

VIRKNINGSEVALUERING I TEORI OG PRAKSIS – FÆLLES UDFORDRINGER OG NYE LØSNINGER?

RASMUS RAVN, PH.D.-STIPENDIAT
JULIA SALADO RASMUSSEN, PH.D.-STIPENDIAT
INSTITUT FOR STATSKUNDSKAB, AAU
SEMINAR 7, DES ÅRSKONFERENCE 2017



AALBORG UNIVERSITY
DENMARK

Program for læringsseminar 7

Torsdag d. 14. september

14.00-14.45	Virkningsevaluering kort fortalt og typiske faldgrupper
14.45-15.00	Pause
15.00-15.45	CMO-konfigurationer samt konstruktion og evaluering af programteori
15.45-16.30	Gruppeøvelser
16.30-17.00	Præsentation af gruppearbejde og opsamling



AALBORG UNIVERSITY
DENMARK

Realistisk evaluering kort fortalt

- Realistisk evaluering (virkningsevaluering) er en teoribaseret evalueringsform, der forsøger at forbinde proces og effekt gennem udarbejdelse og test af programteorier – ofte i form af CMO-konfigurationer
- Programteorier er eksplicite antagelser om, hvad der virker for hvem under hvilke omstændigheder

So, what is a programme theory?



It's just how and why you think the intervention will work.



RAMESESPROJECT.ORG



AALBORG UNIVERSITY
DENMARK

Realistisk evaluering kort fortalt

- Virkningsevaluering er velegnet til at undersøge **hvorfor og hvordan** en indsats virker
- Virkningsevaluering kan både være en **kritik og et supplement** til eksperimentelle og quasi-eksperimentelle evalueringsmetoder
- Der bør anvendes af **både** kvalitative og kvantitative metoder og data i realistisk evaluering.
- Undersøger "**generative mekanismer**". Mekanismer kan aktiveres i højere eller lavere grad. Mekanismerne bidrager med forklaringen på, hvorfor indsatser virker.
- Kontrolgrupper anvendes typisk ikke. I stedet bruges ofte **inter- og intra-programvariation** til at undersøge "hvad der virker, for hvem, under hvilke omstændigheder.



Lidt teorihistorie

- "Teori-baseret evaluering" er knap 50 år gammel (jf. Schumann, 1967)
- Opblomstring blandt evalueringsforskere i slut-1980erne og start-1990erne
- Realistisk evaluering udvikles i slut-1990erne (jf. Pawson & Tilley 1997)
- Bliver mainstream i 2000erne blandt evaluatore og i lærebøger
- Kommer til Danmark i begyndelsen af 2000-tallet og får betegnelsen "virkningsevaluering" (jf. Dahler-Larsen og Krogstrup, 2003)
- Tilbage til rødderne – kritisk realisme og teori-baseret evaluering frem for socialkonstruktivisme (Bredgaard m.fl., 2016)



Formålet med teori-baseret evaluering

*"At undersøge betingelserne for
programmets implementering og de
mekanismer, som forbinder processer og
effekter, som middel til at forstå, hvornår og
hvordan programmer virker"*

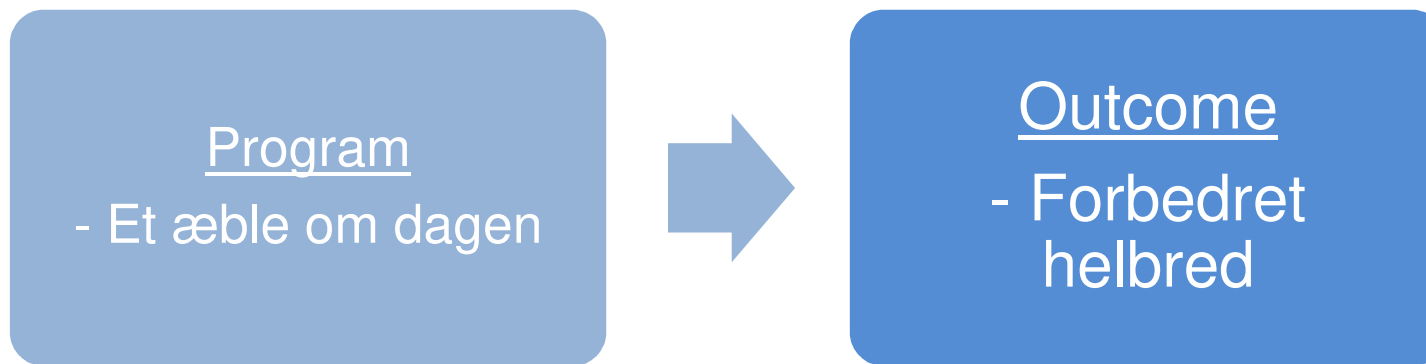
Kilde: C. Weiss (2007): "Theory-based Evaluation: Past, Present and Future, New
Directions for Evaluation, 114, pp. 68-81.

