

Redaktionelt forord

»Nu må vi lige evaluere«, hører man ofte politiske beslutningstagere sige når der har fundet en forandring sted. Men hvorfor egentlig? Mange steder i den offentlige forvaltning har der over de seneste årtier fundet et decideret amokløb sted, hvad angår måle- og dokumentationsiver. Officielt er begrundelsen den, at vi skal sikre os, at de mange skattekrone bruges mest fornuftigt og effektivt. Men sikrer de mange forskelligartede evalueringer så, at skattekrone bruges bedst muligt, eller kan der identificeres andre grunde til, at evalueringsbølgen har nået hidtil usete højder? Det er blot nogle af de spørgsmål, der tages op i dette temanummer af Økonomi & Politik, der er redigeret af *Peter Dahler-Larsen* og *Mads Kristiansen*.

Måske evalueres der blot af symbolske årsager! Måske gennemføres de mange målinger fordi alle andre gør det samme? Er forklaringen den, at evalueringsområdet er gået hen og blevet »big business«? Der er ingen, der længere har fuld indsigt i, hvor mange penge de mange offentlige evalueringsinstitutter, alle de interne evalueringskontorer i ministerierne og kommunerne, samt underskoven af private konsulentfirmaer, bruger på at gennemføre evalueringer af den offentlige sektor. Evalueringssektoren vokser og vokser, og den skaber måske sin egen efterspørgsel.

Er evalueringer i virkeligheden en politisk syltekrukke og en belejlig undskyldning for at forholde sig passiv i forhold til den virkelige verdens problemer? Og har vi i realiteten adgang til data, der er valide og troværdige i forhold til det, vi ønsker at vide noget om? Spørgsmålene hober sig op, og der dannes sig en konsensus om, at evalueringer er ét politisk værktøj blandt andre, og at evalueringspraksissen får grænserne mellem videnskab og politik til at flyde ud.

I dette nummer har vi samlet en gruppe af 11 forskere og praktikere fra Danmark og udlandet, der forsøger at tage pulsen på den evalueringsbølge, der skyller frem og tilbage i den offentlige sektor. Formålet er at tilvejebringe en helt opdateret viden om, hvad vi ved om evaluering i den offentlige sektor og dens praktiske, politiske og organisatoriske effekter. En opsummering af hovedpointerne i de respektive bidrag gives i den indledende artikel af de to temaredaktører. I forlængelse af temanummeret offentliggøres to sammenhængene review-artikler af *Christen Sørensen*, der på deres egen måde forsøger at evaluere omfanget og indholdet af Produktivitetskommissionens samlede værker.

Martin Marcussen
Ansvarshavende redaktør

Hvad kommer der ud af evalueringsbølgen?

**Peter Dahler-Larsen, professor,
Institut for Statskundskab, Københavns Universitet, pdl@ifs.ku.dk**

**Mads Bøge Kristiansen, post.doc.,
Institut for Statskundskab, Københavns Universitet, mbk@ifs.ku.dk**

Spørgsmålet er presserende. Efter flere års erfaringer med evaluering, akkreditering, auditering, kvalitetsmåling, dokumentation, resultatmåling, evidensbasering, effektstyring og deslige må vi konstatere, at der ikke blev tale om en forbigående modebølge, men om et stadigt mere omsiggribende fænomen, der fæstner sig stadig dybere i den offentlige sektor, om end med lokale variationer såvel som knopskydninger.

Skal man forskningsmæssigt give et bud på, hvad der kommer ud af evalueringsbølgen, er to afklaringer påkrævede. Den ene er en nærmere bestemmelse af, hvad der falder inden for begrebet »evaluering«, den anden en indkredsning af, hvad der menes med »kommer ud af«.

Tager man udgangspunkt i Evert Vedungs efterhånden velkendte definition, omfatter evaluering systematisk (metodisk) vurdering af indsats, organisering, præstationer og udfald i forbindelse med offentlig virksomhed, og det er en vurdering, som tiltænkes at spille en rolle i praktiske handlingssituationer (Vedung, 2009: 20). Disse karakteristika er fælles for en lang række af ovenstående tiltag (således også auditering, akkreditering og resultatmåling), uanset om disse i det daglige benævnes evaluering. Der er næppe brug for helt nye

teorier for hver enkelt af de forskellige komponenter i evalueringsbølgen. Tværtimod er der en stor analytisk fordel ved et overordnet teoretisk blik på evalueringsbølgen, nemlig at man kan forstå variationer og forskydninger i form og organisering inden for samme begrebslige rum.

Blandt de vigtigste af disse variationer er: hvem der tager initiativet til evaluering, hvem udføreren er, hvordan evaluering styres og reguleres, samt hvorvidt evaluering er enkeltstående eller vedvarende/systematisk. Evert Vedung taler om hele fire bølger af evaluering (Vedung, 2010). For det første en videnskabelig bølge, hvor »evalueringsforskning« blev rekvireret efter behov. For det andet en dialogorienteret bølge med inddragelse af brugere og interessenter. For det tredje en managementorienteret bølge, hvor vedvarende måling af resultater søges systematisk integreret i styringsmodeller og ledelsespraksis. For det fjerde evidensbevægelsen, hvor ambitionen er at opsamle viden om god praksis baseret på videnskabelige eksperimenter.

Bølgemodellen betyder hverken, at evalueringsforskning er aflyst, eller at udviklingen går i retning af en ensartet praksis. Tværtimod synes mange modeller at byde sig til under en mangfoldighed af skiftende, måske

modeprægede navne, hvilket blot øger behovet for overbliksskabende teoridannelse.

Hvad kommer der så ud af alt dette? Diskussionen om »anvendelse« eller »konsekvenser« af evaluering har stået på i en årrække. Det har vist sig for snævert blot at betragte en instrumentel, formel, planlagt og lineær indflydelse af evalueringsinformation på beslutninger, selv om en sådan naturligvis kan finde sted. Forestillingen om at alle konsekvenser af evalueringer kan studeres i beslutninger (også kendt som »decisionismen«) er delvis forladt. I stedet udvikles nye begreber til at indfange normsættende, strategiske, symbolske og eventuelt utilsigtede effekter af evaluering. Evaluering former rammebetingelser, mentalitet og tankegange via måling, sammenligning og rangordning. Evaluering skaber synlighed for visse resultater, men lader andre perspektiver henlægge ubelyst. Man kan tale om konstitutive effekter af evaluering, når måling og evaluering definerer det kulturelle og strategiske landskab, som politik og ledelse foregår i. En af de strategiske funktioner af et evalueringssystem kan i den forbindelse være at bane vejen for eller at forskubbe andre former for evaluering og dermed andre belysningsmåder.

Når begreber om konsekvenser af evaluering udvider sig i mange retninger, bliver generelle påstande om brug af evaluering meningsløse eller ideologiske. Det gælder både forestillinger om, at evaluering på forhånd og automatisk fører noget godt med sig – og noget dårligt, for den sags skyld.

Til gengæld kan vi blive bedre til at være opmærksomme på og nøgternt afveje de mange forskellige konsekvenser af evaluering og måling i konkrete situationer. Vi må lære gennem mønstergenkendelse snarere end gennem generalisering.

Samtlige artikler i dette temanummer tilbyder os aktuelle vinkler på sammenhængen

mellem organisering og anvendelse af evaluering.

Der åbnes med Thomas Schwandts artikel, der opdaterer vores viden om brug af videnskabelige resultater i politik. Han peger på, at evidens ikke er en fysisk størrelse og ikke giver sig selv. I stedet understreger han betydningen af interaktion, relationer og debat i oversættelsen af videnskabelige resultater til praktisk og politisk handling. I et politisk system, hvor de færreste beslutningstagere fungerer i enenum, synes dette perspektiv relevant også for evaluering, resultatmåling mm.

Tobias Lindeberg ser nærmere på metaforer for anvendelse af evaluering. Officielle forestillinger ser evaluering som et instrument til beslutninger, hvortil et kritisk modbillede har været evaluering som ritual. Men dette modbillede har også begrænsninger. Tobias Lindeberg foreslår i stedet metaforen forsikring. Evaluering kan nemlig bruges til at mindske den politiske risiko for ulykker og sprede det politiske ansvar for de ulykker, der alligevel måtte ske.

De to næste artikler handler om evaluering af større offentlige reformer. Morten Balle Hansen, Karen N. Briedahl, Jan-Eric Furubo og Anne Halvorsen har meta-evalueret evalueringen af den omfattende NAV-reform i Norge. De udleder otte opmærksomhedspunkter, som vi også i Danmark bør tænke igennem før evaluering af større offentlige reformer.

Derefter sammenligner Morten Balle Hansen, Karen N. Breidahl, Gunnar Gjelstrup og Hanne Foss Hansen systematisk evalueringen af den danske kommunalreform med evalueringen af NAV-reformen i Norge. Selv om der er tale om sammenlignelige reformer i sammenlignelige lande, var evalueringernes organisering og konsekvenser ganske forskellige. Forskerne giver bud på hvorfor.

Herefter søger vi ud i daglig forvaltnings-

praksis og fokuserer på evalueringsstyring i danske kommuner, når Eva Moll Sørensen og Søren Kjær Foged undersøger, hvilke konsekvenser økonomiske krisetider har haft for brugen af mål- og resultatstyring. De finder, at kommunerne forsøger at gøre mål- og resultatstyring til et redskab, der kan bruges til at sætte mål for deres strategiske udvikling og skabe grundlag for prioritering af deres indsatser. Det gælder særligt i de mest økonomisk udfordrede kommuner. Samtidig har kommunerne nedtonet eller decentraliseret deres kvalitets- og indholdsstyring.

Fra evalueringsstyring går vi til en evaluering af et styringstiltag, når Søren Kjær Foged evaluerer effekterne af regnskabssanktionerne for danske kommuner. Evalueringen viser, at kommunerne har forbedret budgetoverholdelsen i tråd med sanktionernes formål, men at kommunerne ikke har øget præcisionen af deres regnskaber. Før regnskabssanktionerne overskred kommunerne typisk deres udgifts-

budget. Nu er det omvendt, men uden at den absolutte afvigelse fra budgettet er signifikant anderledes end før sanktionerne. Implikationerne af disse udgiftsreduktioner kan vise sig i såvel færre kommunale arbejdspladser som faldende serviceniveau. Artiklen vidner om, at forskningsbaserede evalueringer hverken er unyttige eller på vej ud.

Sidst ser Peter Dahler-Larsen på medarbejdernes oplevelse af resultatmålinger i to typer af uddannelsesinstitutioner. Resultatmåling skulle gerne give medarbejderne klarhed over, hvad man skal nå, men der sker i meget uens grad i de to grupper. Forklaringer herpå gives, og de velfærdspolitiske implikationer trækkes op. Det er nemlig ikke i alle henseender, at klarhed er et gode.

Litteratur

Vedung, Evert (2009), *Utvärdering i Politik och Förvaltning*, Lund: Studentlitteratur.

Vedung, Evert (2010), »Four waves of evaluation diffusion«, *Evaluation*, 16(3): 263-77.

djøf

Nyhed fra Djøf Forlag



1. udgave 2014
178 sider
Pris kr. 285,-

Regelstaten

Væksten i danske love og bekendtgørelser 1989-2011

af Mads Leth Felsager Jakobsen & Peter Bjerre Mortensen

Den offentlige debat om regler og regeltilvækst nuanceres med denne udgivelse, der viser, år for år, hvor meget omfanget af love og bekendtgørelser er vokset fra 1989 til 2011. Regelmængden er vokset, men væksten har været langt mere begrænset, end man har hævdet i tidligere undersøgelser. Årsagerne til regelvæksten undersøges, og bogen gør op med en række myter om, hvad der forklarer regelvæksten – og peger på, at denne i vidt omfang kan føres tilbage til de folkevalgte medlemmer af Folketinget.

JURIST- OG ØKONOMFORBUNDETS FORLAG

Køb bogen på www.djoef-forlag.dk

Iagttagelser af anvendelse af videnskabelig evidens i politik og praksis

**Thomas A. Schwandt, professor,
University of Illinois, Urbana-Champaign, USA**

Givet at brugen af videnskabelig evidens i praksis og i politisk beslutningstagen ikke grundlæggende kan forstås som en øvelse i teknisk problemløsning, undersøges, hvad brugen af evidens indebærer, når beslutningstagen sker under virkelighedsnære omstændigheder. Artiklen rummer seks pointer: Evidens er ikke det samme som data. Evidens er ikke »basis« eller »grundlag« for beslutningstagen. Videnskabelig evidens er kun én af adskillige typer af viden, som er nødvendig for politisk beslutningstagning. Brug er ikke det samme som en simpel overføring af viden fra producent til forbruger og dens efterfølgende øjeblikkelige og direkte applikation. Anvendelse er et kollektivt, interaktivt fænomen. For at forstå anvendelse, må vi være mere opmærksomme på situeret kognition, kognitive processer og argumentation

De færreste politikere eller offentlige ledere vil være uenige i, at beslutningstagning kan forbedres, hvis de har adgang til et korpus af relevant videnskabelig evidens såvel som til monitorering eller videnstyringssystemer, der leverer pålidelige og præcise data om inputs, aktiviteter, outputs og outcomes af offentlige service. Det er netop ønsket herom, der er drivkraft for reformbevægelser som »evidensbaseret praksis og politisk beslutningstagning« (EBPP) og »datadreven beslutningstagning« (DDDM). Alligevel ved vi, at ingen ekspertviden, hverken fra forskning, evaluering eller nogen anden form for systematisk videnskabelig undersøgelse, forpligter

til handling. For hvis det var sådan, ville de færreste patienter undlade at overholde deres lægers forskrifter om vægttab, regelmæssig motion og rygestop. De færreste beslutningstagere ville støtte sociale tiltag uden evidens for effektivitet og få, hvis nogen, almindelige borgere ville benægte det empiriske grundlag for global opvarmning og klimaforandringer. Kort sagt, »anvendelse« af videnskabelig eller evalueringsbaseret viden er ikke en naturgiven ting.

Således er det ikke overraskende, at der efterhånden findes en betydelig korpus af litteratur i politisk videnskab og i anvendt forskning på områder som uddannelse og folkesundhed, der behandler problemerne ved at bygge bro mellem videnskab og politisk beslutningstagen og mellem videnskab og praksis, og som fokuserer på emner som overførsel af viden, spredning, formidling, oversættelse, implementering og nyttiggørelse.

Meget af litteraturen om EBPP og DDDM betoner den direkte og øjeblikkelige effekt af videnskabelig evidens på politiske beslutninger. Den anlægger også en klassisk, rationel, objektiv og empirisk betragtning på problemløsning: »Et problem eksisterer; en beslutning må tages; information og forståelse er mangelfuld i forhold til at finde en løsning på

problemet eller vælge mellem alternative løsninger; forskning leverer den manglende viden« (Weiss, 1979: 427). At være rationel og objektiv her betyder mere end at opbygge et argument baseret på fornuft og understøttet af evidens. Det betyder at følge en klar, logisk proces for beslutningstagen, som involverer (a) at definere problemet, (b) at identificere beslutningskriterier, (c) at tildele kriterierne vægte, (d) at udvikle alternativer, (e) at evaluere alternativer baseret på videnskabelig evidens og (f) at vælge det bedste (dvs. det optimale) alternativ. Denne rationelle beslutningsproces er også karakteriseret som »decisionism« – »et begrænset antal politiske aktører beskæftiger sig med at træffe kalkulerede valg mellem klart udtænkte alternativer« (Majone, 1989: 12). Det antages at være en overlegen model i forhold til inkrementelle beslutningsmodeller, som betoner, hvordan politik faktisk skabes i mødet med usikkerheder (se f.eks. Haskins og Baron, 2011).

