

Otte opmærksomheds- punkter ved evaluering af politiske reformer

**Morten Balle Hansen, professor, Institut for Statskundskab,
Aalborg Universitet, mbh@dps.aau.dk**

**Karen Nielsen Breidahl, adjunkt, Institut for Statskundskab,
Aalborg Universitet, knb@dps.aau.dk**

**Jan-Eric Furubo, revisjonsråd, Riksrevisionen i Sverige,
jan-eric.furubo@riksrevisionen.se**

**Anne Halvorsen, første amanuensis, Institutt for sosiologi og sosialt arbeid,
Agder Universitet, anne.halvorsen@uia.no¹**

Denne artikel analyserer udfordringer knyttet til evalueringer af større politiske reformer. Den er baseret på en metaevaluering af evalueringen af NAV-reformen i Norge, der både var en betydelig reorganisering af statslig, regional og kommunal forvaltning og en socialpolitisk reform. Metaevalueringer vurderer nytten af en eller flere evalueringer og må ikke forveksles med metaanalyser. Formålet med denne metaevaluering var at udlede nogle generelle principper for organiseringen af evalueringer af store politisk-administrative reformer.

Metaevalueringens ekspertpanel har analyseret sig frem til otte generelle opmærksomhedspunkter knyttet til sådanne evalueringer. Vi præsenterer opmærksomhedspunkterne, diskuterer deres årsager og argumenterer for at Danmark og andre lande kan lære af erfaringerne fra Norge.

I denne artikel undersøger vi, hvilke forhold der kræver særlig opmærksomhed, når opgaven er at evaluere større politisk-administrative reformer.

Vores svar på spørgsmålet bygger empirisk på en metaevaluering af 00ernes største politisk-administrative reform i Norge – NAV-reformen, der blev vedtaget af det norske

storting i 2005. NAV-reformen var både en betydelig reorganisering af statslig, regional og kommunal forvaltning og en socialpolitisk reform. Samtidig med reformens vedtagelse besluttede det norske storting, at den skulle evalueres grundigt. På den baggrund blev en stor forskningsbaseret evaluering organiseret, hvis resultater blev præsenteret på en konference i Oslo i juni 2014. Evalueringen har været genstand for en metaevaluering (Stufflebeam, 2001), og det er resultaterne herfra, der i kondenseret og videreudviklet form præsenteres i denne artikel² (Breidahl, Furubo, Halvorsen og Hansen, 2014). I denne metaevaluering har formålet været at uddrage nogle generelle erfaringer fra evalueringen, der er relevante ved tilrettelæggelsen af fremtidige forskningsbaserede evalueringer af politisk-administrative reformer.

Teoretisk bygger vores svar på en tese om, at politisk-administrative reformer har nogle generelle kendetegn, der medfører, at deres evaluering må forholde sig til og balancere mellem forskellige ofte modsatrettede hensyn, der på forskellig vis kommer til udtryk

som paradokser, dilemmaer, grænser og trade-offs knyttet til evalueringens organisering og anvendelse (Hood og Peters, 2004; Pollitt og Bouckaert, 2011b). Disse kom tydeligt til udtryk i vore empiriske analyser af NAV-reformens evaluering og vi argumenterer teoretisk for, at disse udfordringer har mere generel karakter.

På grundlag af analyserne når vi frem til otte generelle opmærksomhedspunkter knyttet til organiseringen af evalueringer af store politisk-administrative reformer, som beslutningstagere bør være særligt opmærksomme på.

I det følgende indledes med et kort teori-afsnit, hvor fænomenet politisk-administrative reformer afgrænses, der redegøres for den tidligere meget begrænsede forskning på området og analysens centrale begreber klargøres. Herefter redegøres der kort for meta-evalueringens metoder, hvorefter analysens resultater præsenteres i form af otte opmærksomhedspunkter. I forlængelse heraf afrundes artiklen med en kortfattet konklusion.

Teori og tidligere forskning i evaluering af reformer

«... while the fields of public-sector reform and of evaluation have each brought forth a huge body of literature and research, these two realms have been largely treated as separate entities» (Wollmann, 2003: 1).

Den internationale forskningsbaserede viden om evalueringer af større politisk-administrative reformer er begrænset. Som påpeget i en anden artikel i dette temanummer (Breidahl et al., 2015) indikerer den sparsomme litteratur, der findes, at politisk-administrative reformer sjældent evalueres (Pollitt, 1995; Wollmann, 2003) og at de, i de tilfælde hvor de evalueres, typisk evalueres internt som tilfældet er med den danske kommunalreform. Selvom der har været talrige større og mindre

politisk-administrative reformer i de seneste årtier (Pollitt og Bouckaert, 2011a; Wollmann, 2003), så har egentligt forskningsbaserede evalueringer organiseret af et forskningsråd været fraværende og det er således en sjælden case vi analyserer her.