*Hvad virker for hvem, under hvilke
omstændigheder?*



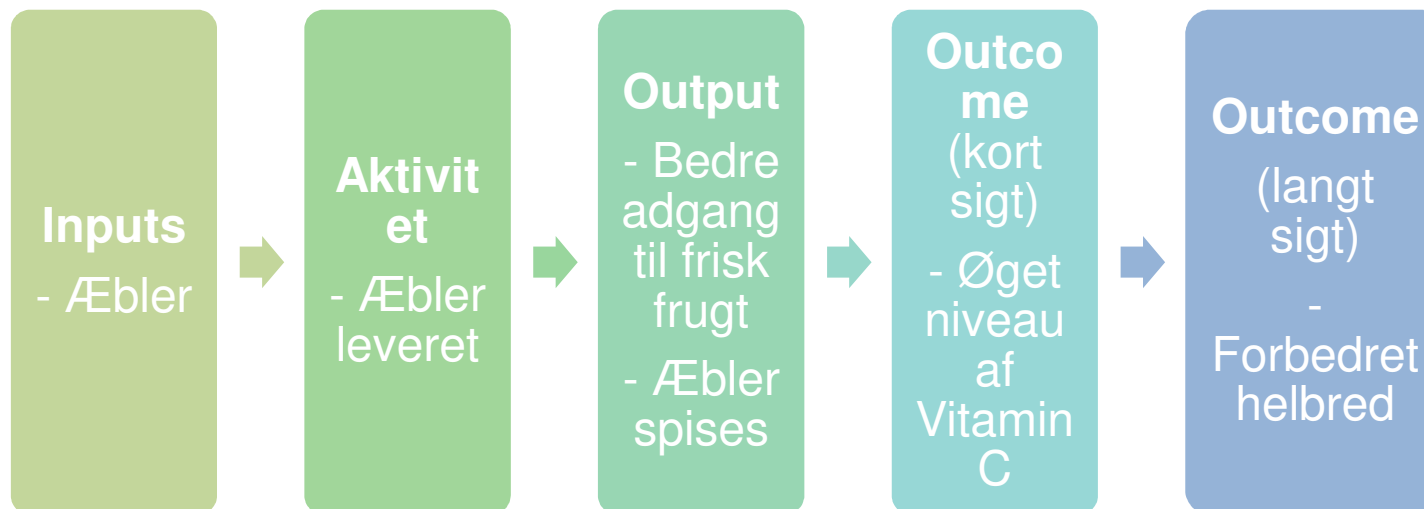
AALBORG UNIVERSITY
DENMARK

Evaluering uden programteori



Kilde: S.C. Funnell & P.J. Rogers (2011): *Purposeful Program Theory*, Josseybass.

Evaluering med programteori



Kilde: S.C. Funnell & P.J. Rogers (2011): *Purposeful Program Theory*, Josseybass.



AALBORG UNIVERSITY
DENMARK

Køreplan i virkningsevaluering

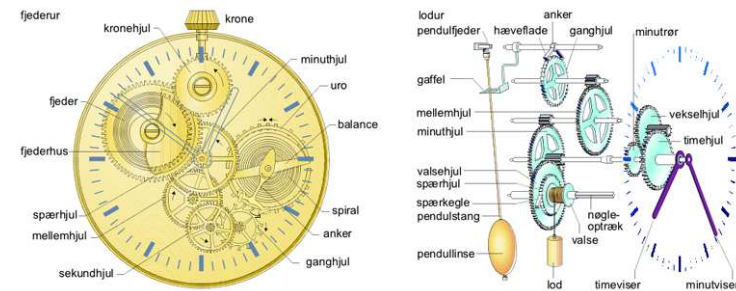
1. Stille evalueringsspørgsmålet
2. Finde kilder til programteorien
3. Opstille programteorien
4. Gøre programteorien klar til evaluering
5. Vælge metode og indsamle data
6. Analysere og konkludere

Kilde: P. Dahler-Larsen (2013): "Evaluering af projekter – og andre ting, som ikke er ting", Syddansk Universitetsforlag (s. 120)