Beslutninger truffet af politikere og offentlige ledere følger muligvis logikken fra den rationelle model i de situationer, hvor beslutninger er programmerede eller simple (hvilket vil sige godt afgrænsede) samt karakteriseret ved meget stabile omstændigheder (enighed om både mål og midler) og ved klare årsag-og-virkningsforhold (Snowden og Boone, 2008). For det meste er det ikke disse vilkår, der karakteriserer beslutningssituationen (hverken den politiske eller praktiske), når videnskabelig viden er i spil i en given situation.

Givet at brugen af videnskabelig evidens i praksis eller i politisk beslutningstagen ikke grundlæggende kan forstås som en øvelse i teknisk problemløsning, virker det rimeligt at undersøge, hvad brugen af evidens faktisk indebærer, når beslutninger træffes under virkelighedsnære omstændigheder. Denne artikel rummer iagttagelser om begreberne »evidens« og »anvendelse« samlet i seks punkter samt en afrunding. Der fokuseres på fælles

problemstillinger ved beslutninger på såvel politisk niveau som praksisniveau, medmindre andet anføres.

Evidens er ikke det samme som data

Hvis vi skal acceptere og handle ud fra den ide, at evidens spiller eller burde spille en vigtig rolle i beslutningstagning, må vi være afklarede omkring, hvad evidens er, og hvad det ikke er. Det har længe været påpeget, at evidens ikke er synonymt med data. Evidens findes ikke, men laves eller konstrueres. Den konstrueres ved at vælge information fra en tilgængelig bestand af viden og indføre denne i et argument, der sigter mod at overtale et bestemt publikum om rigtigheden af nogle påstande eller konklusioner (Majone, 1989; Phillips, 2007; Prewitt, Schwandt og Straf, 2012). Evidens er resultatet af en forståelse af data som relevante for at understøtte eller afkræfte en bestemt påstand. Ydermere er bedømmelsen af relevans et spørgsmål om, hvordan en given evidens passer sammen med andre antagelser, der indgår i argumentet for påstanden (Cartwright, 2006; 2009).

Antag, at vi er i besiddelse af evidens (E) for, at Hr. P's fingeraftryk blev fundet på mordvåbnet på gerningsstedet. Det er kun i lyset af andre supplerende forudsætninger (A_{1-n}) om omstændighederne (e.g., A_1 : Hr. P havde motiv og mulighed; A_2 : der er andre beviser, der placerer ham på gerningsstedet; A_3 : der er andre beviser, der rejser tvivl om hans alibi for, hvor han var på det tidspunkt, hvor mordet blev begået, etc.), at vi kan sige, at (E) giver en grund til at tro (eller er relevant for) hypotesen (H) om, at Hr. P er skyldig i forbrydelsen. Kort sagt er E evidens for H , hvis man er givet argumenterne A_{1-n} for H .

Evidens er hverken »basis« eller »grundlag« for beslutningstagning

Ideen om, at evidens bogstaveligt talt udgør en basis eller et grundlag for politiske beslutninger eller beslutninger i praksis, er problematisk af flere årsager. Vi ved, at vurderin-

gen af, hvorvidt evidens er så stærk, at den udelukker enhver mulighed for fejl, altid er et spørgsmål om fortolkning. En mangeårig tradition for epistemologisk skepticisme minder os om, at vi mangler afgørende beviser for næsten hver eneste antagelse, vi kan komme i tanke om. Evidens for en påstand er tillige altid foreløbig forstået på den måde, at den kan overgås af bedre evidens (Upshur, 2002). Endelig, selv hvis vi accepterer ideen om, at evidens udgør en sikker basis for beslutningstagning, må vi erkende, at evidens som sådan ikke kan være rigtig eller forkert i nogen absolut forstand. Evidens er en fortolkning af information og data, og vores fortolkning kan være fejlbehæftet. Disse karakteristika ved evidens stemmer overens med vores forståelse af den uperfekte karakter af viden. »Evidens-baseret« antyder således et eksklusivt grundlag for beslutningstagen, som ikke er berettiget. At påstå, at evidens burde spille en vigtig rolle i vores beslutninger er en ting; at påstå, at det er grundlaget for vores handlinger en anden.

Videnskabelige beviser er kun én af adskillige typer af viden, som er nødvendig for politisk beslutningstagen.

Hellere end at benytte sproget fra EBPP gør vi klogt i alternativt at tale om evidens-informeret beslutningstagning. En nyere rapport fra U.S. Research Council argumenterer for, at et endnu bedre beskrivende begreb er *evidens-influeret politik*, fordi det signalerer, hvad der forekommer »når videnskabelige beviser indgår i politiske overvejelser om policymuligheder, og dette forekommer langt oftere end de apolitiske, snævert fokuserede aktiviteter, der karakteriserer evidensbaseret policy« (Prewitt, Schwandt og Straf, 2012: 14). Begge disse alternativer til EBPP-sproget minder os om, at der er mere i beslutningstagning i politik (og praksis) end en ligefrem appel til evidens fra forskning og evalueringsstudier. Politik såvel som praksis involverer uundgåeligt forskellige former for viden. Videnskabelig evidens (inklusiv tekni-

ske analyser og evaluering) er en vigtig type viden, men den er i høj grad upåvirket af det faktum, at den generelle politiske kultur begrænser politiske valgmuligheder, og at den lokale sociopolitiske kultur omkring enhver given praksis former de praktiserendes beslutningstagen. Den politiske arena udgør i særdeleshed en anden vigtig kilde til viden i form af »politisk know-how«. Denne omfatter: »At overveje og regulere strategier eller taktik; at sætte dagsordner; at bestemme prioriteringer; at overtale; at fremme sager; at kommunikere kernebudskaber og ideologisk spin; at sikre eller modvirke politisk ansvarsplacering; at bygge alliancer; og selvfølgelig at forhandle kompromisser. Kontekstuelle vurderinger om det mulige og det ønskelige er indbygget i denne form for viden« (Head, 2008: 5). En tredje type viden er også relevant, nemlig de professionelles praktiske indsigt, herunder viden om den konkrete indsats og dens implementering inden for specifikke praksisområder så som uddannelse og socialt arbejde. Denne sidste type viden identificeres som praksisbaseret evidens og kan inkludere personlig refleksion parret med lokal viden om, hvad der er accepteret og effektivt i praksisfællesskaber og lokalsamfund (f.eks. Sundhedsstyrelsen, 2008).

Brug er ikke det samme som en simpel overføring af viden fra producent til forbruger og dens efterfølgende øjeblikkelige og direkte applikation

En aktuel observation af videnskabelig evidens til brug for informeret offentlig amerikansk politik vedrørende småbørnsundervisning påpeger: »En indsprøjtningsteori om videnskabelig indflydelse på politik, som forudsiger direkte, øjeblikkelig og kraftige effekter, er fejlagtig af flere årsager. For det første er forskning kun én af mange inputs i policy processen. For det andet akkumuleres videnskabelig viden gennem adskillige studier, hvoraf nogle drager forskellige konklusioner. For det tredje er et givent studies anvendelighed i et bestemt valg af politik et

spørgsmål om vurdering. For det fjerde bliver videnskabelig viden oversat, kondenseret, ompakket og genfortolket, før den bruges. For det femte er det mere sandsynligt, at brugen af videnskabelig information, når den forekommer, indebærer retfærdiggørelse (styrkelse af en tidligere udtalelse) snarere end overtalelse (dannelse af en ny holdning)« (Gormley, 2011: 978-9).

Et meget udbredt syn på, hvordan videnskabelig evidens bruges i politisk beslutningstagen, stammer fra den mangeårige ide om brobygning mellem to fællesskaber (Caplan, 1979). De to fællesskaber er forskerne eller videnproducenterne på den ene side og beslutningstagere eller brugerne af viden på den anden side. Hver gruppe har sit sprog, sine værdier, sine normer, sit belønningssystem og sine sociale og professionelle tilhørsforhold. Videnproducenter er optaget af standarder og erkendelsesværdier i den videnskabelige proces i bred forstand. Politikere er handlingsorienterede og praktiske. Forskere arbejder med rationel analyse og disciplineret forskning; politikere arbejder med politik og udisciplinerede problemer. De to fællesskaber har fuldkommen forskellige interesser – evaluators, forskere og policy-analytikere koncentrerer sig om specifikke problemstillinger, tekniske detaljer, dataindsamling og om at foretage forsvarlige undersøgelser, mens politikere er involverede i mange flere problemområder og koncentrerer sig om at vinde valg, promovere partier og mobilisere støtte til problemstillinger (Kingdon, 1984).

Forskellige problemer forbundet med brobygningen mellem de to grupper er blevet identificeret i årenes løb. På udbudssiden er man bekymret over, hvorvidt den producerede viden er relevant, troværdig og tilgængelig for beslutningstagere. Problemer i den forbindelse omfatter utilstrækkelig produktion af policy-relevant forskning muligvis grundet mangel på kapacitet; manglende evne til at relatere resultater fra et specifikt

studie til den sociale og politiske historie bag en policy-ide; ineffektiv kommunikation eller indpakning af eksisterende forskning og evaluering; mangel på tilstrækkelige robuste og kumulative konklusioner; metodologisk svag forskning; manglende overholdelse af deadlines og så videre. På efterspørgselssiden kan problemerne inkludere politikernes manglende evne til at formulere klare og operationelle mål; manglende incitament for dem til at bruge evidens og data; uvidenhed om eksisterende forskning og manglende kapacitet til at absorbere eller få mening ud af forskning; en tendens til anti-intellektualisme og mangel på metodiske færdigheder hos politikerne og så videre. På grund af disse problemer er det bredt accepteret, at anvendelse af videnskabelig evidens i politisk beslutningstagning og praksis ikke kan være et spørgsmål om en ligefrem overførsel eller overlevering af viden fra en gruppe af producenter til en gruppe af brugere (National Board of Health and Welfare, 2013). Hvis gabet mellem de to grupper skal lukkes, kræves strategier og taktikker, der behandler de førnævnte problemer. Den måske mest betydningsfulde indsats i denne henseende har været promoveringen af formidlingsstrategier influeret af diffusionsteori. Den seneste formidlingstænkning har haft fokus på streng sikkerhedskontrol af videnskabelige resultater, der opsummerer styrken af evidens på tværs af studier ved brug af meta-analyser og med oversættelsesmekanismer såsom retningslinjer for *best practice* for hurtigt at overføre videnskabelig viden til de tiltænkte brugere (Green et al., 2009). Begrebet *translation science* (brugt i vidt omfang i sundhedsvidenskaberne) signalerer netop en indsats for hurtigt at sprede resultater fra grundforskning til klinisk praksis. Den samme logik ligger til grund for etablering af evidensbaserede registre over »best practices«, der skal informere policy og praksisvalg. Oversættelsesstrategier inden for undervisning, sundhed, strafferet og social velfærd er mest tydelige i den verdensomspændende »Cochrane Collaboration«, »The National

Registry of Evidence-based Programs and Practices in mental health and substance abuse interventions« og »The Agency for Healthcare Research and Quality« i USA, »The Social Care Institute for Excellence« i Storbritannien og »Nordisk Campbell Center« i København. Disse strategier er kildebase-rede, udbudsstyrede og »skub«-orienterede snarere end »træk«-orienterede. De er envejs, idet de beskæftiger sig med at overføre forskning og evaluering fra produktion til anvendelse (policy og praksis). De er baseret på antagelser om rationalitet, en kultur inden for beslutningstagen, der værdsætter rationel adfærd som beskrevet ovenfor.

Anvendelse er et kollektivt, interaktivt fænomen

Den empiriske evidens for, hvorvidt oversættelsesstrategier faktisk fremmer anvendelse, er i bedste fald blandet (Prewitt, Schwandt og Straf, 2012: 45), og der er en stigende erkendelse af, at det kræver mere end envejsmekanismer at overkomme gabet mellem forskning og policy/praksis. Ideen om en social eller interaktiv model for vidensanvendelse blev allerede introduceret i en skelsættende artikel af Carol Weiss i 1979. Her hedder det, at processen for anvendelse ikke er en lineær rækkefølge fra forskning til beslutning, men et rodet sæt af sammenkoblinger og meldinger frem og tilbage, der trodser pæne diagrammer. Med udgangspunkt i den ide har forskere i vidensanvendelse undersøgt mere gensidige og iterative tilgange til anvendelse af viden i praksis og i politisk beslutningstagen under overskrifter som »interaktive modeller« eller »koblings- og udvekslingsmodeller« (*linkage and exchange*, L&E models) (Conklin m.fl., 2008). Anvendelse ses som en proces, der involverer kommunikation i flere retninger og vedvarende forhandling mellem forskere, politiske beslutningstagere, serviceudbydere og sågar offentligheden. Interaktive perspektiver på anvendelse tildeler såvel individer som organisationer vigtige roller som videnforhandlere og politiske iværksættere

(som har det ene ben i forskning og evaluering og det andet i politik) for dermed at reducere gabet mellem de to grupper (e.g. Lomas, 2007; Meyer, 2010).

L&E-modeller udviser fælles karakteristika: (1) de betragter anvendelse som en dynamisk og uforudsigelig proces, ikke en enkelt begivenhed eller forekomst; (2) de antager ikke, at forsknings- og evalueringsresultater nødvendigvis vil blive brugt på særligt rationelle måder; (3) de understreger den rolle, som forbundne kulturelle, organisatoriske, personlige og politiske faktorer spiller i forhold til hvorvidt og hvordan forskning anvendes; (4) de betonere, at viden fra forskning og evaluering rekonstrueres gennem politikeres og praktikeres mentale modeller og begrebsapparater; og (5) de afviser den opfattelse, at forskning formidles og dernæst adopteres, og mener i stedet at den tilpasses og blandes med andre former for viden afhængigt af anvendelseskonteksten (Nutley, Walter og Davies, 2008: 119).

Mangfoldige strategier er relevante for L&E: ansigt til ansigt-udveksling mellem beslutningstagere og forskere; undervisningssessioner for beslutningstagere; netværker og praksisfællesskaber; faciliterede møder mellem beslutningstagere og forskere; interaktive, tværfaglige workshops samt styregrupper til at integrere synspunkter fra lokale eksperter ind i designet, gennemførelsen af og fortolkningen af forskning (Mitton, 2007). Til dato er forskning i disse strategier for L&E primært sket på sundhedsområdet, hvor interessen for forbindelsen mellem forsknings-evidens og klinisk praksis har været særlig stærk siden fremkomsten af ideen om evidensbaseret beslutningstagning i 1990'erne. Ydermere er meget af forskningen fokuseret på, hvordan L&E har indflydelse på praksissiden (brug af ressourcer, inklusiv videnskabelig evidens, af praktikere) snarere end på formuleringen af servicepolitik (ressourceallokering, servicemønstre) og forvaltnings-

politikken (organisatoriske og finansielle strukturer) (Black, 2001). Studier af L&E ser f.eks. på partnerskaber på det organisatoriske niveau mellem forskningsenheder og en forvaltningsgren (Goering m.fl., 2003); på hvordan puljer på nationalt niveau stimulerer videnskabens integration i politik og praksis via støtte til partnerskaber, fælles møder og efteruddannelsesmuligheder, således at praktikere kan deltage i forskningsarrangementer (Smith og Denis, 2014); og anvendelse af det, der kendes som the Interactive Systems Framework (ISF) inden for forebyggelse og praksis. ISF er baseret på tre forbundne systemer – et, der beskæftiger sig med at udlede forskningsresultater om effektive interventioner inden for et givet område af forebyggelse (f.eks. børnemishandling) og oversætte dem til brugervenlige formater; et andet, der leverer efteruddannelse og støtte til generel kapacitetsopbygning (der øger infrastruktur, evner og motivationen for en organisation til at bruge forskning) såvel som skræddersyet støtte til særlige interventioner; og et tredje til fremme af implementering (Brodkowski m.fl., 2013; Wandersman m.fl., 2008). ISF er Udviklet af Center for Disease Control i USA og er målrettet sponsorer, forskere og praktikere, der er interesserede i at arbejde sammen på forebyggelsesområdet.

