

2020

# Realistisk evaluering: Et supplement eller et alternativ til effekt-evaluering?

---

**Rasmus Ravn**

Ph.d.-stipendiat, Institut for Statskundskab, Aalborg Universitet

## Resumé:

*Til at undersøge effekter anses det randomiserede kontrollerede forsøg (RCT<sup>1</sup>) som guldstandard for evidens. De senere år har eksperimentelle og kvasi-eksperimentelle tilgange til effekt-evaluering imidlertid fået konkurrence af teoridrevne evalueringstilgange. Disse nyere tilgange undersøger, hvad der virker, for hvem, under hvilke omstændigheder det virker, og hvorfor en indsats virker. Der er især tale om den "realistiske evaluering", der på dansk ofte benævnes "virkningsevaluering".*

*Artiklens formål er at illustrere, hvordan man som forsker eller evaluator vil kunne designe en evaluering, der kombinerer realistisk evaluering med mere klassisk effekt-evaluering. Der skelnes i artiklen mellem integreret og sekventiel modelanvendelse, og det konkluderes og illustreres, at kombinationsformen har afgørende betydning for, hvilke indsigter man kan nå frem som evaluator.*

*Blandt realistiske evaluatore er der imidlertid dyb uenighed om, hvorvidt modellerne kan og bør kombineres. Ophavsmændene til den realistiske evaluering er kritiske over for modelkombination, fordi de anser den realistiske kausalitetsforståelse som uforenelig med evaluering, hvor den kontrafaktiske situation etableres.*

*Hovedparten af artiklen bruges derfor på at blottlægge og diskutere evalueringmodellernes syn på kausalitet og metodeanvendelse ud fra et realistisk perspektiv. Det sker med henblik på at synliggøre de vigtigste argumenter for og imod at kombinere modellerne.*

---

<sup>1</sup> Randomized Controlled Trial

## Introduktion:

Offentlig politik evalueres som aldrig før. Næsten alle offentlige tiltag underkastes evaluering. Ifølge den svenske evalueringsforsker Evert Vedung befinder vi os i denne tidsperiode i den såkaldte "evidensbølge", hvor man fra politisk hold er begyndt at kræve, at der foreligger empirisk belæg for, at indsatser og politikker virker (Vedung, 2016). Her anses det randomiserede kontrollerede forsøg (RCT-forsøg) som guldstandard for evidens i det såkaldte evidenshierarki.

Gennem de seneste to årtier er teoribaserede evalueringsmodeller imidlertid blevet udviklet og har vundet større indpas blandt evaluatorene. Der er især tale om den såkaldte "realistiske evaluering" (Pawson & Tilley, 1997), på dansk ofte benævnt "virkningsevaluering" (Dahler-Larsen, 2013; Bredgaard, Salado-Rasmussen, & Sieling-Monas, 2016). Gennem anvendelse af programteorier undersøger den realistiske evaluering, hvad der virker, for hvem, og under hvilke omstændigheder det virker. En programteori består af eksplicite antagelser om, hvad der virker, for hvem det virker, og hvorfor det virker.

Effektevalueringer<sup>2</sup>, herunder RCT-forsøget, undersøger, om en intervention virker gennem etablering af den kontrafaktiske situation ved brug af en deltagergruppe og en kontrolgruppe. Udviklingen inden for kontrolgruppen indikerer, hvad der ville være sket for deltagergruppen, hvis de ikke havde modtaget en indsatsen (den kontrafaktiske situation) (Vedung, 2009; Hernán & Robins 2016; 3-10). Ved at sammenholde udviklingen inden for deltagergruppen med udviklingen i kontrolgruppen produceres der i effektevalueringer et kvantificerbart effektmål.

Realistisk evaluering undersøger derimod, hvad der virker, for hvem, under hvilke omstændigheder, og producerer teoretiske forklaringer på, *hvorfor* en indsats virker gennem opstilling og test af programteorier. Som en del af mit ph.d.-forløb har jeg foretaget et litteraturstudie af publicerede realistiske evalueringer. Studiet viste, at realistiske evalueringer sin rene form *ikke* producerer et kvantificerbart effektmål, fordi den kontrafaktiske situation ikke etableres. Praksis er i stedet, at der anvendes sammenligningsgrupper, til at vurdere for hvilke grupper indsatsen har virket bedst, hvilket også pointeres af andre realistiske evaluatorene (Westthorp, 2014).

Blandt visse realistiske evaluatorene er det blevet fremhævet, at de to modeller i forening vil kunne producere et kvantificerbart effektmål, samtidigt med at vi får en dybere forståelse for, hvorfor resultatet er indtruffet (Jamal et al., 2015; Bonell et al., 2013).

I den realistiske lejr hersker der imidlertid stor uenighed om, hvorvidt realistisk evaluering skal anses som et alternativ eller et supplement til effektevaluering. Stridsspørgsmålet går på, om modellerne er forenelige på trods af divergerende kausalitetsforståelser og forskellige videnskabsteoretiske afsæt (Marchal et al., 2013). De realistiske modstanderne af integration, der indbefatter Ray Pawson<sup>3</sup>, fremhæver,

---

<sup>2</sup> I artiklen refererer begrebet effektevaluering udelukkende til eksperimentelle eller kvasi-eksperimentelle tilgange til at undersøge effekter. Tilgange, der anvender rene statistiske metoder til at undersøge effekter, er ikke inkluderet i artiklens brug af begrebet.

<sup>3</sup> Ray Pawson er sammen med Nick Tilley ophavsmand til den realistiske evaluering.

at en evaluering, der kombinerer realistisk evaluering med enten eksperimentelle eller kvasi-eksperimentelle metoder, ikke vil kunne være oprigtigt realistisk (Van Belle et al., 2016).

På trods af kritikken og advarslerne mod at kombinere modellerne er formålet med artiklen at komme med forslag til, hvordan man i designfasen af en evaluering vil kunne kombinere realistisk evaluering og mere klassisk effektmåling. Når modellerne kombineres, mener jeg, der bør skelnes mellem *integreret* og *sekventiel* anvendelse af evalueringsmodellerne. Det skyldes, at kombinationsformen har betydning for, hvilke begrundede slutninger man som forsker kan få frem til på baggrund af undersøgelsen.

Når man i en forskningsmæssig sammenhæng træffer et design eller metodemæssigt valg, vil det altid kunne kritiseres. Det gør sig i særlig grad gældende, når metoder og modeller kombineres. Derfor bør man som forsker eller evaluator være bevidst om den kritik, der kan fremføres, hvis man anvender evalueringsmodeller, der baserer sig på forskellige ontologiske og epistemologiske grundlag, som det er tilfældet for effektevaluering og realistisk evaluering.

Af den årsag beskæftiger hovedparten af artiklen sig med at diskutere de realistiske evaluators stridspunkter omhandlende modelkombination samt med blotlægge de væsentligste argumenter for og imod modelkombination ud fra et realistisk standpunkt. Formålet med artiklen er således *ikke* at levere en definitiv dom over, om realistisk evaluering er et supplement eller et alternativ til effektmåling, eller om modellerne er forenelige. Det er i stedet at synliggøre og diskutere argumenter for og imod modelkombination, og illustrere hvordan en evaluering potentielt vil kunne designes, såfremt man ønsker at kombinere modellerne.

For at tydeliggøre artiklens formål undersøges derfor to sammenhængende spørgsmål:

1. Hvad er de væsentligste *realistiske* argumenter for og imod integreret og sekventiel anvendelse af effektevaluering og realistisk evaluering?
2. Hvordan vil en evaluering kunne designes, hvis man ønsker at kombinere modellerne?

Undersøgelsesspørgsmålene belyses for at besvare artiklens overordnede spørgsmål: "Er realistisk evaluering et alternativ eller et supplement til effektevaluering?"

For at besvare forskningsspørgsmålene præsenteres først kort de mest centrale antagelser i klassisk effektevaluering. Dernæst præsenteres den realistiske evaluering mere uddybende, og forskellene mellem modellerne sammenlignes. En grundlæggende forståelse af modellerne er en forudsætning for at kunne følge den videre diskussion.

Derefter påbegyndes artiklens analysedel, hvor modellernes betragtninger på kausalitet og metode- og modelforenelighed belyses og diskuteres. I forhold til behandlingen af effektevalueringen struktureres diskussionen med udgangspunkt i RCT-forsøget, men bør have samme gyldighed i forhold til metoder, hvor statistisk matching anvendes (kvasi-eksperiment), fordi den underliggende logik er den samme.

Modellernes syn på anvendelse af kvalitative interviews til at undersøge effekter blotlægges i et særskilt afsnit, fordi det kvalitative interview er en velegnet case til at

illustrere stridspunkter mellem modellerne. Det vises ydermere i analysen, hvordan visse realistiske evaluatore er begyndt at indarbejde indsigter fra RCT-forsøget, hvilket har medført betydelig diskussion i den realistiske lejr.

Der vil i artiklen kunne findes argumenter til fordel for at kombinere modellerne, og til slut i artiklen fremlægges bud på, hvordan modellerne vil kunne kombineres i et evalueringsdesign. Her skelnes der mellem integreret og sekventiel anvendelse af evalueringsmodellerne.