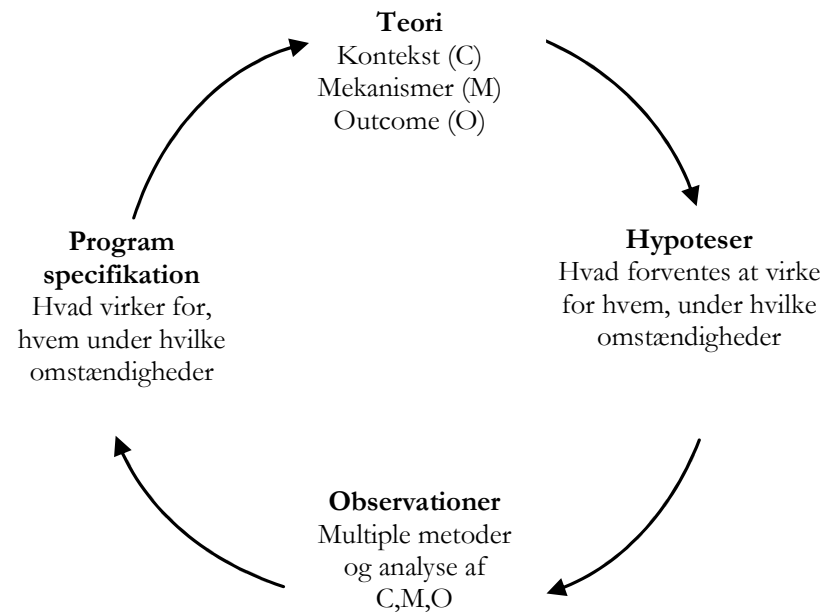
Nøglebegreber

Kontekst + Mekanismer = Effekter

- Mekanisme (virkningen)
 - De indre forhold som får programmer til at virke, når de virker
 - En hypotese (forklaring) om, hvorfor der (potentielt) er forbindelse mellem program og effekt
 - Mekanismer er ofte skjulte under overfladen og følsomme overfor variationer i konteksten.
- Kontekst (moderatorer)
 - De forhold som aktiverer mekanismer, under bestemte omstændigheder
 - En antændingsmekanisme, som regulerer et andet kausalforholds styrke og retning
- Effekter (outcomes)
 - Intenderede og uintenderede konsekvenser af programmer
 - Resultater, som programmer har en realistisk mulighed for at opnå



Den realistiske evalueringscyklus

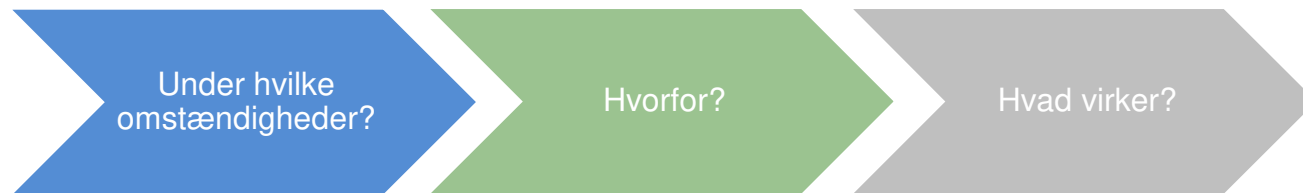


Kilde: R. Pawson & M. Tilley (1997: 85): *Realistic Evaluation*, London: Sage.

Realistisk evaluering (virkningsevaluering)

$$C + M = O$$

Context + Mechanism = Outcome



R. Pawson & N. Tilley (1997): Realistic Evaluation, Sage Publications.



AALBORG UNIVERSITY
DENMARK

Diskuter med sidemand/kvinde (3 min.)

- Hvordan kan I bruge/har brugt virkningsevaluering i jeres arbejdspraksis?
- Hvilke udfordringer kan I forestille jer/har I oplevet ved brug af virkningsevaluering?

Typiske faldgrupper



AALBORG UNIVERSITY
DENMARK

*"The years since the publication of *Realist Evaluation* (Pawson and Tilley 1997) and *Evidence-Based policy: A Realist Perspective* (Pawson 2006) have seen the publication of more than a **hundred papers** claiming to be realist evaluations and over a score of systematic reviews declaring themselves to be realist syntheses. A glaring question arises – are they **really** realist?" (Pawson 2013)*



AALBORG UNIVERSITY
DENMARK

Tre typiske fejl i brugen af "realistisk evaluering"

1. **Manglende forklaringer:** Statistiske

korrelationer og beskrivelser af aktiviteter men ikke forklaringer af programmets resultater.

2. **Anvendelse af én enkelt metode:** Arbejder ikke multimetodologisk. Konstruktion og testning af programteori bør ikke bygge på samme metode og data.

"(...) a classic realist headache – the CMO lottery. Under which column do the hypothesized constructs belong?" (Pawson 2013)

3. **Ingredienslister og kataloger:**

Programaktiviteter forveksles med mekanismer. Programteorierne bliver ikke testbare.

Kilde: Ray Pawson (2013): "The Science of Evaluation", Sage (kapitel 2).



AALBORG UNIVERSITY
DENMARK

Udfordringer i virkningsevaluering

1. Skal programteorier være "virkelighedsbeskrivelser" eller "evalueringsredskaber"?
2. Hvordan identificerer man "generative mekanismer"?
3. Hvordan afgrænses "kontekster" (moderatorer)?
4. Bliver evaluator blind for uforudsete bieffekter?
5. Hvad hvis der er flere aktører med forskellige programteorier på samme tid?
6. Hvad nu hvis programmet er et dynamisk udviklingsprojekt?
7. Er virkningsevaluering et alternativ eller supplement til effektevaluering?

Diskuter med sidemand/kvinde (3 min.)

- Hvordan kan man undersøge/teste mekanismer som evaluator?
 - Bør mekanismer testes?

De typiske faldgrupper (fortsat)

Resultater fra et litteraturstudie af publicerede realistiske evalueringer

Antal studier der udelukkende bruger kvalitative data i analysen	Antal studier der udelukkende bruger kvantitative data i analysen	Antal studier der både bruger kvalitative og kvantitative data i analysen	Antal studier der måler mekanismen kvantitativt	Antal studier der tester mekanismen ved brug af kvantitative metoder
25	2	15	5	3

Vurdering på baggrund af litteraturstudiet:

- Kvantitative og statistiske metoder bruges sjældent
- Mekanismer "måles" og testes sjældent kvantitativt. Ingen studier måler mekanismen kvantitativt og tester den ved brug af intra-program sammenligning baseret på mekanismens "styrke"
- Problem ved kvalitativ test af store-N programmer: *Generalisering af mekanismen til de resterende programdeltagere, der ligeledes har opnået indsatsens mål.*
- Der "overgeneraliseres" i de kvalitative studier.
- Jeg er sjældent overbevist om, at det er de hypotiserede mekanismer, der har bibragt resultaterne.