For at forstå anvendelse må vi være mere opmærksomme på situeret kognition, kognitive processer og argumentation

Interaktive modeller peger på, at opmærksomhed på både kognitive og sociologiske aspekter af anvendelse er påkrævet. Hvordan politikere og praktikere skaber mening ud af videnskabelig evidens, hænger sammen med, hvordan de kobler mentale processer med handling. Kognitiv teori fremhæver den rolle, vores mentale modeller eller skemaer spiller i forhold til vores kapacitet til at inkorporere ny viden og oplevelser (Brown, 1999). På samme tid fremhæver sociologisk teori, at strukturen og dynamikken i bestemte sociale situationer former, hvad vi gør, og på

den måde er det at finde mening i og anvende videnskabelig evidens i beslutningstagning ikke blot et kognitivt, men også et socialt situeret fænomen.

Dette dobbelte fokus eksemplificeres af Spillane og Miele (2007: 48-9), som hævder, at *hvilken* information der lægges mærke til i et bestemt beslutningsprocesmiljø, *hvorvidt* denne information forstås som evidens, der vedrører et bestemt problem eller politisk krav, og *hvordan* den eventuelt kan bruges enten til at formulere et problem eller udvikle en løsning, afhænger af kognitionen hos de individer, der agerer i det pågældende miljø. Imidlertid er denne proces med at skabe mening »ikke udelukkende en solopræstation. Hvad vi lægger mærke til og hvilken mening, vi får ud af denne information, afhænger af vores situation«. De tilføjer, at en situation »ikke bare er bagtæppe for mening, men et definerende eller konstituerende element i menneskelig aktivitet«. På den måde involverer forståelsen af anvendelse også granskning af, hvordan aspekter af en situation såsom organisatoriske rutiner (mønstre af indbyrdes afhængige handlinger blandt flere aktører som f.eks. møder, høringer, kommissioner etc.) og værktøjer (procedurer for beslutningstagning, databaser, evalueringsmodeller, indikatorer, etc.) hjælper med at konstituere praksisser for anvendelsen af videnskabelig viden.

Studier af kognitive processer er relevant i mange henseender i forhold til brug af viden. For det første er der betydelig litteratur om kritisk tænkning, der lægger vægt på at undgå logiske fejlslutninger i argumentation (Battersby, 2013; Fisher, 2001); for det andet hjælper litteratur om heuristik og bias os med at forstå de kognitive fejl, vi laver, når vi fortolker data (Gigerenzer, 2002; Kahneman, 2011); for det tredje er studier af beslutningstagning under virkelighedsnære omstændigheder nyttige i forhold til at forstå, hvordan beslutningstagere håndterer usikkerhed,

og hvordan beslutninger træffes i grupper (Lipshitz et al., 2001).

Disse studier er specielt relevante for at forstå anvendelse, når vi fokuserer på, hvordan videnskabelig evidens er inkorporeret i og givet mening som led i et policy-argument; en tese, der fremføres i en aktuel U.S. National Research Council Report, *Using Science as Evidence in Public Policy*: »Vedtagelsen af en politik kommer ud af en multi-facetteret social proces, der involver flere aktører, der er engagerede i at samles om, fortolke og debattere, hvilken evidens der er relevant for det politik-valg, man står over for, og derefter måske anvender denne evidens til at påstå, at et bestemt politik-valg er bedre end dets alternativer. Denne proces kan bedst forstås som en form for drøftelse eller praktisk ræsonnement, der sigter mod at overbevise om det gavnlige eller skadelige ved politiske valgmuligheder. Politik-argumenter inkluderer generaliseringer, ekstrapoleringer, antagelser, metaforer, anekdoter og andre elementer, der adskiller sig fra og kan modsige videnskabelige argumenter. Fra dette perspektiv er videnskabelig viden »evidens«, når den bruges til at støtte udtalelser, der er relevante for politiske krav eller påstande. »Evidens« findes ikke kun i den verden, hvor videnskab produceres; den viser sig også i den politiske verden, hvor den fortolkes, gives mening og bruges i argumentation for at overbevise nogen om noget« (Prewitt, Schwandt og Straf, 2012: 4).

At forstå, hvad politiske beslutningstagere accepterer som sammenhængende og overbevisende argumenter, der inkorporerer videnskabelig evidens, vil nødvendigvis involvere studiet af psykologiske processer såvel som institutionel logik og normer, organisatoriske rutiner osv., der leder frem til en sådan accept.

Rammebetingelserne for den oplyste politik

Selvom almindelig sund fornuft fortæller os, at evidens fra forskning og evaluering burde bidrage til oplyst diskussion af beslutninger i politik og i praksis, er vi ikke klar over, hvordan dette bidrag faktisk forekommer. Imidlertid er der nogle aspekter eller træk ved »anvendelse«, der synes at være enighed om.

For det første forklarer simpel overførsel eller formidling af viden til beslutningstagere ikke på tilstrækkelig vis, hvordan anvendelse foregår. Fænomenet er i højere grad noget, der udfolder sig i kollektive handlingssystemer bestående af talrige menneskelige aktører (og ofte også adskillige organisationer/forvaltningsgrene), hver med deres egen kognitive og normative referencerammer formet, blandt andet, af deres deltagelse i praksisser lokaliseret i bestemte organisatoriske omstændigheder, der igen formes af den kultur, de normer, rutiner og værktøjer, der er tilgængelige. Dette er den praktiske, politiske og intellektuelle kontekst for anvendelse.

For det andet er det mere sandsynligt, at anvendelse forekommer, hvis der findes procedurer, systemer, etablerede praksisser, teknologier og normer i en forvaltningsgren eller en organisation, der understøtter og fremmer anvendelse, så vel som incitamentsstrukturer, der belønner dette.

For det tredje er studier af engagement, interaktion og kommunikation mellem evaluatore/forskere, politiske beslutningstagere og praktikere lovende i forhold til både at forstå og forbedre anvendelse.

Litteratur

- Battersby, M. (2013), *Is That a Fact?* Peterborough, Ontario, CA: Broadview Press.
- Black, N. (2001), »Evidence Based policy: Proceed With Care«, *BMJ*, 323: 275-9.
- Brown, G. (1995), »What is involved in learning?« i C. Desforges, red., *An Introduction to Teaching*:

- Psychological Perspectives*, Oxford: Blackwell, pp. 11-33.
- Caplan, N. (1979), »The two communities theory and knowledge utilization«, *American Behavioral Scientist*, 22(3): 459-70.
- Cartwright, N. (2006), »Well-Ordered Science: Evidence for Use«, *Philosophy of Science*, 73: 981-90.
- Cartwright, N. (2009), »Evidence-based Policy: What's to be Done About Relevance?«, *Philosophical Studies*, 143: 127-36.
- Conklin, A. et al. (2008), *Briefing on Linkage and Exchange: Facilitating Diffusion of Innovation in Health Services*, Santa Monica, CA: RAND.
- Danish Board of National Health (2008), *Evidence in Health Promotion and Disease Prevention*, Copenhagen: Danish Board of National Health.
- Fisher, A. (2001), *Critical Thinking: An Introduction*, Cambridge, England: Cambridge University Press.
- Gigerenzer, G. (2002), *Calculated Risks*, New York: Simon & Shuster.
- Goering, P. et al. (2003), »Linkage and Exchange at the Organizational Level: A Model of Collaboration Between Research and Policy«, *Journal of Health Service Research and Policy*, 8 (Suppl 2): 14-9.
- Gormley, W.T., Jr. (2011), »From Science to Policy in Early Childhood Education«, *Science*, 333(6,045): 978-81.
- Green, L.W. et al. (2009), »Diffusion Theory and Knowledge Dissemination, Utilization, and Integration in Public Health«, *Annual Review of Public Health*, 30: 151-74.
- Haskins, R. og J. Baron (2011), *Building the Connection Between Policy and Evidence: The Obama-Evidence-based Initiatives*, London: NE-STA. Available at: www.brookings.edu/research/reports/2011/09/07-evidence-based-policy-haskins
- Head, B. (2008), »Three Lenses of Evidence-Based Policy«, *The Australian Journal of Public Administration*, 67(1): 1-11.
- Kahneman, D. (2011), *Thinking, Fast and Slow*, New York: Farrar, Straus og Giroux.
- Kingdon, J. (1995), *Agendas, Alternatives, and Public Policies* (2. udg.), New York: Harper Collins.
- Lipshitz, R. et al. (2001), »Focus Article: Taking Stock of Naturalistic Decision Making«, *Journal of Behavioral Decision Making*, 14(5): 331-52.
- Lomas, J. (2007), »The In-Between World of Knowledge Brokering«, *BMJ*, 334: 129-32.
- Majone, G. (1989), *Evidence, Argument and Persuasion in the Policy Process*, New Haven, CT: Yale University Press.
- Meyer, M. (2010), »The Rise of the Knowledge Broker«, *Science Communication*, 32: 118-27.
- Mitton, C. et al. (2007), »Knowledge Transfer and Exchange: Review and Synthesis of the Literature«, *Milbank Quarterly*, 85: 729-68.
- Nutley, S., I. Walter og H.T.O. Davies (2007), *Using Evidence: How Research Can Inform Public Services*. Bristol, UK: The Policy Press.
- Prewitt, K., T.A. Schwandt, og M.L. Straf (2012), *Using Science as Evidence in Public Policy*, National Research Council, Committee on the Use of Social Science Knowledge in Public Policy, Washington, DC: The National Academies Press.
- Smits, P.A. og J-L. Denis (2014), »How Research Funding Agencies Support Science Integration into policy and Practice: An International Overview«, *Implementation Science*, 9: 28.
- Snowden, D.J., og M.E. Boone (2007), »A Leader's Framework for Decision Making«, *Harvard Business Review*, Nov.: 69-76.
- Spillane, J.P., og D.B. Miele (2007). »Evidence in practice: A framing of the terrain«, i P.A. Moss, red., *Evidence and Decision Making*, 106th Yearbook of the National Society for the Study of Education, Malden, MA: Blackwell, part I, pp. 46-73.
- Swedish National Board of Health and Welfare (*Socialstyrelsen*) (2013), *How Do Government Agencies Use Evidence?*, www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/19163/2013-6-38.pdf
- Upshur, R.E.G. (2002), »If Not Evidence, Then What? Or Does Medicine Really Need an Evidence Base?«, *Journal of Evaluation in Clinical Practice*, 8: 113-119.
- Wandersman, A. et al. (2008), »Bridging the Gap Between Prevention Research and Practice: The Interactive Systems Framework for Dissemination and Implementation«, *American Journal of Community Psychology*, 41: 171-81.
- Weiss, C. (1979), »The Many Meanings of Research Utilization«, *Public Administration Review*, 39: 426-31.

Evaluering som politisk forsikring

Tobias Høygaard Lindeberg, ph.d.,
prodekan ved det Samfundsvidenskabelige og Pædagogiske Fakultet,
Professionshøjskolen Metropol, thli@phmetropol.dk

Artiklen undersøger, hvorvidt forsikring kan tjene som metafor for at forstå evaluering, og hvilke implikationer det har for forståelsen af den rolle, evaluering spiller i samfundet. Forsikring er valgt som rationalistisk modpol til toneangivende analyser af evaluering som ritual. Artiklen argumenterer for, at evaluering kan forstås som forsikring i den forstand, at evaluering fordeler risikoen ved at træffe beslutninger parallelt til fordelingen af økonomisk risiko gennem forsikring. Der påpeges også væsentlige forskelle mellem forsikring og evaluering, der ikke mindst skyldes en modsætning mellem en markedslogik og en hierarkisk logik. Det betyder, at omkostninger ved evaluering – i modsætning til forsikring – ikke primært bæres af den aktør, der høster fordelene. Desuden er risikodeling i en økonomisk sammenhæng typisk ønskeligt, mens det er mere tvetydigt i en politisk sammenhæng. Evaluering anvendes ofte for at give en mere entydig placering af ansvar. Men analyseres evaluering som forsikring, bliver det tydeligt, at evaluering bidrager til at sprede ansvar.

Ritualet som metafor for evaluering¹

En række forskere har påpeget en stigende anvendelse af evaluering i samfundet (Power, 1997, Pollitt, 1998; Wilkinson og Willmott, 1995). Flere og flere spørgsmål om kvalitet og performance bliver besvaret gennem evaluering. I en dansk sammenhæng har der gennem de senere år været reformer af styring, der tydeliggør, at der ikke har været tale om et forbigående modefænomen. Eksemplerne

omfatter smiley-ordninger til fødevarerbutikker (indført i 2001), omdannelse af Statens Bilinspektion (indført i 2005) til en ordning, hvor alle virksomheder, der lever op til bestemte kriterier, kan godkendes til at gennemføre synet, akkreditering af sygehuse (indført i 2005) og videregående uddannelse (indført i 2007) og endelig styringen af socialtilbud til udsatte børn og unge gennem tilsyn (indført i 2014). Læg til dette de tiltag, der er taget, til evidensbaseret de senere år, hvor resultaterne af en række evalueringer søges anvendt som grundlag for politik og praksis (se Schwandt i dette nummer og Rieper og Hansen, 2007).

Et centralt spørgsmål i forhold til evalueringens rolle i samfundet har været, inden for hvilken ramme udbredelsen af evaluering skal analyseres. Evaluering som ritual har været brugt som metafor i flere toneangivende tekster (Power, 1997 og Dahler-Larsen, 2000). Metaforer gør os i stand til at forstå ét begreb gennem et andet ved at fremhæve ligheder og forskelligheder (Lakoff og Johnson, 1980). Ritualet som metafor accentuerer kritikken af evaluering som en praksis, der ikke alene fører til bedre beslutninger gennem bedre viden, men som også har en række symbolske og mytiske implikationer for de sociale sammenhænge, de indgår i. Michael

Power (1997) så audit som rituel verifikation, som fortrænger fagprofessionelle logikker og/eller skaber unødvendigt bureaukrati. Han sammenfatter sin kritik på følgende måde: »As a result, Society will have more of what it does not really need – certifications and non-opinions which are commonly accepted as useless and which are time-consuming and distracting to produce – and less of what it does – valuable but vulnerable judgements based on the best available knowledge to inform decisions in the face of uncertainty« (Power, 2004: 46). Udbredelsen af audit skaber således en række overfladiske og omkostningstunge papirer, der ikke er værdiskabende i forhold til organisationens brugere. I forhold til ritualmetaforen generaliserede Power Pentlands (1993) analyse af finansiel revision som rituel verifikation til at omfatte al revision (audit). Pentlands analyse, der er baseret på et feltarbejde blandt revisorer, fremhæver, at revision kan opfattes som et ritual, der anvendes til at håndtere en usikker og farlig fremtid gennem skabelse af »a sense of order and safety« (Pentland, 1993: 606). I en dansk sammenhæng har Peter Dahler-Larsen (2000) med bogen »Den rituelle refleksion« kastet et kritisk blik på den øgede brug af evaluering i organisationer. Han konkluderer blandt andet, at »evaluering er gennemspilninger af mytiske temaer i den refleksive modernitet« (Dahler-Larsen, 2000: 163). Evaluering er et ritual, der gør det muligt for organisationer at fremstå som refleksive, videndrevne og forandringsparate.