Ved integreret anvendelse skabes der en hybrid mellem effektevaluering og realistisk evaluering, så modellerne udgør én samlet evalueringsmodel. Ved sekventiel anvendelse anvendes modellerne adskilt fra hinanden og udføres i overensstemmelse med modellernes egne principper. Indsigterne fra de to modeller supplerer her blot hinanden.

## Effektevaluering kort fortalt

I effektevaluering er det randomiserede kontrollerede forsøg (RCT) det arketyperiske eksempel på, hvordan der bør produceres evidens og identificeres effekter (Hernán & Robins, 2016; 13-20).

I randomiserede kontrollerede forsøg (og i kvasi-eksperimenter) er begrebet 'den kontrafaktiske situation' centralt, da etablering af en kontrafaktisk situation er midlet til at påvise en effekt (Hernán & Robins, 2016; Vedung, 2009). Den bærende logik er, at man undersøger, hvad der ville ske, hvis man *ikke* havde iværksat en given indsats. I RCT-forsøget etableres den kontrafaktiske situation ved, at et sample tilfældigt inddeles i en deltagergruppe (der modtager en indsats) og i en kontrolgruppe (der ikke modtager den pågældende indsats). Den tilfældige inddeling (randomiseringen) sikrer, at deltagergruppen og kontrolgruppen ligner hinanden på alle parametre, på nær at det kun er deltagergruppen, der modtager indsatsen. *Effekten* er således forskellen mellem andelen af borgere i deltagergruppen og kontrolgruppen, der har opnået det ønskede udfald.

Randomiseringen medfører, at der i et RCT-forsøg er en høj grad af intern validitet, da der ikke eksisterer andre forklaringer på, at et resultat er indtruffet, end at det skyldes indsatsen. Derimod kan der sættes spørgsmålstegn ved graden af den eksterne validitet (vil sammenhængen kunne generaliseres og ligeledes gælde i en anden kontekst?). Vi nærmer os først noget der ligner universel og kontekstafhængig viden (ekstern validitet), når adskillige RCT-forsøg har påvist en effekt af en bestemt type intervention. Det bedste middel til at akkumulere denne form for viden og skabe *evidens* er gennem statistiske metaanalyser, der udregner de gennemsnitlige effekter af en klart defineret type indsats på baggrund af resultaterne af en lang række eksperimentelle studier (Baadsgaard & Krogstrup, 2016)

I effektevalueringer er der tale om en variansbaseret tilgang til studiet af kausalitet (Dahler-Larsen, 2013), fordi det eneste, der i princippet adskiller deltagergruppen fra kontrolgruppen, er, at det udelukkende er deltagergruppen, der har modtaget den pågældende indsats. Kun én "uafhængig" variabel skal i effektevalueringen variere

(modtagelse af indsatsen eller ej). Det bør føre til at variation i den afhængige variabel (outcome) *kun* kan tilskrives indsatsen (Bredgaard, Drejer & Nielsen, 2016; 289).

RCT-forsøget (og kvasi-eksperimentelle tilgange) bygger på en 'alt andet lige' antagelse (Morgan & Winship, 2015; 42–43; Angrist & Pischke, 2015; xi-xv). Gennem randomiseringen antager man, at alle andre faktorer holdes konstante, eller at de øvrige kontekstuelle faktorer påvirker deltager- og kontrolgruppen på samme måde. Man forsøger således "kunstigt" at producere et lukket socialt system (Morgan & Winship, 2015; 7), så man kan tilskrive "effekten" til indsatsen, ved at sammenligne "lige med lige". I effektevaluering har man etableret en kausal sammenhæng, når det udelukkende er indsatsen, der kan forklare forskellen mellem indsats- og kontrolgruppens udfald. Dette kan beskrives som en variansbaseret (eller successionistisk) tilgang forståelsen af kausalitet, fordi alle andre forklaringer på resultatet udelukkes (Bredgaard, Salado-Rasmussen & Sieling-Monas, 2016; 336–337).

I sin mest simple form betyder denne kausalitetsforståelse, at A (indsatsen) fører til B (resultatet), men man ved ikke nødvendigvis præcis, hvad det er, der har produceret resultatet (Dahler-Larsen & Krogstrup, 2003; 97). Kort sagt er det variationen i den uafhængige variabel (indsats eller ingen indsats), der skaber resultatet.

Deraf følger også, at effektevaluering baserer sig på en ontologi, hvor der eksisterer en objektiv virkelighed uafhængigt af observatøren (Ingemann, 2013; 53). Epistemologisk er afsættet for effektevaluering, at vi vil kunne få viden om, i hvilken grad en indsats virker ved brug af raffinerede kvantitative (eksperimentelle) metoder (Rowe & Oltmann, 2016; Bolton, 2008).

Af realistiske evaluatore kritiseres RCT-forsøget ofte for at blive til en "black box", hvor man ikke ved, hvad det er, der foregår i indsatsen, som skaber effekterne (Pawson & Tilley, 1997). Kritikken er dog kun berettiget, hvis man udelukkende ser på "rene" RCT-forsøg.

## REALISTISK EVALUERING

Den realistiske evaluering er udviklet på baggrund af en overvejende videnskabs-teoretisk kritik af RCT-forsøget. Den primære kritik af RCT-forsøget fra realistisk side går på, at RCT-forsøget i sin rene form ikke producerer forklaringer på, hvorfor et resultat er indtruffet. Det konstateres blot, at indsatsen har haft en effekt. Indsatsen forbliver i RCT-forsøget en "black box" (Pawson & Tilley, 1997; 30).

Den realistiske evalueringens videnskabsteoretiske udgangspunkt benævnes som enten videnskabelig realisme eller korrigerbar realisme (Pawson, 2013; 84) og anses som en brobygger mellem positivisme og konstruktivisme (Bredgaard, Salado-Rasmussen & Sieling-Monas, 2016).

Den videnskabelige realisme antager på linje med effektevalueringen, at der eksisterer en virkelighed uafhængigt af observatøren. Modellens ontologiske udgangspunkt er således realistisk, som navnet på modellen også antyder. I lighed med den kritiske realisme (Bhaskar, 1978; Ingemann, 2013; Andersen, 2007) anses virkeligheden dog her som stratificeret og lagdelt. Dele af den eksisterende "virkelighed" skal forefindes på "dybe" niveauer, der ikke kan observeres direkte af forskeren. Det bevirker, at vor

viden om ”virkeligheden” altid vil være usikker, kontekstafhængig og provisorisk (Pawson, 2013).

Af flere fremtrædende realistiske evaluatore er modellens epistemologiske udgangspunkt blevet karakteriseret som tilnærmelsesvis konstruktivistisk (Jagosh, 2017; Westhorp, 2017). Selvom der eksisterer en objektiv virkelighed uafhængigt af observatøren, er vores viden om ”virkeligheden” både individuelt og socialt konstrueret, og selv samme ”virkelighed” udøver indflydelse på, hvilke fortolkninger af ”virkeligheden” der er ”fornuftige”, jævnfør også den kommende tabel 1.

Til forskel fra effektevalueringens anvendelse af variansbaserede kausalteori abonneres der i den realistiske evaluering på den generative kausalitetsforståelse. Antagelsen er her, at ”usynlige” underliggende mekanismer genererer forandringer i både i interventioner og samfundet generelt (Pawson, 2013). Mekanismer er ud fra den realistiske forståelse ”virkelige”, selvom de ikke kan iagttages med det blotte øje, fordi de befinder sig på de ”dybe” niveauer af virkeligheden. Afsættet for den realistiske evaluering er, at mekanismerne kun vil træde i kraft, når bestemte omstændigheder er til stede, altså i bestemte kontekster. Det er således konteksterne, der afgør, om mekanismerne træder i kraft.

Den programteoretiske formel, der i realistisk evaluering anvendes til at undersøge evaluanden (det der evalueres), lyder:

$$\textit{Kontekst} + \textit{Mekanisme} = \textit{Outcome/Resultat}$$

Ovenstående formel benævnes i realistisk evaluering som en CMO-konfiguration (Kontekst-Mekanisme-Outcome-konfiguration). Som en del af en realistisk evaluering udarbejdes og testes ofte flere af disse CMO-konfigurationer. Udarbejdelsen af programteorien kan foregå enten induktivt eller deduktivt (Bredgaard, Salado-Rasmussen, and Sieling-Monas 2016). Ved brug af den deduktive tilgang vil man som evaluator typisk finde inspiration til programteorien ved at konsultere den eksisterende empiriske forskning på området og/eller ved brug af ”formel” teori. Typisk vil der blive anvendt ”middle range teori”, der befinder sig mellem ”grand theory” og den specifikke empiri. Det kan f.eks. være teorier om tillid, samarbejde, læring eller organisationsteori, der kan anvendes til at informere programteorien eller CMO-konfigurationen. Ved den induktive tilgang vil programteorien i stedet blive udledt fra praksis. Ofte vil programteorien udledes gennem interviews med projektmedarbejdere eller frontlinjepersonale, hvilket bevirker, at programteorien tilpasses den lokale kontekst. I praksis vil udarbejdelsen af programteori og CMO-konfigurationer ofte ske på baggrund af begge tilgange (Funnel and Rogers, 2011; 108–116). En kort uddybende bemærkning skal her knyttes til begreberne ”programteori” og ”CMO-konfigurationer”. *Programteorien* er de generelle underliggende antagelser om, hvordan den samlede indsats vil fungere, og hvorfor den vil virke. Programteorien vil ofte udgøres af adskillige *CMO-konfigurationer*. I CMO-konfigurationerne ”zoomer” man ind på et bestemt element af indsatsen med henblik på at belyse det nærmere. Med andre ord ligger CMO-konfigurationer på et lavere abstraktionsniveau end den samlede programteori.