AALBORG UNIVERSITY
DENMARK

Eksempel: Virkningsevaluering af borger-sagsbehandlersamarbejdet på beskæftigelsesområdet

- Illustration af intra-programsammenligning

Den anvendte case og den teoretiske tilgang:

- Beskæftigelsesindsatsen for sygemeldte borgere og svagere kontanthjælpsmodtagere i Hjørring Kommune.
- Målet med indsatsen er, at borgerne skal opleve progression og gøres selvforsørgende.
- Spiller borgernes vurdering af samarbejdet med deres sagsbehandler og jobcenterindsatsen en rolle for opnåelse af progression? (og sidenhen selvforsørgelse)

Working alliance teori:

- Afsættet for teorien er, at et godt og konstruktivt samarbejde med borgerinddragelse og indflydelse fører til bedre resultater (mekanismen).

Alliancen består af tre forbundne elementer:

- Mål
- Aktiviteter
- Relationen

Metode

- Spørgeskemaundersøgelse blandt ledige og sygemeldte borgere i Hjørring Kommune indsamlet i august 2016.
 - N = 507, svarprocent 25,1
 - 8 spørgsmål der havde til formål at måle de tre elementer af alliancen. Fra de 8 spørgsmål konstrueres et indeks, der måler alliancen. Indekset er rekodet til "stærk", "middel" og "svag/negativ".
-
- Obs: Evalueringens kvalitative del medtages ikke i gennemgangen af eksemplet.

Test af mekanismen gennem intra-program sammenligning – eksempler på mulige analyser

Den binære sammenhæng mellem alliancens styrke og selv vurderet progression som følge af den samlede jobcenterindsats. Vist for sygemeldte borgere, kontanthjælpsmodtagere, uddannelseshjælpsmodtagere og forsikrede ledige.

Alliancens styrke	Lav/middel selv vurderet progression	Høj selv vurderet progression
Stærk	28,0 % (52)	72,0 % (134)
Middel	80,0 % (52)	20,0 % (13)
Svag	88,2 % (30)	11,7 % (4)

Ovenstående tabel viser sammenhængen mellem mekanismens styrke og outcome (Selv vurderet progression)
OBS: Her er kontekster på individniveau udeladt!

Test af mekanismen gennem intra-program sammenligning – eksempler på mulige analyser (fortsat)

Selvvurderet fysisk helbred (C) og alliancens styrke (M) sammenlignet med selvvurderet progression (O). Vist for sygemeldte borgere, kontanthjælpsmodtagere, uddannelseshjælpsmodtagere og forsikrede ledige

Kontekst (C) +	Alliancens styrke (Mekanismen)	Lav/middle progression (outcome 0)	Høj grad af progression (outcome 1)
Borgere med et dårligt selvvurderet fysisk helbred (C1)	Stærk (M1)	47,2 % (28)	52,8 % (39)
	Middel (M2)	69,2 % (18)	30,8 % (8)
	Svag/negativ(M3)	100,0 % (17)	0,0 % (0)
Borgere med et godt selvvurderet fysisk helbred(C2)	Stærk (M1)	20,2 % (24)	79,8 % (95)
	Middel (M2)	87,2 % (34)	12,8 % (5)
	Svag/negativ(M3)	76,5 % (13)	23,5 % (4)

En højere andel af borgere borgerne med et godt selvvurderet helbred der har en stærk alliance oplever en høj grad af progression som følge af jobcenterindsatsen sammenlignet borgere med en stærk alliance og et dårligt selvvurderet helbred. Tilsyneladende betyder alliancen mere for borgere med et godt helbred.

Videre logistiske regressionsanalyser viser, at alliancen har betydning for de svageste borgergruppers sandsynlighed for *selvforsøgelse*, men ikke for dagpengemodtagerne. (for hvem spørgsmålet)

Pause 15 minutter



AALBORG UNIVERSITY
DENMARK

CMO-konfigurationer



AALBORG UNIVERSITY
DENMARK

Konteksten i lag

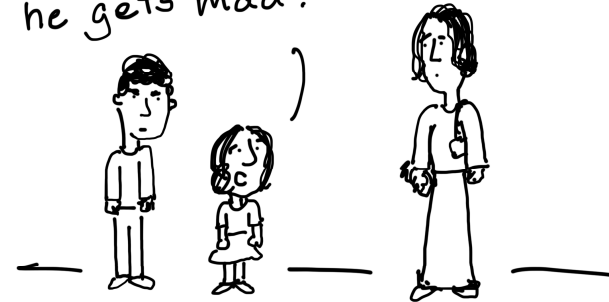
- 1) Individer – interessenters karakteristika og kapaciteter
- 2) Interpersonelle relationer – interessenternes relationer
- 3) Institutionelle kontekst – regler, normer og kulturer som omgiver programmet
- 4) Infrastruktur – den bredere sociale, økonomiske og kulturelle kontekst.