Kendetegn ved politisk-administrative reformer

Store politisk-administrative reformer er kendetegnet ved at berøre mange mennesker, at have flere ofte svært forenelige målsætninger, at udvikle og forandre sig over tid og at være irreversible i den forstand, at de ikke eller vanskeligt kan føres tilbage til udgangspunktet. De politisk-administrative reformer som er gennemført i de seneste årtier i en lang række lande har involveret paradokser, trade-offs og dilemmaer (Hood og Peters, 2004; Pollitt og Bouckaert, 2011b), hvis præcise betydning har varieret og været kontekstafhængig.

Mange reformer besluttet eller formes i et grænseland mellem politik og forvaltning men også mellem institutioner indenfor såvel som udenfor staten. Indenfor de enkelte stater foretages der beslutninger, som har politisk betydning, og som sætter rammerne for politikernes muligheder for at agere og som ikke påvirkes af de folkevalgte politikere.

Målsætningerne er ofte flertydige, og mange gange er det uklart, hvad der beskrives som de faktiske mål. Det kan bl.a. skyldes, at de forudgående politiske kompromisser har haft fokus på formuleringen af målene (March og Olsen, 1983). Hvordan reformen skal præsenteres med henblik på mål og motiv, har ofte været mere kontroversielt end reformens saglige indhold (Baier, March og Saetren, 1986). De officielle målsætninger afspejler dermed ofte ufuldstændigt reformens bagvedliggende ambitioner.

Fordi politisk administrative reformer er komplekse og påvirker mange mennesker og

policyområder, er effekterne i reglen vanskelige at forudsige, og de vigtigste effekter kan nemt være udenfor de officielle målsætninger og målområder (Wollmann, 2003).

Store reformer præges også ofte af at være irreversible. Man kan nok justere reformen, men man kan ikke gå tilbage til det, man gjorde før reformen, uden meget store politiske såvel som økonomiske omkostninger.

Reformens kompleksitet og omfang rejser nogle udfordringer for evaluatoren. For det første vil reformens faktiske karakter ofte blive omformet i implementeringsprocessen og selve evalueringsobjektet er derfor svært at indkredse. For det andet er reformen blot én ud af mange faktorer, som påvirker målopfyldelsen. Ikke bare en række andre statslige foranstaltninger påvirker målopfyldelsen, men også en række andre internationale og indenlandske faktorer. For det tredje kan vi tale om tidsperspektivets dobbelthed: Det tager tid inden effekterne af en stor reform bliver tydelige, men jo længere tid der går, desto sværere bliver det at afdække, hvilke forandringer der skyldes reformen, og hvilken betydning andre faktorer har (Hansen, 2010). Vi kan, for det fjerde, notere, at opfattelsen af reformens forskellige mål og deres indbyrdes relation ikke er indlysende. Og afslutningsvis, den kontekst, som evalueringen kan være tiltænkt at anvendes i, adskiller sig ofte fra den kontekst, som evalueringen blev gennemført under. Med andre ord har »virkeligheden« ændret sig undervejs.

Disse faktorer, og andre, er med til at tydeliggøre, at der er betydelige begrænsninger i den videnanvendelse, som evalueringer af større politiske reformer kan bidrage med. Det gælder særligt spørgsmål som »hvordan gik det?«, »hvad blev resultatet?«, »blev målene opnået?« etc.

Opmærksomhedspunkter, paradokser, dilemmaer, trade-offs og grænser

I det efterfølgende analyseafsnit gennemgås de opmærksomhedspunkter, som ekspertpanelet har fundet frem til i forbindelse med metaevalueringen af NAV-reformen.