En god programteori skal ligeledes være *testbar*. Den skal kunne underkastes empirisk belysning. Det vil ske ved at evaluator allerede i forbindelse med udarbejdelsen

programteorien og CMO-konfigurationerne gør sig overvejelser omkring, hvilke *indikatorer* (empiri eller datagrundlag) der kan anvendes til at belyse de forskellige komponenter.

Uanset hvordan CMO-konfigurationerne udarbejdes, og hvilke indikatorer der anvendes, skal programteorien illustrere, at det er ”mekanismerne”, der producerer resultaterne, men at de kun træder i kraft under bestemte omstændigheder (kontekster). Mekanismen kan opdeles i to komponenter: ressourcer og ræsonnementer/adfærdsændringer.

Som en del af en indsats stilles nogle ressourcer til rådighed (f.eks. viden, pisk eller gulerod), som får deltagerne til at reflektere og ræsonnere, hvilket får deltagerne til at ændre deres adfærd (Dalkin et al. 2015). Mekanismen består således både af ressourcedelen og refleksionen, der får deltagerne til at ændre deres handlinger.

Kontekster I realistisk evaluering er en bredere kategori. Kontekster kan anses som karakteristika ved individet (f.eks. køn, alder, etnicitet, uddannelse med videre). Men kontekster kan også findes på det interpersonelle niveau, på det institutionelle niveau og på makro-niveau (Pawson 2013;:37) i form af f.eks. samarbejdsgrad, normer, kulturer eller økonomiske konjunkturer.

Den realistiske forståelse af begrebet ”effekt” adskiller sig betydeligt fra den klassiske forståelse af begrebet ”effekt”. Ved en nærlæsning af Pawsons tre bøger om realistisk evaluering, nævnes det ikke ét sted, at målet med realistisk evaluering er at undersøge effekter/impact ved at producere et kvantificerbart effektmål. Det omtalte litteraturstudie viste også, at realistiske evalueringer i praksis ikke producerer et kvantificerbart effektmål gennem etableringen af den kontrafaktiske situation. Målet er i stedet at undersøge outcomes (udfald/resultater) gennem et fokus på, hvilke mekanismer og kontekster der har bidraget til at skabe udfaldet.

Den generative kausalitetsforståelse forudsætter, at mennesker ændrer deres adfærd som følge af, at der stilles ressourcer til rådighed for dem som en del af en indsats, hvilket får dem til at reflektere, ræsonnere og handle anderledes. Derfor er der tale om en ”effekt” i realistisk evaluering, når mekanismen og konteksten i forening har produceret et outcome. Gennem test af programteorien *sandsynliggøres* det, at det er indsatsen, der har produceret resultatet. Der fældes ikke en klar dom over, om en indsats virker, eller hvor meget den virker. Det illustreres meget godt af følgende citat af Pawson og Tilley, hvor de opsummerer den realistiske evalueringens formål:

*”Realist evaluations asks not, ‘What works?’ or, ‘Does this program work?’ but asks instead, ‘What works for whom in what circumstances and in what respects, and how?’” (Pawson & Tilley, 2004;:2).*

Citatet viser, at realistiske evalueringer ikke undersøger, *om* et program eller en indsats virker (gennem sammenligning med en kontrolgruppe). Det konventionelle ”effektspørgsmål” berøres ikke direkte i realistiske evalueringer. Ud fra den konventionelle forståelse af effektbegrebet er der i realistisk evaluering ikke tale om en effektanalyse, men snarere en outcome-analyse. Alligevel er realistisk evaluering en form for effektevaluering – effektbegrebet adskiller sig blot betydeligt fra den konventionelle forståelse, som det skal blive uddybet i det følgende.

Til at konkretisere de forskellige effektforståelser kan vi anvende et tænkt eksempel som illustration. Vi kan forestille os en beskæftigelsesindsats, hvor ledige opkvalificeres fra ufaglærte til faglærte gennem en uddannelsesindsats. Den gængse måde at nå frem til, om uddannelse virker, vil være tilfældig inddeling i to grupper. Én gruppe der modtager uddannelsesindsatsen, og én gruppe der ikke gør. Hvis en større andel af borgerne, der modtag uddannelse, er i job efter en periode eller har en højere beskæftigelsesgrad, er der en positiv effekt af indsatsen.

I realistisk evaluering undersøges ”effekten” (outcomes) oftest gennem intra- eller inter-programvariation/programsammenligning (Pawson & Tilley, 1997; Westhorp, 2014). Intra-programvariation betyder, at outcomes for forskellige deltagergrupper undersøges indenfor *samme* indsats. Dermed anskueliggøres det, at indsatsen har virket bedre for nogle deltagere end for andre. Se blandt andet Pawson og Tilley for konkrete eksempler på intra-programvariation (Pawson & Tilley, 1997; Pawson, 2013).

Hvis vi igen tager udgangspunkt i vores tænkte eksempel med en uddannelsesindsats for ledige, vil intra-programmsammenligning kunne foretages ved, at man undersøger, om uddannelsesindsatsen har virket bedre for nogle grupper af de ledige end for andre. I den eksisterende forskning i opkvalificering og efter- og videreuddannelse skelnes der mellem lystdrevet motivation (se f.eks. Klindt & Houman, 2010) og nødvendighedsdrevet motivation (se f.eks. Keller, 2004). Ydermere viser en gennemgang af den eksisterende forskning, at især yngre kvinder i højere grad end andre grupper *ønsker* (lystdrevet motivation) at efteruddanne sig, samt at de i højere grad gør det. Midaldrende mænd er omvendt meget lidt motiverede for at uddanne sig og gør det sjældent – selv under ledighedsperioder (Christensen et al. 1998). Som følge heraf kan man opstille en programteoretisk antagelse om, at (yngre) kvinder vil have det største udbytte af uddannelsesindsatsen, fordi de som følge af deres lystdrevne motivation (mekanismen) vil være tilføjelige til at deltage (mere) engagerede og motiverede i indsatsen. Såfremt en større andel af de kvinder, der modtog uddannelse, er kommet i job, sammenlignet med de mænd, der ligeledes deltog i uddannelse, kan man nå frem til, at indsatsen virkede bedre for kvinder end for mænd på grund af deres højere lystdrevne motivation.

Inter-programvariation anvendes ligeledes i den realistiske evaluering. Her sammenlignes outcomes mellem sammenlignelige indsats, der blot har udspillet sig under forskellige omstændigheder (kontekster), for at anskueliggøre at en indsats virker bedre i nogle kontekster end i andre. Med vores tænkte eksempel in mente kan vi ved inter-program sammenligning forstille os, at vi undersøger, om uddannelsesindsatsen virker bedre i f.eks. vestjyske kommuner end i hovedstadskommuner. Denne antagelse kan vi basere f.eks. basere på deskriptiv statistik, der viser, at andelen af ufaglærte ledige gennemsnitligt er højere i vestjyske udkantskommuner (se f.eks. jobindsats.dk) end i hovedstadskommuner. I tillæg til dette viser en analyse fra Danske Regioner (Danske Regioner 2016), at der i Nord- og Midtjylland er mangel på faglært arbejdskraft inden for både bygge og anlæg samt inden for det merkantile område, mens der i Region Hovedstaden hovedsageligt er mangel på arbejdskraft med en mellemlang teknisk uddannelse. Hvis en større andel af deltagerne i de vestjyske kommuner er kommet i job efter indsatsen, kan man nå frem til, at uddannelse fra ufaglært til faglært virker (bedre) i de (vestjyske) udkantskommuner, og at det skyldes, at uddannelsesindsatsens kompetencer er mere efterspurgt (mekanismen) i Vestjylland.



Inter- og intra programsammenligning har det til fælles, at det ikke afsiges en *endelig* dom over, om indsatsen virker eller ikke virker – jævnfør også ovenstående citat. I stedet undersøges det, om en indsats virker bedre for nogle deltagere end for andre, eller om indsatsen virker bedre i nogle kontekster end i andre – med det forbehold at vores viden grundlæggende er meget usikker.

