Derfor var vi i praksis nødt til at supplere QCA med uddybende, deskriptive analyser af mønstre i data.

QCA-analyse havde således i vores tilfælde væsentlige begrænsninger og spørgsmålet er om kodning af data i Nvivo eller andre programmer til systematisk analyse af kvalitative data, kunne have afhjulpet analysen lige så tilfredsstillende.

Det skal dog samtidig fremhæves, at vores forudsætninger for at gennemføre RS var begrænset af at være evaluatore på en ex-post evalueringssopgave, hvor dataindsamling og analyse var forestået af andre, genstandsfeltet for primærstudierne varierede og analysestrategien derfor mere beroede mere på "det muliges kunst" frem for en "køge-bogs" tilgang, hvor RS blev gennemført med metodologisk stringens.

Det er vores håb, at vi og andre fremadrettet får mulighed for at udrulle en fuld RS-metaevaluering, hvor analysestrategien fra start er tilpasset realistisk syntese.

## STYRKER OG SVAGHEDER VEDRØRENDE ANVENDELSE

RS har i vores øjne især en styrke i sin anvendelsesorientering. Det er en stor styrke, at RS ikke per definition forkaster indsatser *in toto*, hvis de ikke overordnet set er effektive i forhold til et bestemt udfald og med anvendelse af et bestemt evalueringdesign. I stedet giver RS mulighed for at fokusere på at finde sammenhænge / delteorier inden for en overordnet programteori, der har vist sig effektiv under givne forhold og således forklaringer på, hvornår *dele af* den givne indsats er effektiv eller ej. Dette sikrer ikke alene en større anvendelighed af evaluering, men også et mere nuanceret blik på de indsatser, der evalueres. Yderligere kommer implementeringsforhold også i fokus og bidrager med en viden som systematiske reviews om givne indsatser typisk ville have udgrænset.

Kort sagt har RS en særlig styrke set med et anvendelsessigte, fordi det anerkendes, at indsatser virker i komplekse systemer og derfor ikke kan reduceres til et spørgsmål om, "virker indsatsen" eller ej. Dette bidrager nemlig ikke til meget for de politiske beslutningstagere eller praktikere, som også har behov for at vide, under hvilke omstændigheder eller betingelser, en givne indsats er mere eller mindre tilbøjelige til at fungere og således sætte fokus på, hvad der fremadrettet kan gøres for at maksimere chancerne for succes (og minimere risikoen for fiasko). Dette kan også gøres ved at stykke "guldklumperne" sammen fra den indsigt som realistiske syntese skaber (Pawson et al. 2004).

Imidlertid kan fokus på såvel CMO konfigurationer også udlægges som en svaghed, fordi det formidler komplekse forklaringer. - Hvis beslutningstagere efterspørger simple løsninger på komplekse spørgsmål er RS i reel risiko for at blive marginaliseret som troværdig evidens.

Samlet set peger anvendelsen af RS på en lovende tilgang, som imidlertid stadig skal afprøves, dens grænser afsøges og metodologisk raffineres. I en virkelighed, hvor evaluatore oftest ikke har et bredt spektrum af relevante RCT'er at vælge mellem tilbyder realistisk syntese et pragmatisk og videnskabsteoretisk alternativ til gængse statistiske meta-analyser.