Med opmærksomhedspunkter mener vi nogle kritiske punkter, som de centrale aktører i evalueringen – opdragsgiver, evaluator og de der evalueres – må være opmærksomme på, forholde sig til og træffe beslutninger om. Opmærksomhedspunkter i relation til reformer og evalueringer af disse udspringer ofte af en balancegang mellem modsatrettede, men legitime hensyn (Margetts og Hood, 2010; Pollitt og Bouckaert, 2011b). Der kan være tale om mere eller mindre stejle *trade-offs*, hvor opprioriteringen af et hensyn nødvendigvis må betyde nedprioriteringen af et andet. Der kan også være tale om nogle *grænser* – eksempelvis hensynet til uafhængig forskning – som ikke kan overskrides, hvis der skal være tale om en forskningsbaseret evaluering. Der kan også være tale om regulære *dilemmaer*, hvor evalueringen vil blive kritiseret, uanset hvilke resultater den kan fremvise. Nogle gange viser *paradokser* – i betydningen tilsyneladende modsætninger – sig imidlertid netop kun at være tilsyneladende. Hvis de håndteres pragmatisk og løsningsorienteret under hensyntagen til den konkrete kontekst, som evalueringen foregår i, kan de i praksis vise sig at være falske modsætninger.

Datagenerering og analysemetoder

Datagenereringen foregik gennem dokumentanalyser og interviews med centrale aktører i NAV-evalueringen i perioden fra februar til juni 2014.

Med henblik på at få afdækket interessenternes opfattelser og vurderinger af evalueringens organisering, processer, output og anvendelse er der gennemført 12 kvalitative interviews med centrale aktører i evalueringen, som repræsenterer evalueringens centrale in-

teressenter: Arbejds- og socialdepartementet, NAV-direktoratet, Kommunernes centralforbund, Forskningsrådet, styringsgruppen og en række udvalgte forskerne. Disse interview er blevet transskriberet og efterfølgende kodet.

Med henblik på at få afdækket output, sådan som det foreligger i publikationerne, er der gennemført en analyse af de publikationer fra NAV-evalueringen, der var udgivet frem til slutningen af maj 2014 (i alt 87 publikationer). Disse er blevet læst og kodet med henblik på at afdække det indhold, der er relevant for metaevalueringen, bl.a. med fokus på publikationernes fokus, typer af spørgsmål og anvendte metoder.

Resultaterne af de empiriske undersøgelser blev analyseret og drøftet på ekspertpanelets møder i marts, april, maj og juni 2014. På baggrund heraf udviklede ekspertpanelet en række generelle opmærksomhedspunkter, der blev præsenteret på et afsluttende datagenererings seminar for evalueringens centrale interessenter medio juni 2014.

På baggrund af feedback fra seminaret og yderligere drøftelser i ekspertpanelet blev den endelige rapport leveret på dansk 3. juli med de reviderede otte opmærksomhedspunkter. Rapporten blev oversat af det norske forskningsråd og udgivet i september 2014 (Breidahl et al., 2014). Her præsenteres opmærksomhedspunkterne i komprimeret form.

Otte opmærksomhedspunkter ved evalueringen af politisk-administrative reformer

Opmærksomhedspunkt 1: Hvilke fordele og ulemper er der ved en forskningsbaseret evaluering organiseret i Forskningsrådet sammenlignet med alternative modeller?

Evalueringen af NAV-reformen og flere andre politisk-administrative reformer i Norge er organiseret af det norske forskningsråd. Selvom der findes få sammenlignelige studier af organiseringen af evalueringsvirksomheden

(Pollitt, 1995; Wollmann, 2003), så indikerer disse undersøgelser, at denne norske praksis er bemærkelsesværdig og derfor værd at overveje fordele og ulemper ved.

Organiseringen af en evaluering berører et centralt tema i evalueringsforskningen (Dahler-Larsen og Krogstrup, 2001; Vedung, 1997), der især har drøftet fordele og ulemper ved henholdsvis interne og eksterne evalueringer i relation til evalueringernes formål. Idealtypisk sondres der her mellem formålene ansvarliggørelse (kontrol), forbedring og grundlæggende kundskabsudvikling, der i princippet kalder på forskellige organiseringer. Vedung (1997: 117-20) argumenterer således for, at formålene ansvarliggørelse og grundlæggende kundskabsudvikling bedst opnås gennem en form for ekstern evaluering, mens forbedring bedst opnås gennem en organisering i en intern evaluering.

I den norske evaluering var alle tre formål en del af evalueringsopdraget (Breidahl et al., 2014). Metaevalueringen af NAV-reformen indikerer støtte til Vedungs tese i den forstand at evalueringen af NAV-reformen havde størst succes for hvad angår generel vidensudvikling og til dels analyser af reformens effekter. Derimod kneb det med løbende feedback vedrørende forbedringer og justeringer i forbindelse med implementeringen af reformen.

Valget af model for, hvordan evalueringer skal organiseres, kan dog diskuteres i flere perspektiver. Det er ikke altid indlysende, at en specifik »model« skaber bedre forudsætninger for, at evalueringen eksempelvis bidrager til langsigtet kundskabsudvikling end andre modeller.