Et andet eksempel på brugen inter-programsammenligning kan findes i den realistiske syntese (Pawson, 2006). Den realistiske syntese er den realistiske evalueringens bud på en tilgang til systematiske litteraturstudier. Det særlige ved den realistiske syntese er, at det er en *teoribaseret* tilgang til litteraturgennemgange. Tilgangen er udviklet i opposition til statistiske meta-analyser, der på baggrund af forskning, der rangerer højt i evidenshierarkiet og lever op til strenge inklusionskriterier, udregner de gennemsnitlige effekter af en bestemt type indsats. Præcis som i realistisk evaluering udarbejdes, testes og raffineres programteori i den realistiske syntese. I stedet for at undersøge under hvilke omstændigheder et specifikt program virker (som i realistisk evaluering), undersøges det i den realistiske syntese, under hvilke omstændigheder en ”familie” af programmer virker. I stedet for at anvende førstehåndsempiri bruges i stedet eksisterende forskningspublikationer til at teste programteorien. I realistisk syntese undersøges det på denne vis, om indsatser eller programmer, der ligner hinanden, virker forskelligt i forskellige kontekster. Målet er at nå frem til, hvilke omstændigheder der hæmmer eller fremmer virkningerne af en bestemt type indsats. Bevæggrunden for udviklingen af realistisk syntese er, at programmagerne har behov for denne type viden (og ikke kun viden om gennemsnitseffekter), når man skal designe en ny indsats. For et dansk eksempel på brugen af realistisk syntese se f.eks. Salado-Rasmussen og Bredgaard (2016).

I effektmålinger sammenligner man som nævnt ”lige med lige” (kontrolgruppe) gennem f.eks. randomisering eller matching. I realistisk evaluering antager man ikke, at noget er lige. Derfor sammenligner man i stedet resultaterne for forskellige grupper, der har deltaget i *samme* indsats, eller resultaterne fra to (eller flere) indsatser, der ligner hinanden, men udspiller sig i *forskellige* kontekster (intra- eller interprogram variation). Det hænger logisk sammen med, at man i realistisk evaluering undersøger, *for hvem og under hvilke omstændigheder* en indsats virker (bedre). I realistisk evaluering kan man således tale om ”effekter”, når det er sandsynliggjort, at konteksten og mekanismen i forening har bidraget til at skabe de(t) observerede outcome(s).

Når realistisk evaluering udelukkende sandsynliggør, at en indsats har bidraget til at skabe et resultat, betyder det også, at realistisk evaluering har en betydeligt lavere grad af intern validitet end evalueringer, der anvender eksperimentelle eller kvasi-eksperimentelle metoder. Samtidigt forkaster den realistiske evaluering antagelsen om høj ekstern validitet (generalisering på tværs af kontekster), fordi det der virker i en sammenhæng, ikke nødvendigvis gør det i en anden.

I stedet er det vigtige i realistisk evalueringssammenhæng at finde ud af, under hvilke omstændigheder (i hvilke kontekster) mekanismerne aktiveres, og hvorfor indsatsen virker gennem et samspil mellem teori og tilbunds gående kvalitativ og kvantitativ empiri. Hvad angår generaliserbarhed, så er målet i realistisk evaluering at opnå teoretisk/analytisk generaliserbarhed (Collins et al., 2007; Onwuegbuzie & Leech 2010). Generaliseringspraksisser i realistisk evaluering vil tage form af ”læresætninger”

med udgangspunkt i CMO-konfigurationerne (Dahler-Larsen & Krogstrup, 2003; 108): Under de og de omstændigheder er sandsynligheden større, for at indsatsen virker, men det er under forudsætning af, at de og de mekanismer aktiveres, hvilket de og de tiltag kan øge sandsynligheden for (Pawson, 2006; 100; Pawson & Tilley, 2004). Ambitionen om universel generaliserbar viden forkastes derfor i realistisk evaluering.

### **Analyse: Modelforenelighed?**

Til brug i dette analyseafsnit præsenteres Tabel 1 nedenfor. Den illustrerer de væsentligste forskelle mellem effektevaluering og realistisk evaluering. Først bør der lægges mærke til, at undersøgelsesspørgsmålene har forskellig karakter. Effektevaluering undersøger om, og i givet fald hvor meget, en indsats virker. Den realistiske evaluering undersøger i kontrast hertil, hvad der virker for hvem under hvilke omstændigheder og hvorfor. Med andre ord undersøger det, for hvem indsatsen virker (bedst), eller om indsatsen virker bedre i nogle kontekster end andre.

De forskellige undersøgelsesspørgsmål har rod i divergerende forståelser af ”effektbegrebet”. I effektevalueringen er effekten lig forskellen mellem den faktiske og den kontrafaktiske situation. I realistisk evaluering etableres den kontrafaktiske situation imidlertid ikke. Det medfører, at forståelsen af effektbegrebet kan fortolkes som de *observerede inter- eller intra-program forskelle*. I realistisk evaluering bedrives der således også effektmåling, men forståelsen af effektbegrebet adskiller sig betydeligt fra den klassiske forståelse. Ud fra den klassiske effektforståelse vil der imidlertid kun være tale om en outcome-analyse.

Kausalitetsforståelserne er endvidere divergerende, og der anvendes forskellige metoder og datamateriale i de to modeller. Endvidere er der forskellige generaliseringspraksisser forbundet med modellerne, og viden akkumuleres på forskellig vis. Som beskrevet er der modelforskelle i forhold til hvilke data, der anses som gangbare i forhold til at undersøge ”effekter”. I det følgende afsnit uddybes modellernes syn på anvendelse af kvalitative interviews til at undersøge effekter. Det kvalitative interview anvendes som case, fordi det er velegnet til at illustrere de øvrige stridspunkter fra tabel 1.

Tabel 1: Forskelle mellem effektevaluering og realistisk evaluering.

	Effektevaluering	Realistisk evaluering
<b>Undersøgelsspørgsmål</b>	Virker det? (Og hvor meget virker det?)	Hvad virker for hvem, under hvilke omstændigheder, og hvorfor?
<b>Forståelse af effektbegrebet</b>	Effekten er lig forskellen mellem den faktiske og den kontrafaktiske situation	Intra- og inter-program variation:  Virker indsatsen bedre for nogle deltagergrupper end andre? Virker indsatsen bedre i nogle kontekster end andre?  Der fokuseres på <i>observerede</i> outcomes blandt deltagere.
<b>Kausalitetsforståelse</b>	Variansbaseret kausalitetsforståelse (successionistisk)	Generativ kausalitetsforståelse
<b>Metode</b>	Kvantitative metoder	Multimetodisk: Både kvalitative og kvantitative metoder
<b>Data</b>	Kvantitativt datamateriale	Både kvalitativt og kvantitativt datamateriale
<b>Ontologi</b>	Realistisk (Der eksisterer en objektiv virkelighed uafhængigt af observatøren)	Realistisk (Der eksisterer en objektiv virkelighed uafhængigt af observatøren, men den er stratificeret og lagdelt)
<b>Epistemologi</b>	Realistisk  (Vi kan få viden om, i hvilken grad en indsats virker, ved brug af raffinerede kvantitative (eksperimentelle) metoder	Tilnærmelsesvis konstruktivistisk  (Menneskelig viden er socialt konstrueret. Selvom der eksisterer en "virkelighed", er det vanskeligt at få viden om den, fordi virkeligheden er lagdelt, og mekanismerne skal findes på det "dybere" niveau.)
<b>Generaliseringspraksis</b>	RCT-studiet har en høj grad af intern validitet, men der er generaliseres sjældent ud fra et enkeltstående effektstudie.  Viden akkumuleres i stedet på baggrund af kvantitative meta-analyser (systematiske reviews)	Teoretisk generaliserbarhed i de enkelte studier. Der generaliseres med udgangspunkt i CMO-konfigurationerne.  Viden akkumuleres på baggrund af teoribaserede litteraturstudier (realistiske synteser)

Kilde: Egen udvikling med inspiration fra Bredgaard, Salado-Rasmussen & Sieling-Monas (2016)

### KAN KVALITATIVE DATA ANVENDES TIL AT UNDERSØGES EFFEKTER?

Et stridspunkt mellem de to evalueringsmodeller går på, om kvalitativ empiri kan bruges til at undersøge effekter, jævnfør tabel 1. Modellernes syn på dette gennemgås i det følgende, eksemplificeret ved en effektevaluator (i form af en eksperimentalist) og en realistisk evaluator.

En effektevaluator vil i en evalueringmæssig sammenhæng være interesseret i at undersøge, *om* og *i hvilken grad* en indsats har virket. Vedkommende vil derfor af flere grunde afvise brugen af kvalitative data til at undersøge effekter. Som vi erindrer, er

den effektevaluatoriske kausalitetsforståelse variansbaseret. Det er variationen i den afhængige variabel (indsats/ingen indsats) og den tilfældige inddeling i deltagergruppe eller kontrolgruppe, der giver os viden om effekten. Den kontrafaktiske situation vil givet kunne etableres på et *usikkert* grundlag ved brug af kvalitative data. En metode til dette kan være såkaldt skyggekontrol (Vedung 2009;247-251), hvor en ekspert på områdes, spørges om, hvad der ville være sket, hvis indsatsen ikke var blevet iværksat. Ifølge en eksperimentalist er problemet med denne kvalitative tilgang, og kvalitative data generelt, at de er forbundet med et subjektivt bias. Så snart man anvender kvalitative data (interviews, observation, dokumenter), er der vedhæftet et subjektivt bias (enten i indsamlingen af data eller ved fortolkningen af data). Det betyder, at resultaterne kun i lav grad vil være replicerbare, og at konklusionerne vil have en lav grad af validitet. Af den årsag mener eksperimentlisten, at strenge metodiske forskrifter (randomisering) skal overholdes, fordi metoden skal være replicerbar, og fordi der ikke levnes plads til subjektivitet, når metoden anvendes. Kvalitative data vil givetvis kunne anvendes til at undersøge andre forskningsspørgsmål eller generere hypoteser, men kun effektspørgsmålet har egentlig forskningsmæssig relevans.