Ray Pawson (2013), The Science of Evaluation – A realist manifesto

Konteksten

- Konteksten opfattes som noget, der er nødvendigt for at en given kausal mekanisme sættes i spil, og dette betyder at der i forståelsen af konteksten ligger værdifuld viden om mekanismen.
- Man antager i den realistiske evaluering at der er en overskuelig mængde af relevante kontekster
- Hvor meget skal en situation ændre sig, før at vi med rette kan sige, at der er tale om en anden kontekst? I praksis svært at afgøre.

Mum. How come when you ask Henry to do something he does it? But when I ask, he gets mad?



RAMESESPROJECT.ORG



AALBORG UNIVERSITY
DENMARK

Moderator	Mekanisme 1: Rygeplastre virker ved at støtte en motivation til at holde op med at ryge	Mekanisme 2: Rygeplastre tilfører nikotin og gør derved cigaretter overflødige
Kontekst 1: Brugere betaler for rygeplastre	Brugere lægger cigaretterne	Brugere lægger cigaretterne
Kontekst 2: Brugere får rygeplastre gratis	Brugere fortsætter med at ryge	Brugere lægger cigaretterne

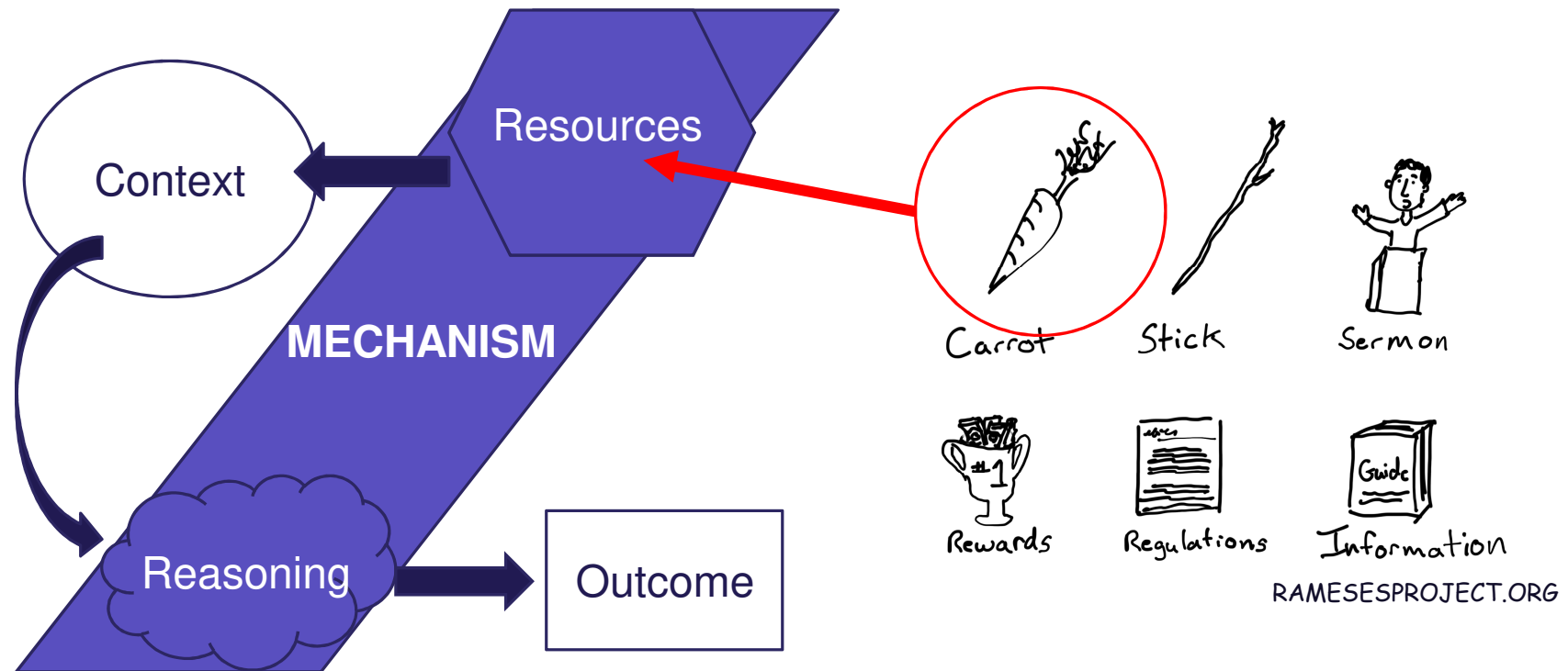
Kilde: Dahler-Larsen (2013)

Mekanisme – hvorfor virker indsatsen?

- Astbury and Leeuw (2010): “*underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest*”.
- Weiss (1997): “(a programme theory). . . deals with the mechanisms that intervene between the delivery of program service and the occurrence of outcomes of interest. It focuses on participants’ responses to program service. The mechanism of change is not the program service per se but the response that the activities generate.”
- Pawson (2013): “(...) programme mechanisms capture the way in which the programme’s resources impinge on the stakeholders’ reasoning”.