En anden diskussion vedrører den eksterne evaluators uafhængighed. Undertiden kritiseres eksterne evalueringer fra konsulenthuse for at være bestilt arbejde (Wildavsky, 1996). En ledelse vil ofte have en interesse i at styre resultaterne af en given ekstern eva-

luering, men alene mistanken om at en sådan manipulation har fundet sted kan være undergravende for troværdigheden af resultaterne fra en evaluering. Her må en organisering i Forskningsrådet have den klare fordel at en sådan mistanke fjernes eller i hvert fald minimeres. Det kræver naturligvis, at hensynet til den frie forskning er et ufravigeligt krav, og at Forskningsrådet står som en garant herfor. Det bør også tilføjes, at den legitimitet der må formodes at være knyttet til en forskningsrådevaluering ikke er unik, men kan leveres af andre institutioner.

Analysen ovenfor af forskellige modeller for organisering af evalueringer kan fremstå noget overfladisk. Håbet er imidlertid, at den kan give anledning til mere præcise analyser af fordele og ulemper ved forskellige modeller, end det har været muligt indenfor rammerne af denne metaevaluering. Norge har i en international forstand valgt en usædvanlig og bemærkelsesværdig model ved at bruge forskningsrådet som organisator. Modellen har i hvert fald været effektiv i forhold til det evalueringsformål, der lægger vægt på at bidrage til langsigtet kundskabsudvikling.

Opmærksomhedspunkt 2: Hvilke aktører bør inddrages ved formuleringen af evalueringsopdraget ved evalueringen af store politiske reformer?

Når opdragsgiver formulerer evalueringsopgaven ved store politiske reformer, kan der med fordel inddrages forskellige aktører. Det rejser det principielle spørgsmål om, hvilke aktører som bør/skal formulere opdraget, og dermed også, hvem der skal have mulighed for at påvirke styringen af selve evalueringen. Det kan dreje sig om dels en inddragelse af forskere og eksperter, dels en inddragelse af evalueringens interessenter, der indgår i styringsgruppen.

En inddragelse af forskere og eksperter kan have den fordel, at de kan bidrage med råd og vejledning, hvilket kan bidrage til en præcise-

ring af opdraget. Udfordringen kan imidlertid bestå i, at det i små lande som Norge og de øvrige skandinaviske lande, pga. relativt små forskningsmiljøer indenfor nogle områder, kan være de samme forskere, som efterfølgende vælger at ansøge om evalueringen. De vil derfor alt andet lige have en konkurrence-mæssig fordel i udbudsprocessen i forhold til andre evalueringskonsortier. Vi vurderer, at der er tale om en afvejning mellem hensynet til en adækvat opgaveformulering og hensynet til en lige og fair udbudsproces. Vi finder, at der er tale om et trade-off mellem to legitime hensyn og at den konkrete håndtering må bero på et ledelsesmæssigt skøn. Det er dog efter vores opfattelse ikke hensigtsmæssigt, hvis habilitetsproblemer forhindrer, at den bedst mulige ekspertise inddrages ved formuleringen af evalueringsopgaven. En øget grad af brug af internationale eksperter – f.eks. fra de andre skandinaviske lande – kan være en del af løsningen.

Inddragelsen af evalueringens interessenter, der indgår i styringsgruppen, i selve formuleringen af evalueringsopgaven kan bidrage til at sikre et ejerskab til og en høj grad af involvering i evalueringen fra styringsgruppens side. Interessenternes inddragelse i formuleringen af evalueringsopdraget er også en anledning til at afklare interessenternes forventninger til og forståelse af, hvad en evaluering er og kræver af ressourcer og involvering fra deres side.

En mulighed, som blev foreslået på meta-evalueringsseminaret og som vedrører inddragelse af potentielle aktører mere generelt, er at invitere alle potentielle evaluatore til et åbent seminar, hvor formuleringen af evalueringsopdraget drøftes. I det omfang man selektivt måtte vælge at inddrage nationale eksperter, der også byder på opgaven, bør der tages højde for denne bias ved tilrettelæggelsen af udbudsprocessen. Der bør gives tid, således at konsortier, der ikke har været involveret, har mulighed for at formulere kon-

kurrencefygtige bud, og udbudsmaterialet bør udformes, så det letter arbejdet hermed.

Opmærksomhedspunkt 3: Hvilke medlemmer bør indgå i styringsgruppen?