Den realistiske evaluator vil derimod mene, at man *skal* anvende kvalitative data, når man undersøger outcomes, og hvad der har bidraget til at producere disse outcomes (Pawson, 2013). Den bedste måde at nå indsigt i, *hvorfor* en indsats har virket for personen (mekanismen), er derfor ved at spørge enten programmagerne, frontlinjemedarbejdere eller programdeltagere. Her indtager evaluator rollen som detektiv (Dahler-Larsen, 2013; 136-137), der som forberedelse til en "retssag" leder efter "spor", der kan overbevise en jury om en persons (u)skyld. Målet er at *sandsynliggøre* på overbevisende vis, at "noget" i indsatsen har bibragt det ønskede resultat.

Når f.eks. det kvalitative interview anvendes i realistisk evaluering, bruges det derfor til at besvare delspørgsmålet: *Hvorfor* virker det (når det virker)? Her vil realistiske evaluatorene mene, at de oplagte svar skal findes hos programmedarbejdere eller programdeltagere, eftersom disse vil ligge ind med "tavs" viden. En realistisk evaluering kan dog heller ikke undvære kvantitative data ifølge Pawson, fordi outcomes ikke kan undersøges kvalitativt – det kræver håndfaste opgørelser i form af kvantitative data (Pawson, 2013; 20–21).

Som udgangspunkt bidrager de forskellige empirikilder således til at besvare forskellige spørgsmål i realistisk evaluering. Denne opfattelse finder vi hos Ray Pawson, hvilket illustreres på sigende vis i følgende citat:

*"As a first approximation one can say that mining mechanisms requires qualitative evidence, observing outcomes is quantitative, and that canvassing contexts requires comparative and sometimes historical data" (Pawson, 2013; 19)*

De forskellige datatyper bidrager således med forskellige elementer i erkendelsesprocessen. Den realistiske evaluator vil derfor mene, at forskellige former for empiri er at foretrække, afhængigt af om det er mekanismer, kontekster eller outcomes, der undersøges.

Den erkendelsesmæssige interesse i realistisk evaluering forbliver at få viden om de "virkelige" mekanismer, der producerer et outcome under bestemte omstændigheder.

Ofte vil det kvalitative interview være det mest velegnede til at få forståelse for, hvorfor indsatsen virker (mekanismen).

Umiddelbart tyder det således på, at modellerne er inkompatible på dataspørgsmålet. Det skyldes, at de klassiske effektevaluatorer afviser kvalitativ empiri (herunder interviews) til at undersøge effektspørgsmålet, mens realistiske evaluatorer fremhæver den kvalitative empiris nødvendighed, fordi en effekt (kausalitet) ikke kan undersøges, uden at der produceres forklaringer på, *hvorfor* effekten er indtruffet.

### **REALISTISK EVALUERINGS TILNÆRMELSE TIL EFFEKTMÅLING? REALISTISKE ARGUMENTER FOR OG IMOD INTEGRATION OG SEKVENTIEL ANVENDELSE**

Den realistiske lejr er yderst splittet med hensyn til, om det er acceptabelt at anvende indsigter fra effektmåling i en realistisk evaluering. De primære stridspunkter er videnskabsteoretiske og vedrører, om man som realist kan acceptere præmissen om, at der kan etableres lukkede sociale systemer og bedrives evaluering, hvor den kontrafaktiske situation etableres (tilfældigt eller statistisk).

For klarhedens skyld vælger jeg at opdele de to grupper i ”traditionelle realistiske evaluatorer” og ”fornyende realistiske evaluatorer”. De traditionelle realistiske evaluatorer er blandt andet repræsenteret ved Ray Pawson, mens ”fornyerne” er repræsenteret ved nyere realistiske evalueringsforskere, der forsøger at videreudvikle tilgangen i mere kvantitativ retning. Fornyerne fremfører, at *modelintegration* er muligt og ønskværdigt og kan praktiseres gennem et såkaldt ”realistisk RCT”. Her anvendes programteori kombineret med et randomiseret kontrolleret forsøg. Argumentet fra fornyernes side er, at man ved at gennemføre et realistisk RCT producerer stærk evidens og bevarer RCT-forsøgets stærke interne validitet. Samtidigt får man et kvantificerbart effektmål, og man opnår viden om, hvilke kontekster og mekanismer der har bidraget til at skabe resultatet (Jamal et al., 2015). En programteori udvikles i det ”realistiske RCT” på baggrund af kvalitative data, som sidenhen operationaliseres og kvantificeres, hvorefter der foretages statistiske analyser, der undersøger mediatorer (mekanismer) og moderatorer (kontekster). Herefter er der mulighed for at gennemføre kvalitative interviews til at undersøge, hvorfor indsatsen i givet fald virkede, fordi *forklaringer* ifølge den realistiske evaluering skal bero på kvalitative data.

Fra den anden lejr (de traditionelle realistiske evaluatorer) argumenterer Pawson sammen med en række andre forfattere imidlertid for, at det randomiserede kontrollerede forsøg bygger på den successionistiske kausalitetsforståelse, der er uforenelig med realistisk evaluering (Marchal et al., 2013). Kritikken går endvidere på, at det ”realistiske RCT” ikke i høj nok grad tager højde for kompleksitet og forsøger at ”kontrollere” konteksten væk (Marchal et al., 2013). Det fremhæves blandt andet af næsten samme forfatterkreds, at alt andet lige forestillingen ikke formår at vise, på hvilken måde variablerne *samvirker* om at skabe forandring, samt at menneskelig handlen og mekanismer ikke kan reduceres til statistiske korrelationer (Van Belle et al., 2016). Der vil altid være stiafhængighed, eksistere feed-back mekanismer, og de generative mekanismer vil ikke kunne identificeres gennem statistiske analyser. Det vil kræve kvalitative udsagn.

Ud fra kritikernes logik (og i realistisk evaluering generelt) introduceres indsatsen i en kontekst (et dynamisk socialt system), der samvirker om at skabe forandring (resultater) for målgruppen. Argumentet er, at det ikke vil kunne indfanges gennem statistiske analyser. Det ”realistiske RCT” vil ifølge kritikerne ikke være i stand til at åbne den sorte boks. Derfor vil det ikke være muligt at teste disse ud fra en realistisk logik (Van Belle et al., 2016).

Argumentationen er fra traditionalisternes side efter min vurdering på visse punkter en smule inkonsistent og rammer i flere tilfælde ved siden af: Programteori kan opstilles både deduktivt og induktivt, og Pawson fremhæver selv, at programteori og mekanismer kan identificeres gennem den eksisterende litteratur (Pawson, 2006). Pawson med flere har givetvis ret i, at mekanismerne ikke vil kunne *identificeres* gennem statistiske analyser. Det realistiske RCT har imidlertid allerede gennemgået en ”kvalitativ” fase, hvor mekanismerne er blevet identificeret gennem konsultation af den eksisterende litteratur eller gennem interviews. Herefter vil relevante indikatorer (variabler) for mekanismerne kunne inkluderes i f.eks. et spørgeskema. Gennem statistiske analyser vil man naturligvis ikke kunne iagttage, hvordan mekanismerne først ændrer deltagernes tankegang og ræsonnementer og sidenhen påvirker deres aktive handlinger. Men dette vil heller ikke være muligt gennem kvalitativ empiri, da mekanismerne er ”usynlige” og eksisterer under overfladen.

Pawson med flere har givetvis en pointe i forhold til, at CMO-konfigurationerne ikke vil kunne testes ud fra en realistisk logik i det ”realistiske RCT”. Som vi husker, har man en temmelig bred forståelse af kontekster i realistisk evaluering. Kontekster kan indbefatte karakteristika ved individet (f.eks. køn, alder, uddannelse osv.) – altså mere eller mindre klassiske kontrolvariabler i statistiske analyser (kontekster på individniveau). Gennem f.eks. lineære regressionsanalyser vil ”kontekst-variablerne” blive holdt konstante, mens man vil finde den ”isolerede effekt” af ”mekanisme-variablerne”. Kontekstens indflydelse vil derfor blive kontrolleret væk, og det vil ikke blive undersøgt, hvilken indflydelse konteksten har på, om mekanismerne aktiveres. Det betyder dog ikke, at kontekstens betydning behøver at blive ignoreret i statistiske analyser. Man kan analysere interaktionseffekter, eller analyserne kan stratificeres på de forskellige ”kontekst-variabler”. Såfremt programteoriens ”mekanisme” (f.eks. tillid) operationaliseres til spørgsmål i et spørgeskema, vil man gennem regressionsanalyser på datamaterialet kunne anvende ”tillid” som afhængig variabel i analysen med henblik på at fremanalysere, hvilke kontekster på individniveau (baggrundsvARIABLE og øvrige kendetegn ved individet) der påvirker, hvor kraftigt mekanismen (tillid) aktiveres. Statistiske metoder indeholder således muligheder for, at kontekstens betydning også kan fremanalyses – såfremt datamaterialet tillader det. Derimod kommer de statistiske analyser hurtigere til kort, når det gælder kontekster på interpersonelle niveau, på det institutionelle niveau og på makro-niveau (Pawson, 2013; 37). Det er betydeligt sværere at inkludere relevante variabler fra kontekstuelle niveauer over individniveau.