Ressourcer og reaktion



Kilde: Dalkin et al. (2015). Figure 1 A CMOC framework

Outcomes – hvad virker?

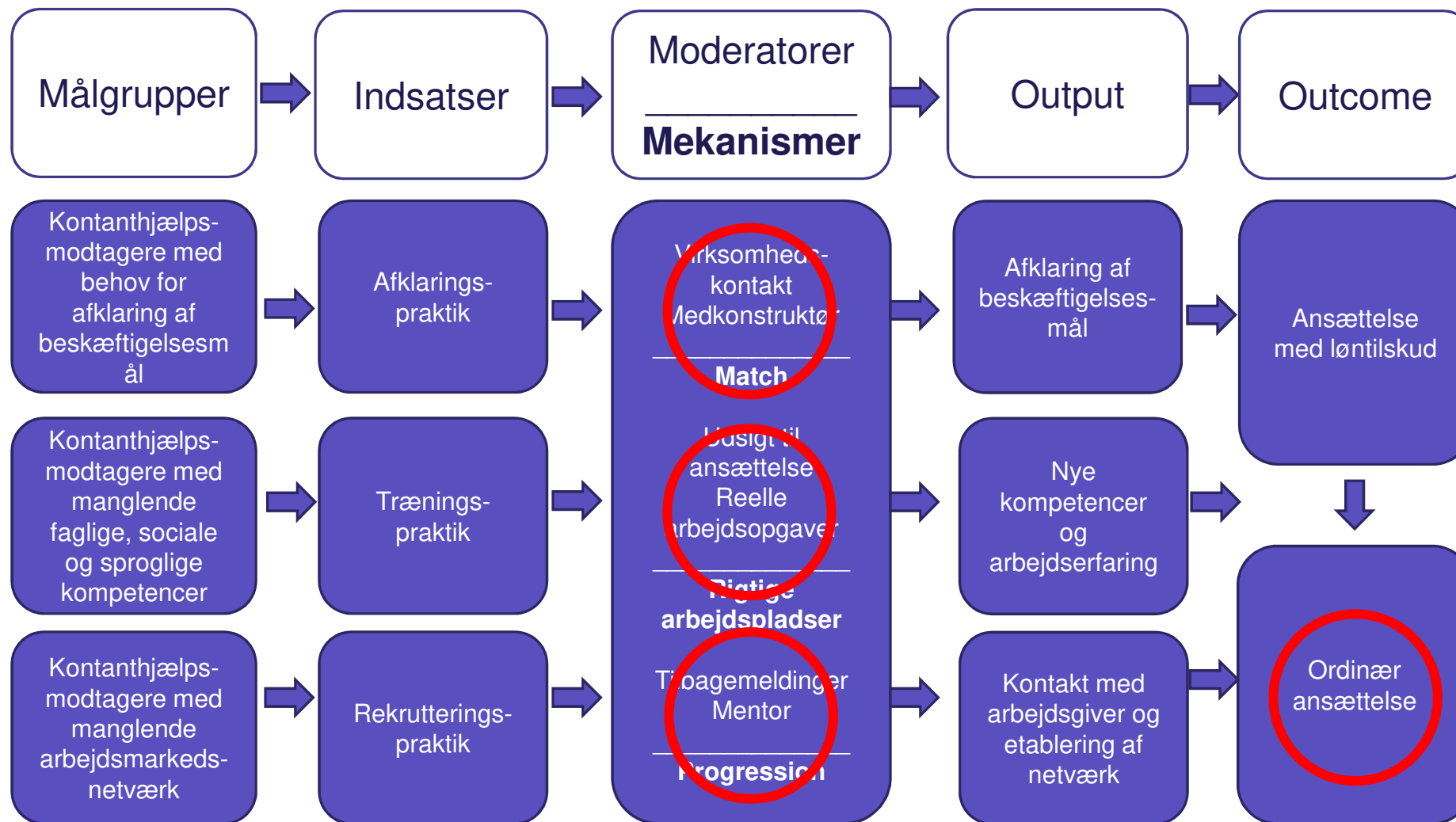
- Outcomes skal testes systematisk, fx ved brug af baseline eller nødvendige og specifikke tegn.
- En programteori bør *ikke* konstrueres og testes gennem de nøjagtig samme data/empiri.

”Realist evaluation presupposes pattern. There will be winners... and losers”

Ray Pawson (2013), The Science of Evaluation – A realist manifesto

Eksempel: Virkningsevaluering af virksomhedspraktik

- En virkningssyntese (realist synthesis) peger på tre mekanismer der har betydning for om virksomhedspraktik leder til beskæftigelse:
 - At borgeren oplever et godt match med virksomheden
 - At borgeren har oplevelsen af at være på en rigtig arbejdsplads
 - At borgeren oplever progression under praktikken.
- Formålet med virkningsevalueringen er at teste programteorien. Kan de tre mekanismer forklare beskæftigelsesgrad?



Indikatorer i survey

Match

- Klart formål med praktikken
- Motivation
- Relation chef
- Relation kollegaer
- Relation jobcenter.