Ved evalueringen af store politiske reformer er der vigtige hensyn, der peger på, at de væsentligste interessenter, såsom NAV-reformen, er repræsenteret i styringsgruppen. Det gælder både repræsentanter for opdragsgiver, de organisationer som skal implementere reformen, og for evaluaterne (forskerne). Herudover kunne også relevante brugerorganisationer repræsenteres i styringsgruppen. En sådan sammensætning kan bidrage til at fremme vidensdeling og anvendelse af resultater og sikre, at der er et forum, hvor de væsentligste uenigheder og praktiske problemer omkring evalueringen kan blive handlet af og afklaret.

Der har været argumenteret for, at opdragsgiver ikke bør sidde i styringsgruppen, idet dette kan repræsentere et brud med armslængdeprincippet. På samme måde kunne der argumenteres for, at andre interessenter, såsom implementeringsorganisationerne, ikke skulle sidde i styringsgruppen. Det kan give problemer, hvis den legitimitet, der er knyttet til autonomi og høj forskningsmæssig kvalitet, kan anfægtes.

Balancegangen mellem opdragsgivers og andre grupper legitime interesser og hensynet til den frie forskning, bør imidlertid ofte kunne sikres gennem de procedurer for samarbejde mellem styringsgruppe og forskere som etableres.

Igen er der dog tale om en afvejning, som må bero på kendetegn ved den konkrete reform, der skal evalueres. Hvis der er tale om en kraftigt politiseret reform, med høj grad af samfundsmæssig mistillid til opdragsgivers rolle, så kan det komme på tale at formalisere armslængdeprincippet gennem sammensæt-

ningen af styringsgruppen. Eventuelt kunne opdragsgiver i sådanne tilfælde dog have en observatørstatus.

Opmærksomhedspunkt 4: Hvilke principper bør kendetegne relationen mellem styringsgruppe og forskere?

I en forskningsbaseret evaluering organiseret af Forskningsrådet vil den langsigtede kundskabsudvikling ofte naturligt være i højsædet, mens den løbende instrumentelle anvendelse af foreløbige resultater, som kunne være nyttige i forbindelse med tilpasning og implementering af reformen, ofte vil skulle have særlig opmærksomhed.

Her kan styringsgruppen, sammen med projektleder og forskningskoordinatorer, spille en væsentlig og legitim rolle som brobygger og formidler mellem forskning og implementeringsproces. En aktiv styringsgruppe kan endvidere facilitere, at ressourcer og kompetencer, som kan være nyttige for evalueringen, stilles til rådighed, hvilket ofte kan have stor positiv betydning for evalueringsprocessen og evalueringens resultaters kvalitet.

Det må imidlertid sikres, at denne brobygger- og formidlingsrolle ikke bryder med armslængdeprincippet og hensynet til høj forskningsmæssig kvalitet. Der er her nogle grænser for indblanding, som må overholdes, for at de kvaliteter, som kendetegner forskningsbaserede evalueringer, bevares og evalueringens legitimitet sikres.

Det er rimeligt, at opdragsgiver og styringsgruppemedlemmer har mulighed for at få informationer om analyser og konklusioner fra evalueringen, før de offentliggøres. Herved kan opdragsgiver og implementeringsorganisationer være på forkant og hurtigt inddrage relevante resultater i den løbende implementering af reformen.

Opdragsgiver og implementeringsorganisationer vil også i nogle tilfælde sidde inde

med faktuel viden, som kan bidrage til at øge kvaliteten af publikationer fra evalueringen.

Men en sådan indblanding må aldrig få karakter af andet end venlige forslag. Der må ikke kunne stilles spørgsmålstegn ved forskernes uafhængighed. Det gælder generelt og det er særligt vigtigt ved evalueringer af større politiske reformer. Ved sådanne reformer vil der ofte være væsentlige interesser knyttet til at påvirke evalueringen i bestemte retninger. Og her må der aldrig være tvivl om forskningens integritet og uafhængighed af opdragsgivers og andre aktørers eventuelle interesser i at skjule problemer ved implementeringsprocessen.

Opmærksomhedspunkt 5: Hvordan håndteres flere og vanskeligt forenelige målsætninger for evalueringen?

Valget af forskningsrådet som organisator af evalueringen betyder, at opdragsgiver – bevidst eller ubevidst – har valgt at prioritere langsigtet kundskabsudvikling som et væsentligt element i evalueringsopgaven. Ved at varetage og prioritere Forskningsrådets længe og veludviklede tradition for at sikre høj kvalitet har man (formodentlig) påtænkt både at sikre evalueringens kvalitet og legitimitet ved at forskningen skulle bidrage til et armslængdeprincip fra opdragsgiver. Samtidig indgår det også som et delformål at bidrage med løbende anvendelighed af evalueringens resultater i forbindelse med implementeringen.