I et forsvar for det realistiske RCT argumenterer de fornyende realistiske evaluatore for, at der ikke er et noget immanent ”positivistisk” forbundet med at gennemføre et RCT-forsøg, som ”traditionalisterne” hævder. Som påpeget af Bonell et al. (2013) medfører metoder ikke bestemte videnskabsteoretiske antagelser – det væsentlige er, hvilke videnskabsteoretiske antagelser forskeren bringer med sig. De argumenterer

videre for, at det er selvmodsigende, at man i realistisk evaluering hylder metodepluralitet, mens man samtidigt hårdnakket afviser, at eksperimentelle eller kvasi-eksperimentelle metoder kan anvendes. Især når det samtidigt accepteres, at naturlige eksperimenter og sammenligninger på tværs af grupper og kontekster kan anvendes.

På spørgsmålet om, hvorvidt effektevaluering og realistisk evaluering kan integreres til at undersøge samme effektspørgsmål, varierer svarene. Pawson et al. ville mene, at det ikke er acceptabelt, fordi effektmåling ifølge Pawson bygger på et (implicit) positivistisk videnskabsideal. Klassisk effektevaluering bygger på etableringen af den kontrafaktiske situation og en deraf følgende successionistisk kausalitetsforståelse, hvilket som udgangspunkt ikke er anvendeligt i realistisk evaluering. Evaluering bør bygge på generativ kausalteori, som også indebærer *forklaring* af, hvorfor resultatet indtræffer. Det medfører, at hverken integration eller sekventiel anvendelse er acceptabelt ud fra "traditionalisternes" opfattelse. Integreres modellerne, vil det føre til en uhensigtsmæssig *sammenfjydning* af kausalitetsforståelserne (Jagosh, 2017).

Bonell et al. (2013) vil imidlertid argumentere for, at realistisk evaluering hylder metode- og datapluralitet, og at det derfor også burde være fuldt ud acceptabelt at anvende de klassiske metoder til effektevaluering som en del af en realistisk evaluering. De holder fast i, at videnskabsteori er vigtigt, men at eksperimentlogikken er forenelig med en realistisk videnskabsteoretisk forståelse. Gennem operationalisering og kvantificering af mekanismerne ud fra kvalitative data vil man kunne bibeholde den generative kausalitetsforståelse, men samtidigt anvende statistiske metoder til at teste, om de har betydning for resultatet i praksis. Kontekstens betydning vil dog i det realistiske RCT af designmæssige grunde blive nedprioriteret (der vil ikke på forhånd kunne inkluderes variabler for alle væsentlige kontekster). Ud fra "fornyernes" betragtning er evalueringsmodellerne således integrerbare og vil ligeledes kunne anvendes sekventielt.

## MULIGE PRAKTISKE IMPLIKATIONER

Hvorvidt det anses som forsvarligt at *integrere* klassisk effektmåling med realistisk evaluering, afhænger naturligvis af, hvem man spørger. Det samme gør sig gældende for sekventiel anvendelse af modellerne.

Det overraskende er, at de fornyende realistiske evaluatore på nogle punkter har mere til fælles med effektevaluatorerne, mens de på andre punkter har mere tilfælles med de traditionelle realistiske evaluatore. Realistisk evaluering er fra flere sider blevet fremhævet som en brobygger mellem både positivisme og socialkonstruktivisme og mellem proces- og effektevaluering (Dahler-Larsen, 2013; Pawson & Tilley 1997). Realistisk evaluering er tydeligvis en brobygger mht. at forbinde proces og "effekt". Det viser sig imidlertid, at traditionelle realistiske evaluatore snarere bygger barrikader frem for broer, når vi retter blikket mod *modelintegration* og *sekventiel anvendelse*, trods den realistiske evaluerings selverklærede hyldest til empiri og metodepluralitet.

De centrale pointer og synspunkter fra den ovenstående diskussion er gengivet i nedenstående Tabel 2 ud fra min egen fortolkning af positionerne:

**Tabel 2: Evalueringspositionernes betragtninger på integration og sekventiel anvendelse af data og evalueringsmodeller.**

	Klassiske effektevaluatorer	Klassiske realistiske evaluatorer	Fornyende realistiske Evaluatorer
Kan kvalitativ og kvantitativ empiri <i>integreres</i> i samme analyse til at besvare <i>effektspørgsmålet</i> ?	Nej	Ja	Ja
Kan kvalitativ og kvantitativ empiri anvendes <i>sekventielt</i> til at undersøge <i>effektspørgsmålet</i> ?	Nej	Ja	Ja
Kan kvalitativ og kvantitativ empiri anvendes <i>sekventielt</i> til at besvare <i>forskellige</i> spørgsmål?	Ja	Ja	Ja
Kan effektevaluering og realistisk evaluering <i>integreres</i> til en hybrid?	Nej	Nej	Ja
Kan effektevaluering og realistisk evaluering anvendes <i>sekventielt</i> til at undersøge <i>effektspørgsmålet</i> ?	Nej	Nej	Ja
Kan effektevaluering og realistisk evaluering anvendes <i>sekventielt</i> til at undersøge henholdsvis effekter og producere forklaringer?	Ja	Nej	Ja

Som tabellen illustrerer, er der en række stridspunkter mellem de tre positioner, men også en række punkter, hvor de er enige. Alle tre positioner erkender, at kvalitativ og kvantitativ empiri vil kunne anvendes sekventielt til at besvare *forskellige* spørgsmål. Effektevaluatorerne vil imidlertid afvise, at kvalitativ empiri vil kunne anvendes til at besvare effektspørgsmålet, fordi den kontrafaktiske situation ikke vil kunne etableres på et tilstrækkeligt grundlag gennem kvalitativ empiri.

Kun de fornyende realistiske evaluatorer vil gå så langt som til at acceptere, at modellerne integreres til at udgøre en egentlig hybrid. Både effektevaluatorerne og de klassiske realistiske evaluatorer vil indvende, at de to kausalitetsforståelser er uforenelige. Heroverfor står de fornyende realistiske evaluatorer, der mener, at kausalitetsforståelserne bør anses som supplerende i stedet for konkurrerende. Argumentet er fra denne side også, at denne form for triangulering vil føre til mere valide konklusioner om årsag, effekt og virkning.

Til gengæld kan der spores en vis grad af enighed om sekventiel anvendelse af modellerne mellem effektevaluatorerne og de fornyende realistiske evaluatorer. Som berørt i afsnittet om effektevaluering, testes hypoteser ofte i forbindelse med RCT-studier. Derfor vil effektevaluatorerne vedgå, at realistisk evaluering kan anvendes som et middel til at producere disse hypoteser og give *provisoriske* forklaringer på, hvorfor en indsats vil virke. De bør blot sidenhen testes kvantitativt i forbindelse med en effektevaluering. De fornyende realistiske evaluatorer er enige i, at effektmålingen kan anvendes til at estimere effekterne af en indsats, og at en sekventiel anvendelse af realistisk evaluering kan bidrage med forklaringer på, hvorfor (og for hvem) indsatsen har virket. De fornyende realistiske evaluatorer vil dog snarere være fortalere for



integreret anvendelse, fordi der i designet af effektevalueringen vil kunne indgå programteoretiske overvejelser, som kan udsættes for kvantitative tests.

Hvad angår den sekventielle anvendelse af effektevaluering og realistisk evaluering, er de klassiske realistiske evaluators igen de mest konservative. Indsigter fra effektmåling afvises notorisk på et videnskabsteoretisk grundlag med henvisning til den succesionistiske kausalitetsforståelse og etableringen af den kontrafaktiske situation.

Ud fra ovenstående præsentation og diskussion af positionerne bør det være tydeligt, at såfremt effektevaluering og realistisk evaluering skal kombineres, er den sekventielle anvendelse den *mindst* omstridte anvendelsesform. To ud af tre positioner kan acceptere dette.

Retter vi nu blikket mod spørgsmålet fra artiklens titel ”Er realistisk evaluering et supplement eller et alternativ til effektevaluering?”, burde det være tydeligt, at de tre positioner har forskellige syn på dette.