REFERENCER

- Astbury B and Leeuw F (2010) Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation* 31(3): 363-81.
- Befani, B., Ledermann, S., & Sager, F. (2007). Realistic evaluation and QCA. Conceptual parallels and an empirical application. *Evaluation*, vol 13(2):171-192.
- Bhatti, Y., Foss Hansen, H. & Rieper, O. (2006). *Evidensbevægelsens udvikling, organisering og arbejdsform. En kortlægningsrapport*. København: AKF Forlaget.
- Dybdal, L., Nielsen, S.B. & Lemire, S.T. (2010). Contribution Analysis Applied. Reflections on Scope and Methodology, *Canadian Journal of Program Evaluation*. Vol. 25 (2): 29-57.
- Gittelsohn J, Kumar MB (2007). Preventing childhood obesity and diabetes: is it time to move out of the school? *Pediatric Diabetes*. Dec.; 8 Suppl. 9:55-69.
- Kane, S.S., Gerretsen, B., Scherpbier, R., Dal Poz, M, & Dieleman, M. (2010). A realist synthesis of randomised control trials involving use of community health workers for delivering child health interventions in low and middle income countries. *BMC Health Services Research*, vol 10:286.
- Krogstrup, H.K. (2011). Kampen om evidens. - Resultatmåling, effekt-evaluering og evidens. København: Hans Reitzels Forlag.
- Lemire, S.T., Nielsen, S.B. & Dybdal, L. (2012). Making Contribution Analysis Work: A Practical Framework for Handling Influencing Factors and Alternative Explanations. *Evaluation. The international journal of theory, research and practice*. Vol 18 (2): 294-309.
- McCormack B, Wright B, Dewar J, Garbett R, Harvey G, Ballantine K (2006). *A Realist Synthesis of Evidence Relating to Practice Development*. NHS Scotland.
- Pawson, R. (2002a). Evidence-based Policy: In Search of a Method. *Evaluation*, vol. 8(2): 157-181.
- Pawson, R. (2002b). Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, vol. 8(3): 340-358.
- Pawson, R. (2003). Nothing as Practical as a Good Theory. *Evaluation*, vol. 9(4): 471-490.
- Pawson, R. (2008). Invisible Mechanisms. *Evaluation Journal of Australasia*. Vol. 8(2): 3-13.
- Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.
- Pawson R., Greenhalgh T., Harvey G., Walshe K (2004). *Realist synthesis: An introduction*. ESRC Research Methods Programme Working Paper Series.
- Pawson R., Greenhalgh T., Harvey G., Walshe K (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, vol. 10(1): 21-34.
- Ragin, Charles C. (1987) *The Comparative Method. Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Ragin, Charles C. (2008) *Redesigning Social Inquiry. Fuzzy sets and Beyond*. London, University of Chicago Press.
- Rihoux, Benoît & Ragin, Charles C. (2009). *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques*, London: SAGE Publications.
- Rieper, O. & Foss Hansen, H. (2007). *Metodedebatten om evidens*. København: AKF Forlaget.
- Rycroft-Malone, J., Fontenla, M., Bigg, D. & Seers, K., (2010). A realistic evaluation: the case of protocol-based care. *Implementation Science*, vol 5:38.
- Rycroft-Malone et al.(2010). Realist synthesis: illustrating the method for implementation research. *Implementation Science* vol 7: 33.
- Sackett, D., Rosenberg, W.M., Gray, J.A., Haynes, R.B., Richardson WS (1996). Evidence based medicine: what it is and what it isn't. *BMJ*, 312:71-2.
- Sager, F. & Anderreggen, C. (2011). Dealing with complex causality in realist synthesis: The promise of qualitative comparative analysis. *American Journal of Evaluation*, vol. 33(4):60.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B..(2012). Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development. *Department for International Development Working Paper 38*. DFID: London
- Summerbell CD, Waters E, Edmunds LD, Kelly S, Brown T, Campbell KJ. Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews 2005, Issue 3*. Art. No.: CD001871. DOI:10.1002/14651858.CD001871.pub2.
- Sundhedsstyrelsen (2010). *Implementeringsforskning om forebyggelse – en baggrundsrapport*. København: Sundhedsstyrelsen.



Sundhedsstyrelsen (2012). *Tværgående evaluering af projekterne i "kommunens plan mod overvægt blandt børn og unge"*. København: Sundhedsstyrelsen.

Winter, Søren C. & Nielsen, Vibeke Lehmann 2008: *Implementering af politik*. 1. udgave, 1. oplag, Academica, Århus.

Wong, G., Pawson, R. & Owen, L. (2011). Policy guidance on threats to legislative interventions in public health: a realist synthesis. *BMC Public Health*, vol 11:222.

---

<sup>i</sup> Teoribaseret evaluering, realistisk evaluering, virkningsevaluering, process tracing, contribution analysis.

<sup>ii</sup> Tak til mine kolleger i KORA: Leif Olsen fra afdelingen Evaluering og Innovation og Hanne Søndergård Pedersen fra afdelingen Styring og Ledelse, for værdifulde kommentarer og forslag. Ansvar for indholdet af artiklen er mit eget.

---

<sup>iii</sup> Artiklen er baseret på konference paper til Dansk Evalueringselskabs årskonference i Kolding, september 2012.

<sup>iv</sup> Vi definerer her meta-evaluering som en systematisk og formel evaluering af evalueringer ud fra en række kriterier om evalueringernes genstandsfelt, designs, fund og videnskabelige kvalitet. Således ses systematiske forskningsoversigter (systematic reviews) som én form for meta-evaluering. Vi definerer meta-analyse som en specific analysestrategi (kvantitativ eller kvalitativ) for at samle og syntetisere resultaterne fra flere undersøgelser (Se bl.a. Sackett, Rosenberg, Haynes & Richardson, 1996).