De udfordringer, som disse forskellige anvendelsesformål skaber, burde have været genstand for en mere omfattende diskussion allerede inden valget af organisationsform blev truffet. På NAV-evalueringens afsluttende konference 12. juni 2014 citerede Peter Dahler-Larsen, lederen af styringsgruppen, således den franske sociolog og antropolog Marcel Mauss for følgende: »I videnskabelige spørgsmål kan man ikke gå for langsomt frem. I praktiske spørgsmål er der ingen tid at spilde.« Selvom citatet sætter tingene på

spidsen, er der en tendens til, at langsigtet kundskabsudvikling stiller andre og delvist modsatrettede krav til evalueringen end de behov, der ligger i en løbende feedback til implementeringsprocessen.

Løbende anvendelse af evalueringresultaterne kan ske på flere niveauer; lige fra en mere overordnet stillingtagen til selve fordelingen af opgaver og ansvar mellem forskellige aktører til forandringer i den måde, de lokale enheder fungerer på. De dele af evalueringen, som gennemføres med henblik på løbende feedback, må derfor målrettes disse forskellige niveauer. Sådant en evaluering kræver en tæt relation mellem dem, som evaluerer, og de forskellige aktører, der skal anvende evalueringresultaterne (Hansen, 2005), og dermed kan autonomi og armslængdeprincippet måske udfordres. Nogle evalueringforskere, som lægger vægt på direkte anvendelse, f.eks. Patton (Patton, 1997), har derfor betonet, at evalueringer i stedet for distance og armslængde kræver nærhed til dem, som løbende skal anvende evalueringen. Og det fremføres endvidere, at evalueringer med sådanne instrumentelle målsætninger må indebære, at en evaluator selv involveres i de virksomheder, som skal evalueres.

Der er her tale om et paradoks – i betydningen tilsyneladende modsætning – når målsætningen både er løbende tilbagemelding og langsigtet kundskabsudvikling. Hvorvidt der er tale om en genuin konflikt mellem kravene til autonomi ved langsigtet kundskabsudvikling og kravene til løbende brugbar feedback, findes der forskellige opfattelser af. Men så forskellige anvendelsesmålsætninger rejser i hvert fald en række udfordringer for evaluator, som man må gøre sig klart. Vi har forståelse for, at det kan være hensigtsmæssigt, ud fra taktiske hensyn i planlægningsfasen at hævde, at forskellige målsætninger skal føres ind i evalueringen.

Evalueringen af NAV-reformen vidner om de

problemer, som opstår, når evalueringer med flere anvendelsesformål skal gennemføres. Selvom det kan lade sig gøre, at en evaluering, som sigter mod langsigtet kundskabsudvikling, også kan udformes på en måde, så den bidrager til løbende instrumentel anvendelse, så vil *denne* form for anvendelse næppe være en forskningsrådsbaseret evaluerings væsentligste styrke. Beslutninger om valget af organisering for kommende evalueringer bør derfor indbefatte overvejelser om, i hvilken udstrækning forskellige typer af organiseringer er befordrende for forskellige anvendelsesmålsætninger.

Opmærksomhedspunkt 6: Hvordan indarbejdes fleksibilitet i evalueringsprocessen?

I teoriafsnittet har vi beskrevet visse kendetegn ved store politiske reformer. Blandt andet blev der peget på, at reformens faktiske karakter i høj grad formes i selve implementeringsprocessen. Dette er i høj grad blevet tydeliggjort ved NAV-reformen. Om det er den intenderede reform eller den faktiske, som evalueres, kommer derfor til at bero på, hvordan evalueringen beslutes og designes. Det er vigtigt, at man efterstræber at evaluere den reform, som faktisk gennemføres.

Der er forskellige ideer hertil fra metaevalueringens empiriske undersøgelse af NAV-reformens evaluering. De vigtigste er vel for det første at forholde sig mere afventende i evalueringens første fase med henblik på at observere, hvor reformen lander i praksis, og for det andet og relateret hertil, at reservere en pulje af midler til uforudsete analyser, der undervejs viser sig at være særligt vigtige. Herved får styringsgruppen, i samspil med projekt- og modullede, en mere væsentlig opgave end ved traditionelle forskningsprojekter med at udvikle evalueringdesignet under indtryk af, hvordan reformen udvikler sig. Ulempen kan være, at man får løbende »politiske spil« om disse midler.