Effektevaluators vil mene, at der er tale om et alternativ, om end et meget dårligt alternativ. Det kan samtidigt være et supplement, der kan anvendes til at generere nye hypoteser, der sidenhen kan efterprøves. De klassiske realistiske evaluators vil mene, at realistisk evaluering udgør et kvalificeret alternativ til klassisk effektmåling, fordi der produceres kontekstnær og direkte anvendelsesorienteret viden. Realistisk evaluering skal ud fra denne position dog ikke anses som et supplement til effektmåling, fordi det vil kompromittere den generative kausalitetsforståelse. Endelig vil de fornyende realistiske evaluators mene, at realistisk evaluering kan være både et supplement og et alternativ til effektmåling. Det vil blot afhænge af, om anvendelsen sker sekventielt eller integreret.

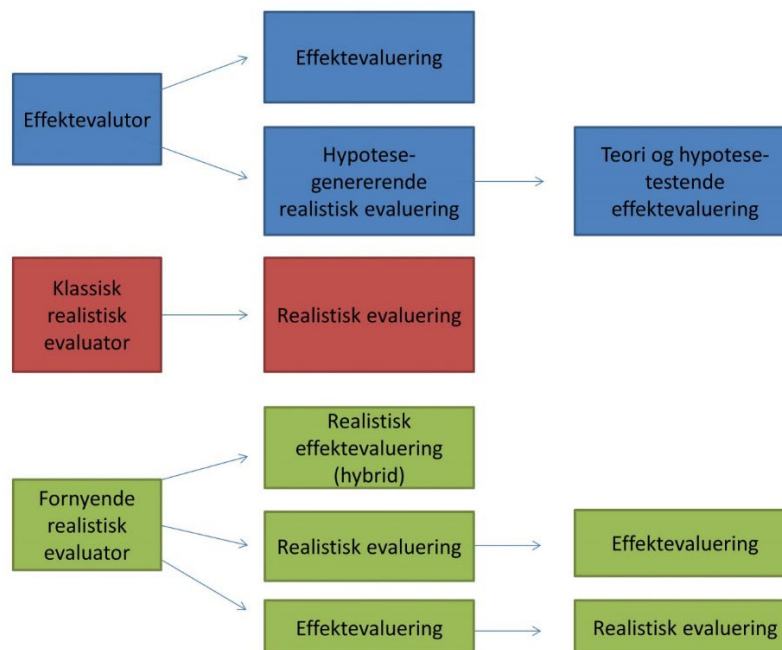
Overvejer man som evaluator at anvende både realistisk evaluering og effektmåling til at undersøge en indsats virkninger, kan denne artikel bruges som et opslagsværk, der kan være en hjælp til at udarbejde et evalueringsdesign. Det kan have følgende praktiske implikationer for evalueringsarbejdet, der er illustreret i Figur1 nedenfor.

Figurens første fase vil bestå i at overveje, hvilken af de tre evalueringspositioner, man tilhører, eller vil arbejde ud fra.

Tilhører man kategorien af effektevaluators, vil man naturligvis kunne gennemføre en klassisk effektevaluering. Søger man også forklaringer på, hvorfor, for hvem, og under hvilke omstændigheder den pågældende indsats virker, vil man som evaluator kunne gennemføre en indledende hypotesegenererende realistisk evaluering. Herefter vil disse hypoteser kunne testes i en teori- og hypotesetestende effektevaluering. Her vil der være tale om *sekventiel* anvendelse af evalueringsmodellerne.

Tilhører man i stedet kategorien af klassiske realistiske evaluators, efterlades man ikke med mange valg. Her vil den eneste mulighed være at gennemføre en traditionel realistisk evaluering, hvor outcomes (effekter) undersøges gennem inter- eller intra programvariation.

**Figur 1: Designforslag til kombination af effektevaluering og realistisk evaluering.**



Er man i stedet tættest på de fornyende realistiske evaluatore, giver det flere designmæssige muligheder. Første mulighed er at *integrere* modellerne til en hybrid, som kan kaldes en ”realistisk effektevaluering”. Det væsentligste i dette design er, at modellerne anvendelse ikke er adskilt i tid, og at ingen af resultaterne fra én evalueringsmodel har forrang for den anden. Der er mange muligheder i designfasen, men her skal blot nævnes ét konkret eksempel. Evalueringen kan designes således, at der gennemføres et RCT-forsøg med tilfældig inddeling i en deltagergruppe og en kontrolgruppe (effektevaluering). Sideløbende hermed kan man med fordel indtænke realistisk evaluering i designfasen af evalueringen og udarbejde en programteori for indsatsen ud fra interviews med indsatsens stake-holders og den eksisterende litteratur, hvor relevante mekanismer og kontekster identificeres. På baggrund af programteorien kan der udarbejdes en spørgeskemaundersøgelse, hvor mekanismerne og konteksterne operationaliseres til konkrete spørgsmål. Spørgeskemaundersøgelsen administreres kun til *deltagergruppen*, og de stiller spørgsmålene, *mens de stadig deltager i indsatsen*. RCT-delen af designet vil muliggøre, at der produceres et kvantificerbart effektmål (deltagergruppens resultater sammenlignes med kontrolgruppens). Efter at indsatsen er afsluttet og effektmålet er produceret, kan evaluator teste programteorien på baggrund af deltagergruppens besvarelser af spørgeskemaet gennem statistiske analyser og intra-programssammenligninger (sammenligning af outcomes for forskellige grupper internt i deltagergruppen). Baseret på programteoretiske antagelser om, hvem indsatsen vil gavne (mest) eller gennem sampling på baggrund af de hypotiserede mekanismer (høj/lav aktivering af mekanismerne/score i spørgeskemaet), vil strategisk udvælgelse af interviewpersoner til kvalitative interviews kunne bidrage til at afdække, *hvorfor* indsatsen i givet fald har virket.

Brugen af den integrerede modelanvendelse medfører, at der produceres to ”effekt-mål”. Det første produceres ved, at deltagergruppens resultater sammenlignes med kontrolgruppens resultater (klassisk effektevaluering). Det andet ”effekt-mål” produceres ved at sammenligne resultater *inden for* deltagergruppen. Intra-program sammenligning anvendes her til at undersøge, *for hvem* det virker internt i deltagergruppen. Det i denne forbindelse væsentligt, at både resultaterne fra effektevalueringen og den klassiske realistiske evaluering tages for pålydende. Et tænkt eksempel kan lyde på, at effektevalueringen når frem til, at der er en svag positiv effekt af indsatsen, mens den realistiske evaluering kan illustrere, at indsatsen virker for nogle deltagergrupper, mens den ikke virker for andre (f.eks. kvinder versus mænd).

Den anden mulighed består i en variant af den *sekventielle* anvendelse. Her vil man f.eks. særskilt kunne gennemføre en realistisk evaluering, hvor programteorien på klassisk vis anvendes og testes gennem intra-programvariation. Herefter kan der gennemføres en separat effektevaluering af indsatsen med etablering af den kontrafaktiske situation gennem inddeling i deltager- og kontrolgruppe. Her er det værd at bemærke, at brugen af evalueringsmodellerne ikke nødvendigvis er adskilt i tid. Der er dog tale om en designmæssig adskillelse, hvor modellerne bruges på deres egne præmisser og gennemføres separat fra hinanden. Denne designtilgang ligner til forveksling det effektevaluatoriske bud på sekventiel anvendelse med gennemførelse af en teori- og hypotesegenerende realistisk evaluering og en efterfølgende teori- og hypotesetestende effektevaluering. Der er dog den afgørende forskel, at den indledende realistiske evaluering i den effektevaluatoriske variant, kun gennemføres som et middel til at generere hypoteser. Dens resultater om effekter (outcomes) tages ikke for pålydende, fordi realistisk evaluering rangerer lavt i evidenshierarkiet. I den ”fornyende” realistiske variant vil effektresultaterne fra begge modeller imidlertid blive taget for pålydende. Indsigterne fra begge modeller vil kunne trianguleres og dermed stå stærkere end ved brug af en enkelt model.

Den tredje og sidste mulighed er et yderligere eksempel på sekventiel anvendelse af modellerne. Her vil effektevalueringen blive gennemført først for at etablere, om og i hvilken grad indsatsen har virket. Herefter vil den realistiske evaluering kunne belyse, for hvem, hvorfor og under hvilke omstændigheder indsatsen i givet fald har virket. Modellerne anvendes her separat på egne præmisser, og indsigterne anvendes til at supplere hinanden. Her ses en klar arbejdsdeling mellem modellerne. Effektevalueringen estimerer effekterne, og den realistiske evaluering producerer forklaringerne på, hvorfor indsatsen har virket (eller ikke virket). Modellerne vil i princippet kunne anvendes sideløbende (men adskilt), men ofte vil modellernes ibrugtagning være adskilt i tid, fordi den realistiske evaluering kobles på som det forklarende element, når vi har viden om indsatsens effekter.

## Konklusion

Artiklen har forsøgt at synliggøre og diskutere de væsentligste realistiske argumenter for og imod *integration* og *sekventiel* anvendelse af realistisk evaluering og effektevaluering. Som først illustreret i tabel 1, adskiller modellerne sig fra hinanden på en række parametre. Undersøgelsesspørgsmålene er vidt forskellige, og modellernes

forståelse af effektbegrebet og kausalitet varierer radikalt. Der er endvidere uenighed om, hvilke data og metoder der kan anvendes til at belyse ”effektspørgsmålet”.