Ritualet er på den ene side et stærkt modbillede til den rationelle selvfremsstilling, der er hæftet på evaluering. Det rituelle perspektiv behandler mytiske og symbolske sider af det sociale liv. At forklare fænomener som ritual vil sige, at det tilsyneladende formål kun er ét element i en dybere logik. Den sande logik er skjult for den udøvende professionelle. Således kommer det rituelle perspektiv i sig selv til at udgøre en kritik. I Pentlands analyse producerer auditering ikke sikkerhed for, at

regnskabsoplysninger er korrekte i en teknisk forstand. I stedet producerer revisionsprocessens ritualer komfort omkring de ellers usikre og farlige regnskabsoplysninger (Pentland, 1993: 619). På den anden side er der væsentlige udfordringer forbundet med at opfatte evaluering som ritualer – både hvis ritualet forstås bogstaveligt og metaforisk. Opfattes evaluering bogstaveligt som ritual, strækkes ritualbegrebet så vidt, at det meste af vores sociale liv bliver ritualer. Det gælder også for forskningsområdet, hvor det er svært at argumentere for, at eksempelvis konferencer og publiceringsprocedurer ikke har ritualets karakter. Udfordringen ved dette er, at det bliver uklart, hvilken gevinst der er ved at kritisere evaluering som ritual, hvis ikke det er mere rituelt end andre rationelle samfundssfærer som videnskab, politik og økonomi. En mindre fordringsfuld tilgang vil være at anvende ritualet som metafor, der skal fremhæve forhold ved en praksis, som ellers er svære at få blik for. Metaforer bruges i analytisk sammenhæng for at forstå et begreb ved at fremhæve ligheder og forskelligheder i forhold til et andet begreb (Lakoff og Johnson, 1980). Her er det imidlertid centralt at være opmærksom på, hvilke dynamikker denne kritik skaber i relation til kritikkens genstand: evaluatorene og dem, der bruger dem. Som påpeget af Lakoff og Johnson (1980) er metaforer ikke uskyldige. De har formative konsekvenser for tanker og handling. »Evaluering som et ritual« indebærer implicit, at evaluator bliver sammenlignelig med en heksedoktor, og beslutningstagernes tillid til evalueringsrapporter bliver sammenlignelig med »de vilde« overbevisning om, at regndans bringer regn. Med andre ord opstilles et modbillede af evaluering, der stærkt divergerer fra evaluators og evalueringsbrugeres mere rationalistiske beskrivelser af evaluering. På denne måde er ritualmetaforen et debatskabende redskab, som samtidig skaber distance til såvel evaluatorene og beslutningstagere.

Denne artikel vil gennemføre en analyse af

evaluering som forsikring med henblik på at illustrere væsentlige aspekter af evaluativ praksis, som falder uden for de selvbeskrivelser, der kommer fra evalueringens udøver og bruger. Forsikring er valgt, fordi den repræsenterer et modsatrettet alternativ til metaforen om ritualet. At anvende forsikring som metafor giver mulighed for at reflektere over evaluering inden for en ramme, hvor rationalitet og beregnelighed er i fokus. Artiklen søger at besvare spørgsmålet om, i hvilken forstand evaluering kan forstås metaforisk som forsikring, og hvilke implikationer det i givet fald har for forståelsen af den rolle, evaluering spiller i det politiske system i almindelighed og i forhold til offentlig styring i særdeleshed. Sammenligningen med forsikring gør, at det ikke mindst er formelle evalueringssystemer og deres udbredelse, som kommer i fokus.

Indledningsvis vil evaluering og forsikring blive indkredset definitorisk. Herefter vil det blive beskrevet, i hvilken forstand evaluering kan forstås som forsikring, ved at drøfte deling af henholdsvis økonomiske og ansvarsmæssige risici. Dernæst vil sammenligningen blive udvidet ved at se på de risici, forsikring skaber, og søge at finde ligheder med disse i relation til evaluering. Derefter følger en diskussion af forskellene mellem evaluering og forsikring, og endelig er der en konklusion og perspektivering.

Evaluering og forsikring

Evaluering har flere betydninger. På den ene side har det en forholdsvis præcis betydning i litteraturen om programevaluering, hvor det dækker en særlig måde at anvende samfundsvidenskabelige metoder til at besvare spørgsmål om kvalitet og effekt. På den anden side har evaluering en bredere betydning, hvor den dækker over en række langt mere forskelligartede praksisser, der trækker på elementer fra audit og kvalitetskontrol. For eksempel har Danmarks Evalueringsinstitut gennemført både auditeringer og akkredite-

ringer som en form for evaluering, selv om disse i vid udstrækning også er beslægtet med revision/audit.² Evaluering bruges her i bred forstand som sociale teknologier, der giver svar på spørgsmål om kvalitet og performance (se Lindeberg, 2007 og Scriven, 1995). Når den brede forståelse af evaluering vælges, skyldes det, at denne artikel diskuterer implikationer af udbredelsen af evaluering i samfundet, og når evaluering udbredes i samfundet, er det ofte blandingsformer, der implementeres. Akkreditering af videregående uddannelse og hospitaler og tilsyn med sociale tilbud er eksempler på, at væsentlige samfundsområder reguleres gennem evalueringsformer, som ikke har deres udspring i lærebøger om bestemte tilgange til evaluering, men i, hvad der vurderes at være hensigtsmæssigt af de myndigheder, der fastsætter proceduren.

En autoriseret forståelse af forsikring findes i International Encyclopaedia of Behavioural and Social Sciences (herefter benævnt »Encyclopaedia«) samt i en lærebog om forsikring (Harrington og Niehaus, 2003). Disse kilder suppleres af samfundsteoriske vinkler på forsikring, der opfatter forsikring som en måde at håndtere risiko på (Luhmann, 1993; Huber, 2002; Rothstein, Huber og Gaskell, 2006). Encyclopaedia definerer forsikring på følgende måde: »when one party agrees to pay an indemnity to another party in the event of the occurrence of a predefined random event generating a damage for the latter« (Gollier, 2004: 7582). Forsikring kompenserer den forsikrede økonomisk, hvis en utilsigtet hændelse, der er defineret i en police, forekommer. På denne måde fordeler forsikring den økonomiske risiko ved forekomst af en utilsigtet hændelse.

Evaluering som forsikring: risiko og beslutningstagning

Det første krav, der skal opfyldes, før vi kan anvende begrebet forsikring som en brugbar metafor for evaluering, er, at vi kan identi-

ficere en ækvivalent til begrebet økonomisk kompensation og fordeling af den økonomiske risiko. For at kunne gøre dette er det nyttigt at anskue risiko i relation til beslutninger.

Hvis noget går galt i det moderne samfund, er det sandsynligt, at nogen vil blive holdt ansvarlig for den uønskede hændelse. Beslutningstagere kan bebrejdes, hvis beslutningen senere anses for at have skadelige virkninger, og da beslutninger altid er truffet på baggrund af ufuldstændig information, vil der eksistere usikkerhed om virkningerne af en beslutning (Luhmann, 1993). Beslutninger skaber risici. Hvis usikkerheden ved beslutninger kan absorberes, kan risikoen ved at træffe beslutninger håndteres.

Usikkerhedsabsorption finder sted »when inferences are drawn from a body of evidence and the inferences, instead of the evidence itself, are then communicated« (March og Simon citeret fra Seidl, 2003: 19). Usikkerhed absorberes således, når en kompleks situation vurderes, og det er vurderingen af situationen, der kommunikeres. Evaluering kan betragtes som sådanne teknologier til at absorbere usikkerhed og dermed dele risiko. Evaluatoren tager ansvaret for en større eller mindre del af beslutningsgrundlaget, og ved at gøre dette tager evaluatoren en del af ansvaret. Indførelsen af akkreditering af nye uddannelser på det videregående uddannelsesområde kan illustrere dette. Før akkrediteringen blev indført, skete godkendelse af nye uddannelser ved en ministerbeslutning baseret på sagsbehandling i ministeriet. Vurderingen af kvaliteten af den foreslåede uddannelse var ministeriets og dermed ministerens. Efter at akkreditering blev indført, blev ansvaret for at vurdere kvaliteten af uddannelser Akkrediteringsrådets. Rådet træffer beslutninger baseret på en evaluativ proces. På denne måde bliver ansvaret for beslutningen om en ny uddannelse delt mellem et politisk niveau, der har rammesat akkrediteringsinstitutionens arbejde gennem lovgivning, og akkrediteringsinstitutionen,

der gennemfører evalueringer inden for rammerne af denne lovgivning. På samme måde vil en beslutning, der i væsentligt omfang støtter sig til resultatet af en evaluering i forhold til at videreføre eller lukke en bestemt indsats, dele ansvaret for beslutningen med evaluatoren.

Evaluatoren påtager sig en del af ansvaret for en beslutning, nemlig den del af ansvaret, der er relateret til beslutningsgrundlaget. Med engelske begreber transformeres blamability til deniability. Mere generelt kan forsikring beskrives som et middel til at omdanne risiko for forekomsten af en uønsket hændelse til risikoen ved ikke at tage en forsikring (Luhmann, 1993: 46). På samme måde kan evaluering forstås som et middel til at omdanne risikoen for at træffe en dårlig beslutning til risikoen for ikke at benytte evaluering til at skabe et grundlag for beslutningen. For at kunne fungere foretager forsikring en reduktion af verden, idet forsikring håndterer alle fordringer som pengekrav (Huber, 2002: 3). Hvis en forsikringstager mister et job eller et arvestykke, er kompensationen ikke et nyt job eller et nyt arvestykke, men penge. Denne reduktion muliggør forsikring af områder, der ellers ikke ville kunne forsikres. Evaluering kan ikke på samme måde reducere alle spørgsmål til penge. Hvis en beslutning har negative effekter, giver evaluering ikke mulighed for økonomisk kompensation. Når evaluering forstås som forsikring, kan ansvar siges at være det, alle spørgsmål reduceres til. Ligesom forsikringen absorberer økonomiske tab i tilfælde af en utilsigtet hændelse, vil evaluering absorbere ansvar i tilfælde af en utilsigtet hændelse. Det ændrer ikke ved, at det er risikabelt at træffe beslutninger. Men da risikoen forbundet med beslutningsgrundlaget og risikoen for selve beslutningen kan beskrives som forskellige processer, kan ansvaret fordeles mellem forskellige beslutningstagere.

Hverken for forsikringstager eller for forsik-

ringsgiver er verden så simpel, at når blot der er indgået en kontrakt, er problemet løst. For forsikringsgiver opstår der en række risici som følge af forsikringer, og for forsikringstager er spørgsmålet om, hvorvidt en hændelse er dækket af forsikringen, ikke endeligt, før forsikringsgiver anerkender eller tvinges til at anerkende, at det er tilfældet. Det giver mulighed for at udvide brugen af forsikringsmetaforen i relation til evaluering.

Selvproducerede risici: udbredelse af metaforen

Forsikringen absorberer ikke kun risici, forsikring producerer selv nye risici. Disse risici, der opstår som følge af forsikringsgivning, betegnes som interne risici i modsætning til eksterne risici, dvs. risiko for en utilsigtet hændelse, der er dækket af forsikringsordninger (Huber, 2002: 4). Det ikke kun forsikring, der på samme tid håndterer og producerer risici. Det samme gælder for risikoreguleringsredskaber i bredere forstand. Risikoregulering skaber på samfundsniveau risiko for, at risikoreguleringen er forkert skruet sammen. På organisationsniveau opstår der også risici, eksempelvis for, at risikohåndteringen vil identificere kritisable forhold, som skaber et internt behov for at håndtere denne risiko (Rothstein, Huber og Gaskell, 2006; Power, 2007). Der er således grund til at antage, at sådanne selvproducerede interne risici ikke er begrænset til forsikring. I det følgende diskuteres to typer af interne risici inden for forsikring. For det første risici, der skabes ved, at et højt antal forsikringshændelser sker samtidigt, og for det andet risici der skabes ved, at forsikringstageres ændrede adfærd som følge af, at de har tegnet en forsikring.

Forsikring af forsikring

Når et forsikringsselskab accepterer at kompensere en forsikret i tilfælde af en utilsigtet hændelse, er accepten baseret på en beregning af sandsynligheden for, at en given hændelse indtræffer. Men for forsikringsgiver er der en risiko for, at et uventet højt antal hæn-

delser opstår, hvilket vil gøre virksomheden urentabel eller endda medføre insolvens. Et uventet højere antal hændelser kan for eksempel ske ved en tilfældighed eller på grund af, at de forsikrede hændelser skyldes samme fænomen, som det fx er tilfældet ved skader forårsaget af orkaner eller oversvømmelser.

For at øge rentabiliteten og undgå insolvens tegner forsikringsgiver genforsikringer (Huber, 2002: 6-7). Genforsikring er en fordobling af forsikring – en forsikring af forsikring. Genforsikring kan for eksempel bestå i, at genforsikreren forpligter sig til at kompensere forsikringsselskabet for alle krav fra en bestemt type forsikringer (Harrington og Niehaus, 2003: 91, for en mere udførlig beskrivelse se Riley, 1997: 5-24). Uanset form er effekten af genforsikringen, at den økonomiske risiko for utilsigtede hændelser bliver yderligere fordelt.

For forsikringstager betyder genforsikring håndtering af den risiko, den forsikrede påtager sig ved at vælge et forsikringsselskab, der kan vise sig at være insolvent. Genforsikring er en måde at øge sandsynligheden for, at forsikringsselskabet rent faktisk vil være i stand til at honorere løftet om compensation, i fald der indtræffer en forsikret hændelse for den forsikrede.

I forhold til evaluering giver tanken om, at en evaluator går »ansvars-insolvent«, ikke umiddelbart mening, men en evaluator kan miste sin legitimitet og dermed være dårligere til at opsuge risiko ved ansvar. Det gælder for eksempel, hvis evaluators uafhængighed beklækkes, som det skete i en kortlægning af unge- indsatsen, hvor et konsulentfirma blev kritiseret for ikke at være uvildige, men tværtimod lade opdragsgiver skrive konklusionerne (Bjerregaard, 2009). For beslutningstager kan genforsikring ansues som et værn mod risikoen for, at den ansvarlige for en given beslutning vælger en forkert evalueringsmodel.