Under alle omstændigheder må selve re-

formens foranderlighed i implementeringsprocessen anses som et vigtigt spørgsmål i evalueringerne. Dette aspekt er der ikke i tilstrækkelig grad blevet taget højde for i beslutningen om og udformningen af NAV-evalueringen. Det er derfor noget, som mere udførligt bør diskuteres ved kommende evalueringer.

Opmærksomhedspunkt 7: Hvor ambitiøs bør man være med henblik på at integrere de forskellige forskningsmoduler i en evaluering?

I Meta-evalueringens empiriske undersøgelse af evalueringen af NAV-reformen har flere forskere rejst spørgsmålstejn ved, hvor ambitiøs man skal være med hensyn til at koordinere og integrere forskellige forskningsmoduler i et større samlet forskningsprojekt, eller om man i højere grad skulle operere med en mere løst koblet projektkoordinering eller flere uafhængige forskningsprojekter. Argumentet imod ambitiøs integration er, at man kan spilde meget tid og mange ressourcer på at forsøge at integrere forskellige forsknings-traditioner, hvor det er meget vanskeligt, hvis ikke umuligt, at bygge bro mellem disse. Argumentet for ambitiøs integration er, at forskellige forskningsprojekter, hvor sammenhængen mellem dem er godt gennemtænkt, kan styrke troværdigheden af det samlede forskningsdesign.

Det er et væsentligt opmærksomhedspunkt, som må afhænge af en vurdering af den konkrete evalueringsopgave. Metaevalueringens empiriske analyser og NAV-evalueringens afsluttende konference indikerer dog, at man her, på trods af en del frustrationer undervejs, er nået et ganske langt stykke med koordineringsstrategien. Procesevalueringerne leverer således nogle troværdige bud på de processer, som kan forklare resultaterne fra effektevalueringerne. Forskellige nyere tendenser indenfor evalueringsteori (Chen, 2014; Hansen og Vedung, 2010; Pawson, 2013) peger endvidere på, hvorledes procesevalueringernes kortlægninger af årsag-virkningssammenhænge

ofte er helt afgørende for at kunne sige noget om, om effekter der afdækkes i effektmålinger skyldes reformen eller andre forhold. Og endnu vigtigere, hvilke kendetegn ved reformen og dens kontekst der forårsager effekterne.

Opmærksomhedspunkt 8: Bør mulighederne for at evaluere være styrende for beslutningen om og implementeringen af reformer?

NAV-evalueringens udfordringer med at afdække effekter aktualiserer diskussionen om, hvorvidt der bør tages hensyn til evalueringen i forbindelse med tilrettelæggelsen af politiske reformer. Bør reformer ideelt organiseres sådan, at der kan laves »ordentlige« effektmålinger organiseret som kvasi-eksperimenter med (ideelt set) randomiserede kontrolgrupper og før-og-efter-målinger? Det er en gammel ide fra 60'erne, og praksis siden da indikerer, at det sjældent – om nogen sinde – lader sig gøre ved store politiske reformer.

For det første, burde der i dag herske enighed om, også mellem de forskellige sider i evidensdebatten, at mulighederne for med sikkerhed at udtale sig om effekter af store politiske reformer er meget begrænset. De forskellige institutter (f.eks. Cochrane og Campbell), som arbejder med disse spørgsmål, har vist, at langt de fleste evalueringer, også dem som er gennemført af veletablerede forskere, ikke bidrager med en sådan kundskab. Den anden side af debatten har bl.a. problematiseret selve spørgsmålet om, hvilken type af viden, som skal regnes for evident og hvad der grundlæggende skal forstås ved evidens. Vi betragter derfor grundlæggende disse standpunkter som urealistiske, ikke bare i synet på, hvordan politik bliver til, men også i synet på de muligheder, som produktionen af kundskab skaber.

For det andet er dette standpunkt problematisk ud fra et demokratisk afsæt, da det indebærer, at politikernes krav skal underordnes videnproduktionen i samfundet og de

krav, der fremføres af evaluatorene. Det rejser spørgsmålet om, hvilken rolle effektmålinger bør have, og hvordan de bør organiseres i forbindelse med store politisk-administrative reformer. Det er ikke et spørgsmål, der kan besvares generelt, men det er et væsentligt opmærksomhedspunkt ved tilrettelæggelsen af evalueringer af politiske reformer.

Konklusion – erfaringerne fra Norge

Metaevalueringens otte opmærksomhedspunkter er som sagt fundet gennem en analyse af en enkelt og sjældent case – evalueringen af NAV-reformen i Norge.