Svaret på, hvorvidt realistisk evaluering udgør et supplement eller et alternativ til effektevaluering, afhænger af, hvilken af de tre undersøgte positioner man spørger. Effektevaluatorene vil mene, at det udgør et alternativ, om end et meget dårligt alternativ. Ud fra denne position kan det dog også udgøre et supplement, der kan anvendes til at generere hypoteser, der sidenhen vil kunne testes. De klassiske realistiske evaluatore vil imidlertid mene, at realistisk evaluering udgør et yderst kvalificeret alternativ til konventionel effektevaluering. Endelig vil de fornyende realistiske evaluatore mene, at realistisk evaluering kan både kan være et supplement og et alternativ til effektmåling.

Ud fra sidstnævnte position vil det mest ønskværdige være, at man som evaluator *integrerer* modellerne til at udgøre en egentlig hybrid. Her vil effekterne kunne estimeres gennem etablering af den kontrafaktiske situation, mens der samtidigt produceres forklaring på, hvorfor, for hvem, og under hvilke omstændigheder indsatsen har virket – blandt andet gennem sammenligning af intra-program variation. Ud fra denne position vil modellerne også kunne anvendes *sekventielt* gennem designmæssige adskillelse af modellerne. Den første mulighed er her, at begge modeller anvendes separat til at undersøge effekterne af indsatsen, hvor man som evaluator efterfølgende triangulerer resultaterne. Den anden mulighed består i, at der laves en klar arbejdsdeling mellem modellerne, hvor effektevalueringen estimerer effekterne, og den realistiske evaluering leverer forklaringer på, hvorfor resultatet er indtruffet.

Såfremt man ønsker at kombinere realistisk evaluering og effektevaluering, vil man som evaluator kunne rådføre sig med denne artikels brødtekst og finde anvendelseparate argumenter for og imod integreret og sekventiel anvendelse af evalueringmodellerne. Rutediagrammet giver ydermere et hurtigt overblik over designmuligheder. Endelig burde det være tydeligt, at *kombinationsformen* har afgørende betydning for, hvilke indsigter man som evaluator når frem til.

## Litteratur

- Andersen, Svend Aage (2007). "Kritisk realisme som perspektiv i socialt arbejde - en introduktion og forskningsoversigt." *Social Skriftserie No. 8*. Aarhus: Den Sociale Højskole.
- Angrist, Joshua D. & Jörn-Steffen Pischke (2015). *Mastering 'Metrics - The Path from Cause to Effect*. New Jersey: Princeton University Press.
- Bhaskar, Roy (1978). *A Realist Theory of Science*. London: Verso.
- Bolton, Derek (2008). "The Epistemology of Randomized, Controlled Trials and Application in Psychiatry", *Philosophy, Psychiatry, & Psychology*, vol. 15 (2), pp. 159–65.
- Bonell, Chris, Adam Fletcher, Matthew Morton, Theo Lorenc & Laurence Moore (2013). "Methods Don't Make Assumptions, Researchers Do: A Response to Marchal Etal", *Social Science and Medicine*, vol. 94, pp. 81-82.
- Bredgaard, Thomas, Ina Drejer & Kristian Nielsen (2016). "Målopfyldelsesevaluering og effektevaluering", pp. 283-306 i Thomas Bredgaard (red.), *Evaluering af offentlig politik og administration*. København: Hans Reitzels Forlag.
- Bredgaard, Thomas, Julia Salado-Rasmussen & Stella Mia Sieling-Monas (2016). "Virkningsevaluering", pp. 331-367 i Thomas Bredgaard (red.), *Evaluering af offentlig politik og administration*. København: Hans Reitzels Forlag.
- Baadsgaard, Kelvin & Hanne Kathrine Krogstrup (2016). "Metaanalyser og evidens", pp. 69-74 i Thomas Bredgaard (red.), *Evaluering af offentlig politik og administration*. København: Hans Reitzels Forlag.
- Christensen, Lillian, Søren Dupont, Tricia Gale & Leif Hansen (1998). *Travle, halvgamle mænd i uddannelsessamfundet - en undersøgelse af nogle 40-60-årige mænds motivation og barrierer i forhold til deltagelse i voksenuddannelse*. København: Roskilde Universitet.
- Collins, Kathleen M. T., Anthony John Onwuegbuzie & Qun G. Jiao (2007). "A Mixed Methods Investigation of Mixed Methods Sampling Designs in Social and Health Science Research." *Journal of Mixed Methods Research*, vol. 1 (3), pp. 267–94.
- Dahler-Larsen, Peter (2013). *Evaluering af projekter - og andre ting, som ikke er ting*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, Peter & Hanne Kathrine Krogstrup (2003). *Nye veje i evaluering*. København: Hans Reitzels Forlag.
- Dalkin, Sonia Michelle, Joanne Greenhalgh, Diana Jones, Bill Cunningham & Monique Lhussier (2015). "What's in a Mechanism? Development of a Key Concept in Realist Evaluation." *Implementation Science*, vol. 10 (1), pp. 1–7.
- Danske Regioner (2016). *Kvalificeret arbejdskraft i hele Danmark*. København: Danske Regioner.
- Funnel, Sue C. & Patricia J. Rogers (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco: John Wiley & Sons.
- Hernán, Miguel A. & James M. Robins (2016). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, Forthcoming.
- Ingemann, Jan Holm (2013). *Videnskabsteori for økonomi, politik og forvaltning*. Frederiksberg: Samfundslitteratur.
- Jagosh, Justin (2017). "Email Correspondence from Member of the RAMESES JISC Online Research Community." <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=RAMESES;20207597.1705>.

- Jamal, Farah, Adam Fletcher, Nichola Shackleton, Diana Elbourne, Russell Viner & Chris Bonell (2015). "The Three Stages of Building and Testing Mid-Level Theories in a Realist RCT: A Theoretical and Methodological Case-Example." *Trials* vol. 16 (466).
- Keller, Hanne Dauer (2004). *Kvinder mellem uddannelse og arbejde. hvidbog om barrierer for kortuddannede kvinders deltagelse i erhvervsfaglig uddannelse*. Institut for Uddannelse, Læring og Filosofi, Aalborg Universitet.
- Klindt, Mads Peter & John Houman (2010). *Barrierer og løftestænger for kortuddannedes opkvalificering*. København: Nationalt Center for Kompetenceudvikling.
- Marchal, Bruno, Gill Westhorp, Geoff Wong, Sara Van Belle, Trisha Greenhalgh, Guy Kegels & Ray Pawson (2013). "Realist RCTs of Complex Interventions - An Oxymoron", *Social Science and Medicine*, vol. 94, pp. 124-128.
- Morgan, Stephen L. & Christopher Winship (2015). *Counterfactuals and Causal Inference*, 2. ed. New York: Cambridge University Press.
- Onwuegbuzie, Anthony John & Nancy L. Leech (2010). "Generalization Practices in Qualitative Research: A Mixed Methods Case Study", *Quality & Quantity* 44 (5): 881-92. doi:10.1007/s11135-009-9241-z.
- Pawson, R. & Nick Tilley (1997). *Realistic Evaluation*. London: SAGE.
- Pawson, Ray (2006). *Evidence-Based Policy - A Realist Perspective*, 73-104. London: SAGE.
- Pawson, Ray (2013). *The Science of Evaluation - A Realist Manifesto*. London: SAGE Publications Ltd.
- Pawson, Ray & Nick Tilley (2004). *Realist Evaluation*. Community Matters. <http://www.communitymatters.com.au/gpage1.html>
- Rowe, Michael & Carmen Oltmann (2016). "Randomised Controlled Trials in Educational Research: Ontological and Epistemological Limitations", *African Journal of Health Professions Education*, vol. 8 (1)
- Salado-Rasmussen, Julia & Thomas Bredgaard (2016). "Virknings syntese - en alternativ tilgang til metaanalyse." *Økonomi & Politik*, vol. 89 (3), pp. 68-82.
- Van Belle, Sara, Geoff Wong, Gill Westhorp, Mark Pearson, Nick Emmel, Ana Manzano & Bruno Marchal (2016). "Can 'Realist' Randomised Controlled Trials be Genuinely Realist?" *Trials* vol. 17 (313), pp. 1-6.
- Vedung, Evert (2009). *Utvärlding I Politik Och Förvaltning*, 3. ed. Lund: Studentlitteratur.
- Vedung, Evert (2016). "Fire Evalueringsbølger", pp. 37-68 i Thomas Bredgaard (red.), *Evaluering af offentlig politik og administration*. København: Hans Reitzels Forlag.
- Westhorp, Gill (2014). *Realist Impact Evaluation - An Introduction*. Overseas Development Institute. A Methods Lab Publication.
- Westhorp, Gill (2017). "Email Correspondence from Member of the RAMESES JISC Online Research Community." <https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1703&L=RAMESES&F=&S=&P=19327>