Ligesom genforsikring fordobler forsikrings-teknologi, er der talrige eksempler på fordobling af evaluering – både i de tilfælde, hvor en evaluering selv evalueres, og i tilfælde, hvor evaluering stiller krav til evaluanden om selvevaluering. Det mest markante eksempel på, at evaluering selv gøres til genstand for evaluering, er kvalitetssikringsområdet. En virksomhed kan vælge at følge ISO 9001-standarden i tilrettelæggelsen af sin interne kvalitetssikring. Denne interne kvalitetssikring kan blive certificeret på baggrund af en evaluering fra et certificeringsselskab. Certificeringsselskabet bliver selv evalueret og akkrediteret af et nationalt akkrediteringsorgan, der selv anerkendes af et regionalt organ på baggrund af peer review. For at fuldende værket anerkendes det regionale akkrediteringsorgan af et globalt akkrediteringsorgan. Det vil sige, at der ud over virksomhedens egen kvalitetssikring er fire niveauer af evaluering af evaluering.³ Eksemplet er ekstremt, men ikke enestående. Da akkreditering af det videregående uddannelsesområde blev etableret i Danmark, var det skrevet ind i lovbemærkningerne, at det danske akkrediteringssystem skulle leve op til de europæiske standarder på området, og dette skulle dokumenteres gennem optagelse i det europæiske register for kvalitetssikringsorganer, hvilket forudsætter en evaluering. Akkrediteringsinstitutionen opnåede denne status i 2010. På samme måde er IKAS, der er ansvarlig for akkreditering på sundhedsområdet, akkrediteret af det internationale akkrediteringsorgan The International Society for Quality in Health Care. Et tredje eksempel på, at evaluering af evaluering bliver en integreret del af evaluativt understøttet styring, er de sociale tilsyn, der blev oprettet i 2014. I lovbemærkninger er det beskrevet, hvordan socialstyrelsen varetager en auditfunktion for at følge praksis i de sociale tilsyn. En anden type genforsikring er de evaluering af evalueringer, der foregår i relation til arbejdet med metastudier. Eksempelvis Campbell Collaboration har udarbejdet standardiserede systemer for vurdering

af effektevalueringer.⁴ Her er det risikoen for at udvikle politik eller praksis på baggrund af dårlige enkeltevalueringer, der håndteres gennem evaluering af evaluering.

Evaluering af evaluering anvendes også til risikospredning af ansvar, når evaluator stiller krav til evaluanden om selvevaluering. Som Micheal Power har beskrevet, kræver auditering, at der eksempelvis er interne evalueringssystemer, der kan gøres til genstand for auditering (Power, 1996). Auditering kan kun fungere, hvis der allerede er evalueringer, som der kan arbejdes med. Gennem kravet om intern evaluering deler evaluator risiko med evaluanden ved at kræve, at denne selv vurderer, i hvilket omfang organisationen lever op til de opstillede kriterier. Interne evalueringer er en central del af akkreditering både på sygehuse og videregående institutioner. Inden for rammen af forsikringsmetaforen kan det siges, at evaluator installerer sig selv som genforsikringsaktør ved at kræve, at kun selvevaluerede evaluander kan evalueres.

Ligesom genforsikring distribuerer økonomisk risiko videre end det enkelte forsknings-selskab, distribuerer evaluering af evaluering ansvar videre end den enkelte evaluator. På denne måde kan risikoen for beslutningsprocessen distribueres og deles i netværk af evaluering og dermed gøre det vanskeligere at tildele ansvar. Anklages en uddannelse for at være af ringe kvalitet, vil ministeren ikke i samme grad kunne klandres for beslutningen om godkendelse af den specifikke uddannelse, idet en del af ansvaret kan forskydes til akkrediteringsinstitutionen. Akkrediteringsinstitutionen kan videre forskyde ansvaret, idet deres procedure lever op til relevante europæiske standarder og i øvrigt baserer sig på selvevaluering foretaget af uddannelsesinstitutionen.

Moralfare

I litteraturen om forsikring opfattes informationsasymmetri som årsag til flere problemer

for forsikringssselskaberne. Informationsasymmetri betyder, at den forsikrede har privilegeret adgang til oplysninger om egne aktiviteter, for eksempel om det ville have været muligt at gøre mere for at undgå en utilsigtet hændelse. Denne type problemer diskuteres under betegnelser såsom moralfare (moral hazard), ugunstig udvælgelse (adverse selection) og bedrageri. Det centrale i denne sammenhæng er, at disse selvproducerede risici skaber omkostninger for forsikringsgiver, som må indregnes i prissætningen på forsikringer. Særligt moralfarebegrebet er relevant i relation til evaluering.

Moralfare er et begreb, der anvendes i forsikringslitteratur til at beskrive »the effect of insurance on the insured's incentive to reduce expected losses« (Harrington og Niehaus, 2003: 183), altså at den forsikrede ændrer adfærd på grund af forsikringen. Moralfare illustreres med et eksempel på en mand, der, netop som han er klar til at gå i seng i sit hjem, indser, at han har efterladt sin computer på sit arbejde. Hvis hans computer er forsikret, vil han være mindre tilbøjelig til at tage tilbage til sit arbejde for at hente den, end hvis den ikke er forsikret (Harrington og Niehaus, 2003: 183).

Nogle undersøgelser tyder på, at et lignende fænomen er fundet inden for evaluering. Evaluering synes at ændre adfærd hos den evaluerede på en utilsigtet måde. Så tidligt som i 1956 konkluderede en forskningsoversigt, at »quantitative performance measurements [...] are seen to have undesirable consequences for over-all organisational performance« (Ridgway, 1956: 247). Nyere studier er nået frem til lignende konklusioner i forhold til negative effekter ved performance measurement (Austin, 1996; Thiel og Leeuw, 2002). Problemet kaldes målforskydning, hvilket vil sige, at organisationen er fokuseret på at opfylde de kriterier, den bliver målt på, snarere end på organisationens egentlige formål, der ofte er beskrevet mere abstrakt.

Ofte er det intentionen, at performancemåling skal være motiverende. Det er ønskeligt at få mere af det, der måles og belønnes. I de tilfælde vil problemet ikke være, at organisationen får mere af det, der måles, men at det, der ikke måles, nedprioriteres mere, end det er ønskeligt. Det er meget få organisationer (om nogen), der kan ønske en bogstavelig efterlevelse af balanced score card-mottoet: What you measure is what you get (Kaplan og Norton, 1992).

I forsikringsteori er moralfare produceret af forsikringsvirksomheden. Ingen forsikring – ingen moralfare. Det er ikke en bieffekt, forsikringsgiver bliver overrasket over forekommer, men en omkostning, der kan indregnes i profitabilitetsberegninger. I en evalueringskontekst kan dette inspirere til at betragte målglidning som en intern selvskabt risiko ved brug af evaluering. Det kan betragtes som en omkostning ved evaluering, i det målglidning vedvarende må holdes i skak gennem for eksempel ledelsesfokus på organisationens egentlige formål. En central forskel mellem evaluering og forsikring er, at evaluerings-systemer ikke nødvendigvis skaber risiko for målglidning hos den ansvarlige for beslutninger, men for aktører andre steder i systemet. Det er sygehusene, ikke sundhedsministeren, der gennem akkrediteringerne får en række krav at leve op til og dermed incitament til målforskydning.

Evaluering er ikke forsikring: marked vs. hierarki

En central forskel mellem evaluering og forsikring er, at forsikring virker på et marked, mens evaluering som forsikring typisk skal virke i hierarkiske relationer.

Omkostninger

Når der tegnes en forsikring, ligger omkostningerne hos forsikringstager, og de er økonomiske. Forsikringspræmien er den afgørende omkostning. For evaluering er der ofte væsentlige omkostninger forbundet med eva-

luering for andre end den beslutningstager, der forsikres. Det skyldes, at evalueringen meget ofte har andre end beslutningstageren som genstand. Det er hospitaler, uddannelsesinstitutioner og andre organisationer, der er genstand for evaluering, mens det politiske niveau forsikres. Eller det er ledelsen i en organisation, der forsikres gennem evaluering, der har de udførende dele af organisationen som objekt.

Det er ikke gratis at være genstand for evaluering. Det kræver typisk, at der afsættes ressourcer til at deltage i interview eller finde data. Baserer evalueringen sig på selvevaluering, kan de væsentligste omkostninger bestå i interne konsulenter, der skal organisere selvevalueringsprocesserne, og ledere og medarbejdere, der skal bidrage med input eller registreringer af alt fra beskrivelser og vurderinger af egen praksis over registrering af data med henblik på, at de kan indgå i evalueringer til deltagelse i workshops om styrker, svagheder, muligheder og trusler. Det betyder, at hvor fordele og omkostninger i forsikring placeres hos den samme aktør, er fordele og omkostninger i relation til evaluering placeret hos forskellige aktører.

Omfordeling af ansvar

Når evaluering anskues som forsikring, bliver det en gevinst, at risikoen ved at træffe beslutninger kan spredes, hvorved ansvar kan deles. Her støder gevinsten i metaforen imidlertid sammen med andre samfundsmæssige hensyn. Spredningen af økonomiske risici i forhold til et ufordelagtigt mønster i forsikrede hændelser er i god tråd med en markedsløik og bidrager til at stabilisere markedet. Men hierarki er en mekanisme til placering af ansvar. Uagtet hvor komplekse beslutningsprocesser i praksis er, medvirker hierarkiet til, at ansvaret for en beslutning kan placeres relativt entydigt hos en person – i politiske systemer i sidste ende en politisk valgt minister eller borgmester. Det er grundlaget for et demokratisk, bureaukratisk system. Det be-

tyder, at i det omfang, evaluering virker som forsikring og med succes fordeler risikoen for beslutningstagning og dermed ansvar, modvirkes en entydig ansvarsplacering.

Når evaluering fremhæves som en mekanisme, der fordeler ansvar mellem beslutningstager og evaluator, vendes diskussionen af evaluering og ansvar på hovedet i forhold til diskussionen om accountability (Munro og Mouritsen, 1996). Evaluering blev fremmet – særligt i 1990'erne – som en ansvarlighedsgørende (accountability-skabende) mekanisme. Det indikerer, at ansvar i relation til evaluering ikke blot er et spørgsmål om ansvarsdeling mellem beslutningstager og evaluator. Det implicerer også, at de aktører i de producerende enheder, der udgør evalueringsgenstanden, stilles til ansvar. Som beskrevet ovenfor viser det sig eksempelvis, når evaluator stiller krav til evaluanden om selvevaluering og dermed skaber en risiko for evaluanden ved dårlig performance. På denne måde forskydes ansvaret ikke blot væk fra beslutningstageren i toppen af hierarkiet til en evaluator, men ofte til de aktører, der udgør evaluanden.

Manglende udbetaling af forsikringssummen

I teorien skal den forsikrede ikke bekymre sig, når forsikringen er tegnet: Hovedparten af risikoen for en uønsket hændelse er opslugt af forsikringen. Men indtræffer hændelsen, så afhænger den faktiske kompensation ikke kun af beslutningen om, at der er tegnet en forsikring, men også af, om forsikringsgiver beslutter eller tvinges til at beslutte (f.eks. ved at forsikringstager klager til en myndighed), at det er en hændelse, der er omfattet af forsikringsbetingelserne.

I praksis vil spørgsmålet om, hvorvidt en evaluering dækker en specifik problemstilling, ofte være mindre entydigt end, om en hændelse er omfattet af forsikring. Finansiell revision har leveret en række spektakulære sager, hvor investorer har lagt revisorvurderinger til

grund for investeringer i virksomheder, som enten har vist sig at have en uhæderlig forretningspraksis (Enron-sagen kunne være et eksempel på dette) eller en mindre robust forretning, end investor havde fået indtryk af (O.W. Bunker kan vise sig at være et nytligt eksempel på dette). Et eksempel fra en anden form for evaluering er den kritik, der ramte social- og integrationsminister Karen Hækkerup, da hun anbefalede fire metoder på baggrund af, at de var evidensbaserede.⁵ Såvel toneangivende forskere som socialrådgivernes formand kritiserede grundlaget for anbefalingen for at være for spinkelt (Dansk Socialrådgiverforening, 2012). Hverken forsikring eller evaluering kan opsuge al risiko ved en uønsket hændelse.

Fra ritual til forsikring

Denne artikel har undersøgt, i hvilken forstand evaluering kan forstås som forsikring, og hvilke implikationer det har for forståelsen af den rolle, evaluering spiller i samfundet. Der er tilstrækkelig mange ligheder til, at centrale dele af evalueringspraksis kan forstås inden for rammen af forsikringsmetaforen. De væsentligste lighedspunkter er opstillet i tabellen nedenfor.

Det er ikke i sig selv værdifuldt at finde metaforer for en samfundsmæssig praksis. Metaforene skal kunne skabe indsigter, der har implikationer for forståelsen af den rolle, evaluering spiller i samfundet. Ritualer som metafor blev indledningsvis kritiseret for at opstille et modbillede til evaluering, der divergerer så stærkt fra evaluators og evalu-

eringsbrugerens mere rationalistiske beskrivelse af evaluering, at dialog vanskeliggøres. Ved at anvende forsikring som alternativ metafor fremhæves de rationalistiske aspekter af evaluering, men samtidig skaber metaforen et blik for aspekter af evalueringspraksis, som kan have et potentiale for at kvalificere forståelsen af evaluering.

Metaforer gør os i stand til at forstå ét begreb gennem et andet ved at fremhæve lighed og forskellighed. Det gælder også i forhold til at forstå implikationer. Den væsentligste lighed mellem evaluering og forsikring er spredningen af ansvar. Her synliggør metaforen, at på trods af, at evaluering ofte tages i brug for at give en mere entydig ansvarsplacering, kan evaluering også bidrage til det modsatte, nemlig at ansvar spredes. På den baggrund bør det ikke overraske, at evaluering er et yndet styringsinstrument, hvis anvendelse stiger: På en og samme tid formår evaluering at stille de producerende enheder i den offentlige sektor til ansvar, samtidig med at ansvaret for svære beslutninger utydeliggøres højere oppe i hierarkiet.

Metaforen bidrager til forståelse ved at synliggøre forskelle, og en central forskel er opgørelsen af omkostninger. Forsikringsmetaforen minder om, at omkostninger ikke blot er de omkostninger, der er forbundet med kerneydelsen (dækning af tab ved visse hændelser). Omkostningerne er også udgifter til forsikringstagers ændrede adfærd (bedrageri og moralfare) og de omkostninger, der er forbundet med at minimere de selvskabte risici.

Tabel 1: Forsikring vs. evaluering

	Forsikring	Evaluering
Funktion	Fordeling af økonomiske risici (penge)	Fordeling af politiske risici (ansvar)
Håndtering af risiko for insolvens	Genforsikring	Evaluering af evaluering
Moralfare (moral hazard)	Uforsigtig omgang med forsikrede genstande	Målforskydning: underfokusering på områder, der ikke er dækket af evaluering

Da det vil være vanskeligt at få et præcist billede, fordi omkostninger og fordele er spredt på mange aktører med forskellige perspektiver og forskellige muligheder for indflydelser, vil en uhildet cost-benefit-betragtning af evaluering kræve, at der er interesse for at få afdækket alle omkostninger hos den aktør, der bestiller evalueringen eller vedtager evalueringssystemet. Det kunne være forfriskende indimellem at forsøge at anlægge et mere nøgternt cost-benefit-perspektiv på evaluering for at få et oplyst grundlag uhildet refleksion, også selv om en forventning om et retvisende kronor og øre-billede af, om evaluering kan betale sig, ikke kan indfries.