Vi har formuleret dem som spørgsmål, som beslutningstagere vedrørende evalueringen af store politisk-administrative reformer bør forholde sig til, og vi har forsøgt at vise de typiske balancegange og afvejninger, som kendetegner opmærksomhedspunktet.

Det er vores vurdering, at de otte opmærksomhedspunkter har generel relevans og vil være helt essentielle at overveje og tage højde for ved organiseringen af fremtidige evalueringer af politisk-administrative reformer. Vi kan lære meget af erfaringerne fra Norge.

Noter

1. Vi vil gerne takke reviewerne, deltagerne i Center for Organisation, Management og Administration (COMA) på Aalborg Universitet, og Steinar Kristiansen fra Norges forskningsråd for værdifulde kommentarer til tidligere versioner af denne artikel.
2. En metaevaluering må ikke forveksles med metaanalyser. En metaevaluering er slet og ret en analyse af værdien af en eller flere evalueringer (Stufflebeam, 2001), der tænkes anvendt i fremtidige beslutningsprocesser

Litteratur

- Baier, V.E., J.G. March og H. Saetren (1986), »Implementation and ambiguity«, *Scandinavian Journal of Management Studies*, 2(3-4): 197-212.
- Breidahl, K.N., G. Gjelstrup, M.B. Hansen og Hanne

- F. Hansen (2015), »Evaluering af politisk-administrative refoermer: En komparativ analyse af evalueringen af de største reformer i Danmark og Norge«, *Økonomi og Politik*: 34-45.
- Breidahl, K.N., J.-E. Furubo, K. Halvorsen og M.B. Hansen (2014), »Opmærksomhedspunkter ved evaluering af politiske reformer«, Rapport fra metaevalueringen af NAV-reformen i Norge, in M. ekspertpanel, ed., Oslo: Norgesforskningsråd.
- Chen, H.T. (2014), *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective*, Thousand Oaks, California: SAGE Publications.
- Dahler-Larsen, P. og H.K. Krogstrup (2001), *Tendenser i Evaluering*, Odense: Odense Universitetsforlag.
- Hansen, M.B. (2005), *Anvendelse af analyser i beslutningsprocesser*, København: Center for Evaluering og Medicinsk Teknologivurdering, Sundhedsstyrelsen, pp. 1-19.
- Hansen, M.B. (2010), »Marketization and Economic Performance. Competitive Tendering in the Social Sector«, *Public Management Review*, 12(2): 255-74.
- Hansen, M.B. og E. Vedung (2010), »Theory-Based Stakeholder Evaluation«, *American Journal of Evaluation*, 31(3), 295-313.
- Hood, C. og G. Peters (2004), »The middle aging of new public management: Into the age of paradox?«, *Journal of Public Administration Research and Theory*, 14(3): 267-82.
- March, J.G. og J.P. Olsen (1983), »Organizing Political Life – What Administrative Reorganization Tells us About Government«, *American Political Science Review*, 77(2): 281-296.
- Margetts, H., P. 6 og C. Hood, red. (2010), *Paradoxes of modernization: Unintended Consequences of Public Policy Reform*, Oxford: Oxford University Press.
- Patton, M.Q. (1997), *Utilization-focused Evaluation. The New Century Text*, Thousand Oaks: SAGE.
- Pawson, R. (2013), *The science of evaluation: a realist manifesto*, London: SAGE publications.
- Pollitt, C. (1995), »Justification by Works or by Faith? Evaluating the New Public Management«, *Evaluation*, 1(2): 133-54.
- Pollitt, C. og G. Bouckaert (2011a), *Public Management Reform. Comparative Analysis of New Public Management, Governance and the Neo-Weberian State*, Oxford: Oxford University Press.
- Pollitt, C. og G. Bouckaert (2011b), »Trade-Offs, balances, limits, dilemmas, contradictions and paradoxes«, i C. Pollitt og G. Bouckaert, red., *Public Management Reform. Comparative Analysis of New Public Management, Governance and the Neo-Weberian State*, Oxford: Oxford University Press, pp. 182-205.
- Stufflebeam, D.L. (2001), »The metaevaluation imperative«, *American Journal of Evaluation*, 22(2): 183-209.
- Vedung, E. (1997), *Public Policy and Program Evaluation*, New Brunswick: Transaction Publishers.
- Wildavsky, A. (1996), *Speaking Truth to Power. The Art and Craft of Policy Analysis*, New Brunswick: Transaction Publishers.
- Wollmann, H. red. (2003), *Evaluation in public-sector reform: concepts and practice in international perspective*, Cheltenham UK: Edward Elgar Publishing.