Rigtige arbejdspladser

- Inddragelse i sociale arrangementer
- Samme type arbejdsopg.
- Meningsfulde opg.
- Klart formål med opg.
- Klare opg.
- Indflydelse
- Lære nye ting
- Tilfredshed med opg. samlet set
- Reel chance for ansættelse.

Progression

- Sidemandsoplæring
- Feedback.



Mekanismernes betydning for beskæftigelse

Rigtige arbejdspladser	β
Inddragelse i sociale arrangementer	0.046*
Samme type arbejdsopgaver som kollegaer	0.088**
Reelle chancer for at blive ansat på praktikvirksomheden	0.122***

"Jeg skulle være i praktik som rør lægger, men det var slet ikke det jeg fik lov til at lave, jeg blev sat til alt muligt andet, forfalden arbejde såsom oprydning efter de andre. Hente og bringe værktøj, grave sand væk med håndskovl. Jeg var der som arbejdsmand. Det var tydeligt fra starten at de ikke havde interesse i at ansætte mig. Når jeg ikke fik lov til noget som helst med rørlægning er det meget svært at vise at jeg faktisk er rigtigt god til det".



AALBORG UNIVERSITY
DENMARK

"Imødekomme kollegaer og chef, meningsfulde opgaver, godt arbejdsmiljø. Blev behandlet som en kollega på lige fod med de andre medarbejdere og fik ansvar - der blev udvist tillid til, at jeg kunne klare opgaverne trods det, at jeg hverken er uddannet eller har erfaring inden for området. Fik også gode anbefalinger med mig".

Gruppeøvelse

1. Udvælg en indsats, som I ønsker at virkningsevaluere
2. Lav en programteori
3. Overvej evalueringsdesign
4. Diskuter jeres design – hvad er de største muligheder og faldgrupper?
 - Dels ud fra gængse metodekriterier (validitet, reliabilitet og generaliserbarhed)
 - Dels ud fra virkningsevaluerings forskrifter (specifikation af CMO-konfigurationer).