Artiklen har beskæftiget sig med evaluering i meget bred forstand. Afslutningsvis er det centralt at overveje, om der er typer af evaluering, der i særlig grad egner sig til at distribuere ansvar – med de fordele og ulemper, det medfører. Er audit mere egnet end programevaluering? Er summativ evaluering mere egnet end formativ evaluering? Et sted at begynde at besvare denne type spørgsmål er ved at konstatere, at muligheden for konsolidering af evaluering gennem genforsikring i form af metaevaluering forudsætter en standardiseret tilgang. Der er grundlag for at opstille den hypotese, at jo mere standardiserede og jo mere elaborerede systemer for metaevaluering, der eksisterer, des mere egnede er evalueringer som forsikring. Det betyder også, at der er grund til at overveje, inden for hvilke dele af samfundslivet det er ønskeligt at lade beslutningstager benytte sig af stærke forsikringsmekanismer, hvor det er ønskeligt at lade evaluativ viden antage mindre kodificerede former eller helt undgå at bruge evaluering og i stedet støtte sig til beslutningstageres dømmekraft.

Noter

1. Tak til Peter Dahler-Larsen for værdifulde kommentarer til denne artikel.
2. For en nærmere analyse af forholdet mellem revision/audit og programevaluering, herunder en

beskrivelse af forskelle og ligheder mellem de to, se Lindeberg (2007a) og for en systematisk sammenligning af programevaluering, kvalitets sikring og revision/audit se Lindeberg 2007b.

3. Dette eksempel analyseres i Lindeberg 2007b.
4. For illustration af dette se Campbell Collaborations Resource Center på deres hjemmeside, link: http://www.campbellcollaboration.org/resources/research/the_production.php
5. Anbefalingen skete med henvisning til en rapport udarbejdet af Rambøll Management Consulting. Rapporten hed Analyse af de økonomiske konsekvenser på området for udsatte børn og unge og udkom i 2012.

Litteratur

- Austin, Robert D. (1996), *Measuring and Managing performance in organizations*, Dorset House Publishing.
- Bjerregaard, Anne (2009), *Støjbergs embedsmænd rettede i unge-rapport*, Altinget, www.altinget.dk/artikel/stoejbergs-embedsmaend-rettede-i-unge-rapport
- Dahler-Larsen, Peter (2000), *Den rituelle refleksion: om evaluering i organisationer*, Odense Universitetsforlag.
- Dansk Socialrådgiverforening (2012), *Forskere kritiserer ministers metodeforslag*, nyhed på Dansk Socialrådgiverforenings hjemmeside, www.socialraadgiverne.dk/Default.aspx?ID=9269
- Gollier, C. (2004), »Insurance«, i Neil J. Smelser og Paul B. Baltes, red., *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier B.V, pp. 7582-7.
- Harrington, Scott E. og Gregory R. Niehaus (2003), *Risk management and insurance*, McGraw-Hill Education.
- Huber, Michael (2002), *Conceptualising Insurance: risk management under conditions of solvency*, Discussion Paper 9, Centre for Analysis of Risk and Regulation, London School of Economics.
- Kaplan, Robert S. og David Norton (1992), »The Balanced Scorecard: Measures that Drive Performance«, *Harvard Business Review*, 70(1): 71-9.
- Lakoff, George og Mark Johnson (1980), *Metaphors We Live By*, The University of Chicago Press.
- Lindeberg, Tobias (2007a), »The Ambiguous Identity of Auditing«, *Financial Accountability & Management*, 23(3): 337-350.
- Lindeberg, Tobias (2007b), *Evaluative technologies: Quality and the Multiplicity of Performance*, Copenhagen Business School, ph.d.-serien.

- Luhmann, Niklas (1993), *Risk: A sociological Theory*, Aldine de Gruyter.
- Munro, Rolland og Jan Mouritsen, red. (1996), *Accountability: Power, Ethos and the Technologies of Managing*, International Thomson Business Press.
- Pentland, Brian T. (1993), »Getting comfortable with the numbers: Auditing and the micro-production of macro-order«, *Accounting, Organizations and Society*, 18(7-8): 605-20.
- Pollitt, Christopher (1998), »Evaluation in Europe: Boom or Bubble?«, *Evaluation*, 4(2): 214-24.
- Wilkinson, Adrian og Hugh Willmott, red. (1995), *Making Quality Critical: New Perspectives on Organisational Change*, Routledge.
- Power, Michael (1996), »Making things auditable«, *Accounting, Organizations and Society*, 21(2/3): 289-315.
- Power, Michael (1997), *Audit Society: Rituals of verification*, Oxford University Press.
- Power, Michael (2004), *The Risk Management of Everything: Rethinking the politics of uncertainty*, Demos.
- Power, Michael (2007), *Organized Uncertainty: Designing a World of Risk Management*, Oxford University Press.
- Ridgway, V.F. (1956), »Dysfunctional consequences of performance measurements«, *Administrative Science Quarterly*, 1(2): 240-7.
- Rieper, Olaf og Hanne Foss Hansen (2007), *Metode-debatten om evidens*, AKF Forlaget.
- Riley, Keith (1997), *The nuts and bolts of reinsurance*, LLP.
- Rothstein, Henry, Michael Huber og George Gaskell (2006), »A theory of risk colonization: the spiraling regulatory logics of societal and institutional risk«, *Economy and society*, 35(1): 91-112.
- Schwandt, Thomas A. (2015), »Iagttagelser om anvendelse af videnskabelig evidens i politik og praksis«, Økonomi og Politik, dette nummer.
- Scriven, Michael (1995), »The logic of Evaluation and Evaluation Practice«, *New Directions for Evaluation: Special Issue: Reasoning in Evaluation: Inferential Links and Leaps*, (68): 49-70.
- Seidl, David (2003), *The dark side of Knowledge*, Working paper 2003-7, Ludwig-Maximilians-Universität München, Munich School of Management.
- Thiel, Sandra van og Leeuw, Frans L. (2002), »The Performance Paradox in the Public Sector«, *Public Performance & Management Review*, 25(3): 267-81.

Otte opmærksomhedspunkter 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 teoriafsnit, 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 Sætre, 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 evalueringen 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 interessenterenes 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 forskere. 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 forskningsrådsevaluering 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 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 evalueringsinteressenter, 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 konkurrencemæssig 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 evalueringsinteressenter, 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 evaluators 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-

kurrencedygtige 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, sådan som tilfældet også var ved evalueringen af NAV-reformen, er repræsenteret i styringsgruppen. Det gælder både repræsentanter for opdragsgiver, de organisationer som skal implementere reformen, og for evaluatorene (forskere). 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 gruppers 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 evalueringsresultaternes 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 spørgsmålstejn 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 armlængdeprincip 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 spille.« 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 armlængdeprincippet 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 armlængde 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 evalueringes 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 besluttet 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 moduledere, en mere væsentlig opgave end ved traditionelle forskningsprojekter med at udvikle evalueringsdesignet 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 spørgsmålstegn 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 spille meget tid og mange ressourcer på at forsøge at integrere forskellige forskningstraditioner, 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 evaluatore. 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 sjælden 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: Norges forskningsråd.
- 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.

Evaluering af politisk-administrative reformer: En komparativ analyse af evalueringen af de største reformer i Danmark og Norge

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

**Gunnar Gjelstrup, lektor, Institut for Statskundskab,
Københavns Universitet, gg@ifs.ku.dk**

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

**Hanne Foss Hansen, professor, Institut for Statskundskab,
Københavns Universitet, hfh@ifs.ku.dk**

To af de største politisk-administrative reformer i Norden i 2000'erne var reformen af arbejds- og velfærdsadministrationen i Norge (NAV-reformen) og kommunalreformen i Danmark. De to reformer har en række kendetegn til fælles – også med andre store politisk-administrative reformer. De blev vedtaget centralt af Stortinget/Folketinget, de berørte store dele af den offentlige sektor og befolkningen, de havde flere og samtidig – i nogen grad – vanskeligt forenelige målsætninger og de udviklede og forandrede sig over tid. De to reformer var også begge kendetegnet ved at formelle strukturændringer var centrale og berørte såvel central- som lokal- og regional administration. Sidst, men ikke mindst er begge reformer blevet evalueret, hvilket internationale undersøgelser har peget på har været undtagelsen snarere end reglen for reformer af denne karakter. Det er derfor interessant at analysere de to evalueringer for at forstå, hvordan det gik til, at reformerne blev evalueret, og hvilken betydning disse evalueringer fik. Artiklen viser, at trods reformernes lighedstræk blev evalueringerne

organiseret forskelligt og fik varierende betydning. Mens evalueringen af NAV-reformen blev organiseret som en forskningsbaseret ekstern evaluering, blev evalueringen af kommunalreformen organiseret som en intern evaluering, hvor centrale interessenter fik evalueringsansvar. Og mens evalueringen af NAV-reformen overvejende blev anvendt vidensopbyggende og debatskabende, blev evalueringen af kommunalreformen anvendt instrumentelt til justering af arbejdsdelingen mellem forvaltningsniveauer på udvalgte opgaveområder. Artiklen afsluttes med refleksioner over, hvordan disse resultater kan tolkes.

Hvorfor evalueres sammenlignelige reformer forskelligt?

Både i Norge og i Danmark har der i de senere år været gennemført evalueringer af to omfattende politisk-administrative reformer. I Norge drejer det sig om reformen af arbejds- og velfærdsadministrationen (NAV-re-

formen), i Danmark om kommunalreformen. Begge reformer var et resultat af politiske beslutninger gennemført i Stortinget/Folketinget. Begge reformer var komplekse, idet de sigtede mod at opnå vanskeligt forenelige målsætninger. Og begge medførte betydelige strukturerændringer, der berørte såvel statslige, som regionale og lokale forvaltningsniveauer.

Den internationale forskningsbaserede viden om evalueringer af større politisk-administrative reformer er sparsom. Den litteratur, der findes, peger på, at reformer med karakteristika som de to omtalte sjældent evalueres (Pollitt, 1995: 135). Herudover peger resultaterne af et komparativt landestudie på, at i de tilfælde, hvor sådanne reformer evalueres, organiseres disse evalueringer sjældent som eksterne evalueringer gennemført af evaluatore, der ikke har været involveret i implementeringen af reformerne, men derimod som interne evalueringer:

»External evaluation of public sector reforms in terms of methodologically rigorous (social science-based) evaluation research is still more or less a rarity throughout all countries, particularly with a comprehensive mandate« (Wollmann, 2003: 250).

Det er derfor interessant at se nærmere på, dels hvordan det overhovedet gik til, at de to evalueringer kom til veje, dels om de blev organiseret som eksterne eller interne evalueringer, og hvilken betydning valget af organisering fik. Artiklen fokuserer på tre konkrete forskningsspørgsmål: 1) Hvorfor og hvordan blev de to evalueringer initieret? 2) Hvordan blev de fulgt op og anvendt? 3) Hvilke ligheder og hvilke forskelle er der evalueringerne imellem, og hvordan kan disse forstås? Mens de to første spørgsmål besvares gennem en beskrivende analyse af de to evalueringers genstand, formål, organisering, konklusioner og anvendelse, besvares det tredje spørgsmål gennem en analyse forankret i fire forskellige grundlæggende organisations- og styringsteoretiske positioner – et rationelt, instrumen-

telt perspektiv, et politisk interessebaseret perspektiv, et institutionelt, kulturelt perspektiv og et kaosperspektiv.

Artiklen er struktureret som følger: I det næste afsnit præsenteres analysens teoretisk-analytiske referenceramme i form af de fire organisations- og styringsteoretiske positioner samt analysens metodiske grundlag. Herefter præsenteres de to cases i en beskrivende analyse. Først gøres der kort rede for indholdet i de to reformer, derefter for de to evalueringers genstand og formål, organisering, konklusioner samt anvendelse. Derefter følger den teoretiske tolkning af den beskrivende analyses resultater. Til sidst opsummeres artiklens centrale konklusioner, og disse tages op til en perspektiverende drøftelse.

Fire perspektiver på organisation og styring

Den komparative analyse af de to evalueringer vil, som allerede nævnt, tage udgangspunkt i fire forskellige grundlæggende organisations- og styringsteoretiske positioner: Et rationelt, instrumentelt perspektiv, et politisk interessebaseret perspektiv, et institutionelt, kulturelt perspektiv og et kaosperspektiv.¹

Det rationelt, instrumentelle perspektiv er forankret i forestillingen om, at der findes en systemmæssig mål-middel rationalitet. I denne kontekst vil de systemfastsatte målsætninger bestemme, hvilke midler der vælges, når der tages beslutninger om problemløsning (Scott & Davis, 2006: Kapitel 2; March, 1995: 51-70). Overført til evaluering indebærer dette perspektiv, at evalueringer forventes organiseret og designet i overensstemmelse med de overordnede formål, der er fastlagt i forhold til at gennemføre disse. Konkret forventes den politiske ledelse at stå samlet om formålet med at evaluere, således at dette formål kan udmøntes i organisering og design, der derefter kan udgøre en ramme for en analyse og vurdering.

Det politisk, interessebaserede perspektiv er forankret i et smallere rationalitetsbegreb. I dette perspektiv har individer, grupper og organisationer målsætninger, som de hver for sig forsøger at fremme gennem koalitionsdannelse (Cyert og March, 1992). I dette perspektiv opfattes reformprocesser, men også evalueringsprocesser som arenaer for forhandling. Forventningen inden for dette perspektiv er, at en evaluerings design og organisering er et resultat af en forhandling mellem de deltagende aktører, ligesom selve evalueringsprocessen og de konklusioner, der produceres i denne, er et resultat af forhandlingsprocesser.

Det institutionelle, kulturelle perspektiv betragter beslutninger og adfærd på individ-, organisations- og organisationsfeltniveau som norm- og kulturstyret (Schein, 1985). Aktørerne agerer, som de oplever, at det forventes af dem, og følger en logic of appropriateness (March og Olsen, 1989). Overført til evaluering betyder dette, at evalueringer designes og organiseres på basis af, hvad der betragtes som en passende måde at designe og organisere på.

Kaosperspektivet understreger betydningen af uklarhed, kompleksitet og foranderlighed for organisationers og organisationsfelters udvikling. Organisationer og organisationsfelter betragtes som emergerende, dvs. de forandres i ikke-lineære processer, defineret som processer der ikke kan forstås på baggrund af den historiske udvikling men derimod alene på baggrund af deres løst-koblede struktur (March, 1995; Weick, 2001; Stacey, 2011). Beslutningstagnation er i dette perspektiv præget af tilfældigheder forårsaget af tidsspecifikke koblinger mellem aktørers opmærksomhed og ressourcer på den ene side og problem- og løsningsstrømme på den anden. Overført til evaluering forventes evalueringsprocesser at være emergente og uforudsigbare processer, der er karakteriseret ved komplekse dynamikker og tilfældigheder.

De fire perspektiver vil blive anvendt både i analysen af evalueringernes organisering og design og i den bredere analyse af reformprocesserne og deres opfølgning.

Analytisk anvendes de fire perspektiver i en udfyldningsstrategi (Roness, 1997), til tider også benævnt en komplementerende strategi (Grøn, Hansen og Kristiansen, 2014). Konkret betyder dette, at teorierne supplerer hinanden. Vi vil med dette afsæt diskutere, om forskellige teorier kan bidrage til at skabe forståelse sekventielt i forhold til de to evalueringers forskellige faser samt om teorierne hver især kan bidrage til forståelse af en del af undersøgelsens genstandsfelt, herunder primært om der er forskel på, hvilke teorier der bidrager til forståelse af de to cases.

Dokumenter og interviews med nøglepersoner

Artiklens empiriske analyser baserer sig på et bredt udsnit af dokumenter, herunder evalueringsoplæg, rapporter, lovforslag, referater etc. Med henblik på at afdække de involverede interessenters opfattelser og vurderinger af evalueringerne og deres forløb er der herudover i begge lande gennemført semistrukturerede interviews med en række nøglepersoner (13² i Norge i foråret 2014 og 4³ i Danmark i foråret 2013). Interviewene har centreret om temaer som evalueringens organisering, processen, konklusionerne, anvendelse etc. (for en nærmere præsentation af det bagvedliggende empiriske materiale se Breidahl et al., 2014: 12; Hansen og Gjelstrup, 2014: 388f).

To reformer – to evalueringer:

En sammenligning

I det følgende præsenteres resultaterne af den beskrivende komparative analyse. Først præsenteres kort hovedtræk ved de to reformer, derefter karakteristika ved de to evalueringer.

De to evalueringer: Formål og indhold

Hovedtræk ved de to reformer er opsummeret i tabel 1:

Tabel 1: De to reformer i hovedtræk

	NAV-reformen	Kommunalreformen
Vedtaget	2005	2004-2006
Ikrafttrædelse	1. januar 2006	1. januar 2007
Hovedelementer	Samling af tre forskellige tjenester i én forvaltning	En inddelings-, opgave- og finansieringsreform
Reformens overordnede formål	1) En stærkere arbejdslinje, 2) En mere brugerorienteret tilgang og 3) Opnå effektiviseringsgevinster	1) Skabe en bæredygtig og fremtids-sikret offentlig sektor, 2) Løse opgaverne tæt på borgerne, 3) Opnå synergigevinster
Beslutning om evaluering	Indbagt i beslutning om reform	Efter politisk uenighed beslutning om at iværksætte evaluering i 2011
Evalueringsperiode	7 år fra 2007 til 2014	2012-2013

NAV-reformen i Norge trådte i kraft 1. januar 2006 og bestod i en omfattende politisk reform af velfærdsadministrationen. Reformen betød, at tre forskellige tjenester – tidligere Trygdeetaten, Aetat og den kommunale social tjeneste – blev samlet i en forvaltning. I forlængelse heraf omfattede reformen en integration af en række velfærdssektorer samt introduktionen af såkaldte partnerskaber mellem Arbejds- og velfærdsadministrationen på den ene side og kommunerne på den anden. De første initiativer til at reformere velfærdsadministrationen blev taget i 2001 og den endelige beslutning om at gennemføre reformen i 2005. Der var bred konsensus herom i Stortinget, da Fremskridtspartiet var det eneste parti, der var fortalere, for en alternativ model. Som følge heraf blev en ny organisation – Den norske arbejds- og velfærdsadministration – lanceret. Implementeringen af reformen strakte sig over en lang periode – fra 2006 til 2010 – og flere ændringer fandt sted undervejs, bl.a. i 2008, hvor såkaldte forvaltningsenheder blev lanceret. I 2005, hvor NAV-reformen blev vedtaget af Stortinget, blev det samtidig vedtaget, at der skulle gennemføres en evaluering af reformen – *Evalueringen af NAV-reformen* (St.prp. nr. 46: 2004-2005), i det følgende benævnt NAV-evalueringen. En beslutning, der ikke var genstand for politiske uenigheder. Evalueringen forløb over 7 år fra 2007 til 2014.

Kommunalreformen i Danmark trådte i kraft 1. januar 2007 og bestod af en inddelings-, en opgave- og en finansieringsreform, hvis forskellige elementer blev vedtaget af Folketinget i perioden 2004-2006 (se nærmere i Indenrigs- og Sundhedsministeriet, 2005). Inddelingsreformen medførte, at antallet af kommuner blev reduceret fra 271 til 98, og at de 14 amter blev nedlagt og erstattet af 5 regioner. Opgavefordelingsreformen medførte en ændret opgavefordeling mellem stat, regioner og kommuner, mens finansieringsreformen medførte ændret finansiering af opgaveløsningen. Implementeringen af reformen strakte sig over en periode og ændringer fandt sted undervejs bl.a. blev hele beskæftigelsesindsatsen i 2009 overdraget til kommunerne. Der var politisk uenighed om, hvorvidt reformen skulle evalueres, og først i 2011 besluttede en ny regering at iværksatte en evaluering af dele af reformen. Evalueringen blev gennemført 2012-2013. Evalueringen benævnes i det følgende KR-evalueringen.

Som det er blevet påpeget ovenfor, er der tale om to reformer, som ligner hinanden på en række centrale områder, bl.a. fordi de – på hver deres måde – var administrative og lagde op til en ny opgavefordeling mellem stat og kommune. I forlængelse heraf knytter evalueringerne an til at belyse væsentlige og helt centrale politiske spørgsmål. Men herud-

over er der betydelige forskelle de to evalueringer imellem. Som fremgået er en evaluering af den norske reform indbagt i beslutningen om at gennemføre reformen, mens der i Danmark i en årrække er politisk uenighed om, hvorvidt der skal gennemføres en evaluering. I det følgende vil yderligere forskelle mellem de to evalueringer blive afdækket, idet evalueringerne vil blive sammenlignet på fire felter: (1) evalueringernes genstand og formål, (2) evalueringernes organisering, (3) evalueringernes konklusioner og (4) evalueringernes anvendelse. Komparationen bygger på Breidahl et al. (2014) og Gjelstrup og Hansen (2014).

De to evalueringer: Genstand og formål

I evalueringsslitteraturen kan en evalueringens genstand bl.a. præciseres ud fra, om den har fokus på helheden eller enkelte dele samt et policy- eller et organisatorisk fokus. Tilsvarende introducerer litteraturen mange formål en given evaluering kan have, eksempelvis kontrol af om intentionerne blev opnået, dokumentation af effekter, påpegning af problemer undervejs med henblik på justeringer eller et oplysende formål i form af langsigtet vidensopbygning (se f.eks. Vedung, 1998; Hansen, 2005).

NAV-evalueringens genstand var hele den omfattende reform af den norske velfærdsadministration og alle aspekter ved den. Genstanden var bred og omfattede både helheden, enkeltdelene, policy og organisering. Tilsvarende havde NAV-evalueringen både til formål at vurdere både graden af målopfyldelse (er der kommet flere i arbejde samt bedre brugerorientering og effektiviseringer?) og implementeringsprocessen, at foreslå justeringer og at være oplysende gennem at skabe en langsigtet vidensopbygning. Den brede genstand og det mangesidede formål blev på et tidligt tidspunkt fastlagt i et »Mål- og rammedokument« (Arbejds- og inkluderingsdepartementet, 2006), der blev lagt meget arbejde i udformningen af, og det lå i sagens

natur, at der skulle evalueres. Undervejs blev en lang række aktører inddraget i forbindelse med afholdelsen af en række planlægningsmøder, bl.a. Forskningsrådet, en række forskere og repræsentanter fra det senere Arbejds- og Velfærdsdirektorat.

Til forskel fra NAV-evalueringens bredde og helhedsorientering var KR-evalueringens genstand og formål ikke fastlagt fra starten, men udviklede sig over tid i fem faser:

- 1) Efter strukturreformens iværksættelse var der, som tidligere nævnt, en politisk diskussion om den burde evalueres eller ej. Den daværende opposition fremlagde forslag om en ret bred evaluering og stillede spørgsmål til den ansvarlige minister om der burde gennemføres en sådan evaluering. Der var imidlertid ikke flertal for en evaluering (Folketinget, 2009, 2010; Haarder, 2011). I denne fase blev genstanden for en evaluering formuleret bredt, og formålet med at evaluere blev ikke præciseret.
- 2) Regeringsskiftet i 2011 indebar, at Socialdemokraterne, Det Radikale Venstre og Socialistisk Folkeparti indskrev en evaluering i regeringsgrundlaget, herunder et fokus på fire opgaveområder: det specialiserede socialområde; natur- og miljøområdet; regional udvikling; samt sundhedsområdet (Regeringen, 2011: 64). Evalueringens genstand var forsat bred, og formålet ikke præciseret.
- 3) Kommissoriet for evalueringen blev offentliggjort i februar 2012 og ministeren, Margrethe Vestager, udtalte bl.a. »at det er relevant at se på, om det var det rigtige snit, der blev lavet ved kommunalreformen i 2007« (Økonomi- og Indenrigsministeriet, 2012a). Konkret skulle evalueringen munde ud i en vurdering af »... den nuværende opgavefordeling mellem kommuner, regioner og stat samt overveje relevante justeringer af opgavefordelingen« samt en vurdering af »...

de eksisterende samarbejdsstrukturer og redskaber til koordination og styring mv. i den offentlige sektor samt overveje mulige justeringer heraf» (Økonomi- og Indenrigsministeriet, 2012b). Med kommissoriet indsnævredes evalueringens genstand. Evalueringens genstand var ikke længere helheden, men fire policy-områder, samt ikke et fokus på indhold i policy, men på de dele af organiseringen, der handlede om relevante snitflader mellem dele af systemet stat-regioner-kommuner. Samtidig var formålet indsnævret til at være et serviceeftersyn med henblik på justeringer og hverken en undersøgelse af, i hvilken grad intentionerne bag reformen var indfriet, eller af reformens effekter.

- 4) Under udvalgsarbejdet formuleredes en række tværgående tematikker (bl.a. politisk styring og bæredygtighed), som evalueringen også skulle belyse. Dermed udvidedes evalueringens genstand og formål, idet de bredere temaer lagde op til en vurdering af reformens grad af målopfyldelse, fx omkring øget bæredygtighed.
- 5) Evalueringens resultater blev offentliggjort i marts 2013, og hovedkonklusionen (se afsnittet om konklusioner) rummede en vurdering af graden af målopfyldelse. Hermed bekræftede hovedrapporten den ovenfor påpegede glidning i retning af en mere helhedsorienteret genstand for evalueringen, og at evalueringens formål også var at vurdere om reformens målsætninger var opfyldt.

De to evalueringer: Organisering

I evalueringslitteraturen sondres mellem intern og ekstern evaluering (se for eksempel Mathison, 2005: 150-1 og 206-8). NAV-evalueringen er et eksempel på en særlig form for ekstern evaluering, hvor Arbejds- og inklusionsdepartementet havde det overordnede ansvar og Norges Forskningsråd ansvaret for udvælgelse af evaluører. Mål- og rammedokumentet dannede grundlag for et udbudsmateriale, en tilbudsrunde, vurdering af

tilbud og en proces hvor der blev sammensat to konsortier (ledet af henholdsvis Arbejdsforskningsinstituttet og Rokkansenteret), hvor sidstnævnte også fik det samlede projektlederansvar. Udvælgelsen skete ved at det Norske forskningsråd nedsatte en ekspertgruppe til at vurdere ansøgningerne og indstille til en nedsat styringsgruppe bl.a. bestående af tre – af Forskningsrådet – udpegede medlemmer, centrale interessenter (herunder implementeringsorganisationerne) og embedsmænd (opdragsgiver). Efter udvælgelsen blev evalueringen organiseret i syv moduler fokuserende på hvert sit delaspækt af den samlede evaluering og med hver sin leder. Det samlede budget for evalueringen i perioden 2007-14 var i alt ca. 50 mio. NOK (senere justeret til 45 mio. NOK).

Til forskel herfra var KR-evalueringen intern i den forstand, at ansvaret blev placeret i et relativt smalt hovedudvalg og underudvalg (for hvert af de fire nævnte opgaveområder) sammensat af repræsentanter fra centrale ministerier (Økonomi/Indenrigs- og Finansministeriet), relevante sektorministerier samt Kommunernes Landsforening og Danske Regioner. Hovedudvalget blev sekretariatsbetjent af de centrale ministerier, og underudvalgene blev sekretariatsbetjent af sektorministeriet i samarbejde med de centrale ministerier. Organiseringen afspejler, at kommunernes og regionernes interesseorganisationer fik en central placering, og at sektorministerierne fik en platform, hvorfra de kunne spille ind. I praksis fungerede den organisatoriske konstruktion på den måde, at underudvalgene både var fora for faglige diskussioner og for forhandlings- og interesse-spil. Endelig afholdt udvalget en række høringer, bl.a. med politikere og chefer fra kommuner og regioner samt interesseorganisationer. Et forsigtigt skøn over udgifterne til evalueringen baseret på anslået timeforbrug hos de involverede medarbejdere og chefer i ministerier, kommunale organisationer mm.

er, at evalueringen har kostet et mindre tocifret millionbeløb.

De to evalueringer: Konklusioner

NAV-evalueringens formidling fandt sted løbende og på flere fronter, herunder bl.a. via evaluatorernes deltagelse i den offentlige debat og i en række møder arrangeret med brugerne. Selve resultaterne blev dokumenteret i ca. 87 publikationer, hvoraf størstedelen var rettet mod en norsk forskningsoffentlighed. Størstedelen af studierne fokuserede på implementeringen af reformen mens en mindre andel havde fokus på reformens målopfyldelse. Samlet set resulterede evalueringen i en påpegning af en række interessante generelle perspektiver, bl.a. at det for mange kommuner var alt for omfattende at integrere tre velfærdstjenester under et tag, at der var en række udfordringer forbundet med realiseringen af partnerskaber mellem det statslige og lokale niveau, at der opstod en række dilemmaer mellem at ansætte personer med generalist- versus specialistkompetencer i NAV-kontorerne, og at reformen næppe har fået flere i arbejde eller ført til øget brugerorientering og effektivisering. Blot for at nævne et lille udpluk af de mange forskelligartede resultater. Evalueringen har kun i begrænset omfang resulteret i input til implementeringsprocessen og forslag til justeringer. I stedet er evalueringens væsentligste hovedresultat et bidrag til langsigtet videnopbygning og til demokratisk debat, både i kraft af de mange publikationer og den løbende formidling af evalueringens fund.

KR-evalueringens resultater blev dokumenteret i 13 publikationer og ledsaget af en pressemeddelelse fra den ansvarlige minister i forbindelse med offentliggørelsen. KR-evalueringens hovedkonklusion var følgende (Økonomi- og Indenrigsministeriet, 2013: 19-20): »Samlet set er det udvalgets vurdering, at der med kommunalreformen var skabt en ramme for en mere robust offentlig sektor, der bedre er i stand til at møde de aktuelle

og kommende års udfordringer. Et af de centrale mål med reformen var at skabe kommuner og regioner med en større faglig og økonomisk bæredygtighed. Desuden var der en forventning om, at kommunalreformen kunne understøtte en omkostningseffektiv opgaveløsning med indhøstning af stordriftsfordele etc... På baggrund af udvalgets analyse, er det vurderingen, at kommuner og regioner i disse år er i færd med at realisere de nævnte målsætninger fra reformens indførelse« Denne hovedkonklusion legitimerede reformen: udviklingen var på rette vej i forhold til reformens målsætninger. Derudover resulterede KR-evalueringen også i, at der på de fire opgaveområder helt i tråd med kommissorierne fremkom mange specifikke vurderinger og forslag. Udvalgene påpegede snitfladeproblemer på tre af de fire områder. På det specialiserede socialområde blev foreslået en national koordinationsstruktur, hvor Socialstyrelsen gives en rolle i forhold til at kvalitetssikre og videreudvikle den specialiserede viden og indsats samt at der politisk bør vælges mellem (1) uændret opgavefordeling (som udvalget tydeligvis ikke selv troede på ville løse problemerne); (2) en kommunal model; og (3) et regionalt ansvar for tilbud til konkret afgrænsede målgrupper. På natur- og miljøområdet præsenteredes også flere konkurrerende, fremtidige organiseringsmuligheder bl.a. en fremtidig statslig eller en udvidet kommunal løsning (omkring godkendelse og tilsyn af virksomheder), og på råstofområdet bl.a. at flytte opgaver fra kommunerne til regionerne. Omkring genoptræning (sundhedsområdet) blev bl.a. foreslået skærpet regional myndighedskompetence, visitationsretningslinjer og øget gennemsigtighed.