Skabelon 1. Pilediagrammet



Skabelon 2. Udfaldskæden

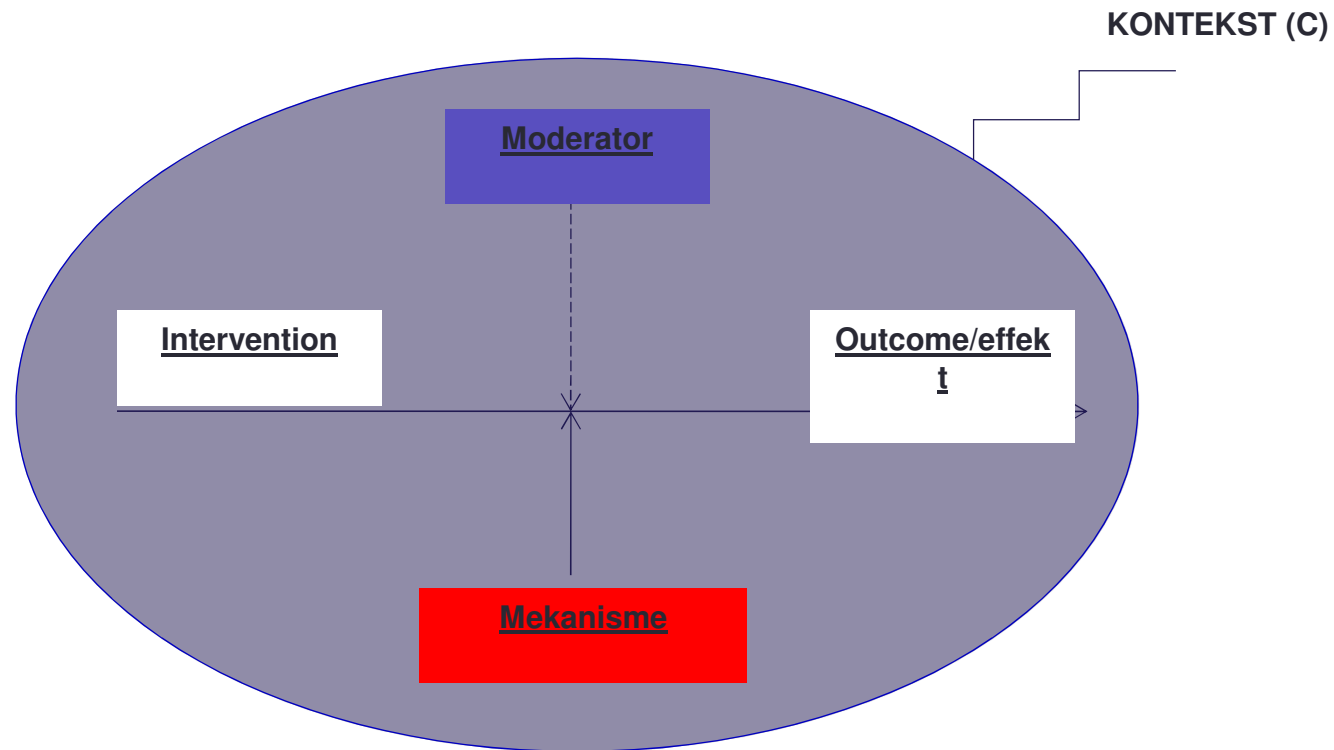


Skabelon 3. Den realistiske matrice

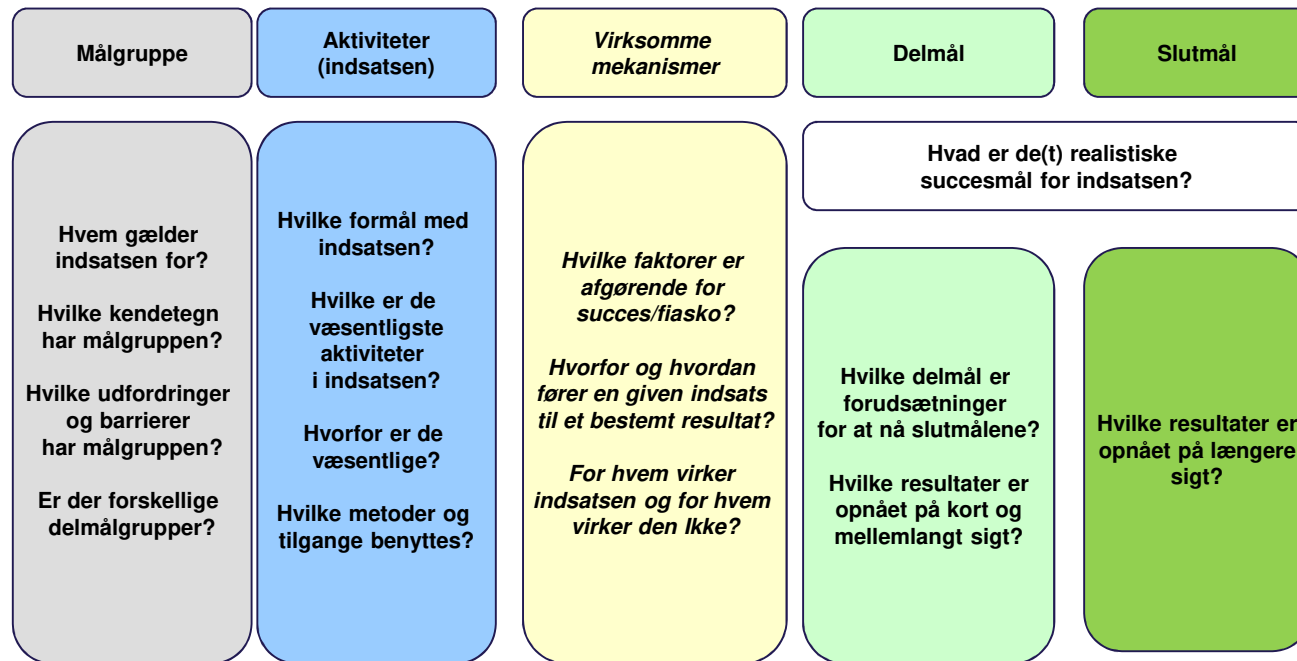
Kontekst (C)	+ Mekanisme (M)	= Outcome (O)
Skoler med mange overvægtige børn	+ Udlevering af æbler (ressource) som indeholder C-vitamin og erstatter (reaktion) usunde fødevarer	= Elever får bedre helbred



Skabelon 4. CMO-konfigurationen



Skabelon 5. Rutediagrammet



Skabelon 6. Sproglig formulering af programteori (*if-then* antagelser)

- Hvis der leveres gratis æbler til skolerne
- Som stilles frem i spisefrikvarteret
- Og eleverne ser æblerne
- Og spiser dem
- Og dermed undlader at spise deres normale mellemmåltid (fx usundt snack)
- Så vil det forbedre deres niveau af C-vitamin og deres BMI
- Hvorved de opnår bedre helbred.

Hvis du vil vide mere...

- R. Pawson og N. Tilley (1997): *Realistic Evaluation*, Sage publications.
- R. Pawson (2006): *Evidence-based policy – A realist perspective*, Sage Publications.
- R. Pawson (2013): *The Science of Evaluation – A Realist Manifesto*, Sage Publications.
- S.C. Funnell & P.J. Rogers (2011): *Purposeful Program Theory*, Josseybass.
- P. Dahler-Larsen & H.K. Krogstrup (2004): *Nye veje i evaluering*, systime Academic.
- H.K. Krogstrup (2011): *Kampen om Evidens*, Hans Reitzels Forlag.
- T. Bredgaard m.fl. (2011): *Hvad virker i aktiveringsindsatsen*, Beskæftigelsesregion Nordjylland.
- T. Bredgaard (red.) (2016): *Evaluering af offentlig politik og administration*, Hans Reitzels Forlag.

Kontaktinformation

Julia Salado-Rasmussen, jsr@dps.aau.dk, tlf. 2065 4021

Rasmus Ravn, rr@dps.aau.dk, tlf. 2091 0468

Forskningscenter for Evaluering (www.fce.aau.dk)

Center for arbejdsmarkedsforskning (www.carma.aau.dk)