De to evalueringer: Anvendelse

I evalueringsfeltets selvforståelse er det en præmis, at evalueringer iværksættes med sigte på at spille en rolle i praktisk handling. NAV-evalueringen er primært anvendt til en langsigtet videnopbygning, oplysning og demokratisk debat, og kun i begrænset om-

fang til at påvirke og justere implementeringsprocessen.

Til forskel herfra er KR-evalueringen umiddelbart anvendt på den måde, at den danske regering i juni 2013 fremlagde et oplæg svarende til evalueringsrapportens forslag til justeringer, hvilket senere samme måned førte til en bred politisk aftale (Regeringen, 2013a, 2013b). Ud over justeringerne er evalueringen også anvendt på den måde, at den har legitimeret reformen og lukket sagen i den forstand, at den tidligere S-SF-R-opposition ikke længere ønsker at rulle dele af 2007-reformen tilbage. På lidt længere sigt afhænger evalueringens betydning dels af den igangværende administrative opfølgning, dels af om Venstre efter et eventuelt regeringsskifte (i 2015) fortsat ønsker at nedlægge regionerne.

Fire perspektiver bidrager til at forstå forskellene mellem de to evalueringer

På basis af den beskrivende analyse diskuteres det i det følgende, hvordan de afdækkede forskelle de to evalueringer imellem kan forstås i lyset af de tidligere præsenterede teoretiske grundpositioner. Tabel 2 sammenfatter de fire teoriers bidrag til forståelse af de to cases.

Det *rationelt, instrumentelle* perspektiv bidrager i meget forskelligt omfang til forståelse af de to evalueringer. Hvor perspektivet giver et betydeligt bidrag til forståelse af procesforløbet i den norske case, giver det et begrænset bidrag til forståelse af det meste af procesforløbet i den danske case. Instrumentel rationalitet prægede i Norge procesforløbet, idet beslutningen om at gennemføre en evaluering i høj grad var baseret på et ønske om at sikre feedback og bidrage til langsigtet

Tabel 2: Teoretiske fortolkninger af de to cases

	NAV-evalueringen	KR-evalueringen
Rationelt, instrumentelt perspektiv	Betydeligt bidrag til forståelse af evalueringens organisering og design. Kun indirekte bidrag til forståelse af anvendelsen af evalueringens resultater, da disse blev anvendt oplysende snarere end instrumentelt beslutningsorienteret.	Begrænset bidrag til forståelse af evalueringens procesforløb og design. Et vist bidrag til forståelse af dele af organiseringen. Betydeligt bidrag til forståelse af den politiske opfølgning og anvendelse af evalueringens resultater.
Politisk, interessebaseret perspektiv	Begrænset bidrag til forståelse af evalueringen. Dog bidrag til forståelse af forhandlingsprocesserne forskergrupperne imellem om det konkrete design af evalueringen og arbejdsdelingen.	Betydeligt bidrag til forståelse af processen frem til beslutningen om at gennemføre evalueringen, til evalueringens organisering og til formulering af evalueringens konklusioner.
Institutionelt, kulturelt perspektiv	Bidrag til forståelse af det faktum at det var naturligt at gennemføre evalueringen og til den valgte organisering af evalueringen (begge blev taget for givent).	Bidrag til forståelse af evalueringens organisering (korporativ kultur) såvel som design (pragmatisme).
Kaosperspektiv	Begrænset bidrag til forståelse af evalueringen. Dog bidrag til forståelse af den proces hvor resultatområder blev formuleret.	Betydeligt bidrag til forståelse af evalueringens processens emergerende og uforudsigbare forløb, hvor fokus henholdsvis indsnævres og udvides undervejs.

vidensopbygning. Ligeledes kan beslutningen om at lade Norges forskningsråd stå for evalueringen for gennem en udbudsrunde at kunne vælge det bedste evalueringsprojekt tolkes som en rationel proces, en proces, som i øvrigt blev fulgt op med en beslutning i 2014 om at gennemføre en meta-evaluering i form af en evaluering af evalueringen med sigte på at skabe læring om, hvordan den offentlige sektor bedst planlægger, organiserer og gennemfører reformevalueringer. Omvendt var hverken forløbet op til beslutningen om at gennemføre en evaluering, beslutningerne om evalueringens design og ej heller selve evalueringsprocessen forløb i Danmark præget af instrumentel rationalitet. Modsat var derimod mønsteret i den efterfølgende anvendelsesproces. I Norge blev evalueringsresultaterne anvendt oplysende og debat-skabende snarere end instrumentelt beslutningsorienteret. I relation til den danske case giver det rationelt instrumentelle perspektiv derimod et betydeligt bidrag til forståelse af den umiddelbare politiske opfølgning og anvendelsen af evalueringsresultaterne i lovgivningssammenhæng. Evalueringens organisering kan i en vis udstrækning også forstås instrumentelt, især trækkene med det smalle hovedudvalg og den stærke koordinering fra centrale ministerier, om end hovedindtrykket af organiseringen er, at den byggede interesser ind i processen.

Også det *politiske, interessebaserede perspektiv* bidrager i meget forskelligt omfang til forståelse af de to evalueringer. Hvor perspektivet giver et betydeligt bidrag til forståelse af den danske case, giver det et begrænset bidrag til at forstå den norske. Mens der i Danmark i flere år var politiske diskussioner om, hvorvidt der overhovedet skulle gennemføres en evaluering, var der i Norge modsat ingen diskussioner om dette. Her var det en selvfølge, at reformen skulle evalueres. Mens rapportudarbejdelse og formulering af konklusioner var forhandlingsprocesser i den danske case, var dette analytiske processer

i den norske. Kun i den fase, hvor forskergrupperne skulle blive enige om design og arbejdsdeling synes interesseforankrede forhandlingsprocesser at have karakteriseret den norske proces.

Det *institutionelle, kulturelle perspektiv* bidrager til forståelse af begge cases men i relation til forskellige aspekter. I den norske case kan det, at det er en selvfølgelighed at gennemføre en evaluering og at ansvaret for at gennemføre denne henlægges til den evalueringslegitime organisation Norges forskningsråd forstås som en afspejling af en forvaltningsmodel med en veludviklet evalueringskultur. I den danske case kan det, at parterne i form af Kommunernes Landsforening og Danske Regioner inddrages på centrale poster i arbejdsgrupper og evalueringsudvalg, forstås som en afspejling af den danske tradition for korporatisme. Et kulturtræk ved den danske forvaltningsmodel har historisk været, at velorganiserede interesser inddrages i betydeligt omfang i såvel udarbejdelse af love og regler som i implementering (Hansen og Jørgensen, 2009; Grøn og Hansen, 2014). Ligeledes kan evalueringens design forstås som en afspejling af den danske tradition for pragmatisme i betydningen at lægge vægt på det umiddelbart praktisk gennemførlige. Om disse variationer i kulturtræk landene imellem mere overordnet kan tolkes som udtryk for, at politik og administration er tæt sammenvævet i den danske forvaltningsmodel, mens der er en større grad af arbejdsdeling mellem politik og administration i den norske er et tema for yderligere diskussion.

Det fjerde perspektiv *kaosperspektivet* bidrager som de to første i meget forskelligt omfang til forståelse af de to cases. Hvor perspektivet giver et betydeligt bidrag til forståelse af evalueringsprocessens uforudsigelige og emergerende forløb i den danske case, giver det et begrænset bidrag til forståelse af forløbet som sådan i den norske case. Dog er der også i den norske case faser i procesfor-

løbet, hvor usikkerheden er stor og tilfældigheder synes at spille ind. Dette gælder frem for alt den fase, hvori resultatområderne med ganske forskelligartede evalueringsformål fastlægges.

Konklusion

Analyserne har vist, at trods de to reformers lighedstræk blev evalueringerne organiseret meget forskelligt og fik varierende betydning. Mens evalueringen af NAV-reformen blev organiseret som en forskningsbaseret ekstern evaluering i en proces, der i betydeligt omfang var præget af instrumentel rationalitet, blev evalueringen af kommunalreformen organiseret som en intern evaluering, hvor centrale interessenter fik evalueringsansvar og dermed en platform til at udøve interessevaretagelse. Evalueringsprocessen blev i denne kontekst mere en forhandlingsproces end en analytisk proces. Samtidig har analyserne vist, at mens evalueringen af NAV-reformen blev anvendt vidensopbyggende og debatskabende, blev evalueringen af kommunalreformen anvendt direkte instrumentelt beslutningsorienteret til justering af arbejdsdelingen mellem forvaltningsniveauer på udvalgte opgaveområder.

Analysen illustrerer, hvordan valg knyttet til evalueringsorganisering fastsætter rammer for evalueringsprocessers karakter og lægger spor ud for anvendelsen af evalueringsresultaterne. I den norske case fører den eksterne organisering til evalueringsresultater, der betragtes som interessante på et generelt plan og derved skaber læring men som samtidig ikke opleves som særlig anvendelige i forhold til justering af reformens implementering. I den danske case fører den interne organisering modsat til, at der skabes forslag til reformjusteringer, der viser sig at være både politisk acceptable og mulige at gennemføre og samtidig har støtte fra de aktører, der efterfølgende skal deltage i implementeringen af disse. Det politiske og emergende aspekt knyttet til den danske evalueringsorganisering og design udgør dermed et væsentligt

grundlag for evalueringsresultaternes efterfølgende instrumentelle anvendelse.

Analyserne afdækker også, at de to evalueringsprocesser adskiller sig markant fra hinanden, hvad angår tidsperspektivet. Mens NAV-evalueringen løber over syv år gennemføres evalueringen af kommunalreformen på et år. Når omfattende politisk-administrative reformer, hvor implementering foregår over flere år, besluttet evalueret på tidspunktet for beslutningen af reformen, kan evalueringsprocessen planlægges, så den følger implementeringsprocessen. Dette sker i Norge, hvilket understreger det rationelt instrumentelle perspektivs forklaringskraft. Når der omvendt er politisk uenighed om, hvorvidt omfattende politisk-administrative reformer skal evalueres eller ej, og en evaluering først besluttet gennemført en årrække henne i implementeringsfasen, er tålmodigheden i relation til at vente på evalueringsresultater begrænset. Der opstår et tidspres for relativt hurtige resultater. Dette sker i Danmark, hvilket understreger det politiske, interessebase-rede perspektivs forklaringskraft.

Analysen har også peget på, at det institutionelle, kulturelle perspektiv bidrager til forståelsen af forskellene mellem de to cases. I Norge er det en selvfølgelighed at evaluere NAV-reformen og opgaven overdrages til Norges forskningsråd, som er en organisation med en veludviklet evalueringskultur. I Danmark er der længe politisk uenighed om, hvorvidt KR-reformen bør evalueres. Da beslutningen tages, organiseres evalueringen internt forankret i pragmatiske overvejelser om, hvad der umiddelbart er praktisk gennemførligt inden for den tidshorison, som den politiske dagsorden sætter. Den norske case med dens eksterne og forskningsbaserede karakter bryder dermed med resultaterne af de indledningsvist omtalte internationale undersøgelser, mens den danske case bekræfter disse. Er denne forskel udtryk for overgribende nationale kulturforskelle? I hvert fald er det tanke-

vækkende, at et lignende mønster tegner sig i relation til de evalueringer, som begge lande for nyligt har besluttet at gennemføre i relation til deres indsatser i Afghanistan. I Norge skal et udvalg bestående af uafhængige eksperter heriblandt en række forskere vurdere »Norges helhetlige innsats i Afghanistan i perioden 2001-2014« (Kongelig resolusjon 2014). I Danmark er en længere politisk diskussion af, hvordan en evaluering på området kunne organiseres, endt med en beslutning om en altovervejende intern »Erfaringsopsamling vedr. den samtænkte danske Afghanistanindsats« (Regeringen, 2014), hvor Forsvarsministeriet og Udenrigsministeriet gives ansvar for processen og hvor den militære indsat ikke sættes under lup. Noget tyder på, at Norge har en politisk mere risikovillig og mere oplysningsorienteret evalueringskultur forankret i værdier knyttet til uafhængighed.

Noter

1. Det rationelt, instrumentelle og det politisk interessebaserede perspektiv er diskuteret i Grøn & Kristiansen (2014), det institutionelt, kulturelle og kaosperspektivet i Grøn, Hansen & Kristiansen (2014). I teksten nedenfor gives enkelte henvisninger til central originallitteratur, der har inspireret diskussionen.
2. Arbeids- og socialdepartementet, Arbeids- og velferdsdirektoratet (AV-direktoratet), Kommunenes interesseorganisation (KS), Forskningsrådet, styringsgruppen og en række udvalgte forskere
3. Tre medlemmer af det hovedudvalg, der var ansvarlig for evalueringen samt et medlem af en af de fire arbejdsgrupper, der udarbejdede underliggende delrapporter.

Litteratur

- Arbejds- og inkluderingsdepartementet (2006), *Mål og rammedokument for evaluering av NAV-reformen*, Oslo.
- Breidahl, K.N., J.-E. Furubo, K. Halvorsen, and M.B. Hansen. 2014. »Opmærksomhedspunkter ved evaluering af politiske reformer. Rapport fra metævalueringen af NAV-reformen i Norge«, Norgesforskningsråd, Oslo.
- Cyert, R.M. og J.G. March (1992), *A Behavioral Theory of the Firm*, Cambridge: Blackwell Business.

- Folketinget (2009), *Forslag til vedtagelse V8 Om regioner og kommuner*, 11.11.09.
- Folketinget (2010), *Forslag til vedtagelse V69 Om den kommunale økonomi*, 07.05.10.
- Gjelstrup, G. og H.F. Hansen (2014), »Evalueringsstyring i en krisetid – kommunalreformen«, i C.H. Grøn, H.F. Hansen og M. Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag, pp. 385-416.
- Grøn, C.H. og H.F. Hansen, (2014), »Offentlig styring«, i C.H. Grøn, H.F. Hansen og M. Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag, pp. 87-105.
- Grøn, C.H., H.F. Hansen, og M. Kristiansen (2014), »Kultur-, symbol- og kaosperspektiver på styring«, i C.H. Grøn, H.F. Hansen og M. Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag, pp. 131-60.
- Grøn, C.H. og M. Kristiansen (2014), »Rationelle perspektiver på styring«, i C.H. Grøn, H.F. Hansen og M. Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag, pp. 107-29.
- Haarder, B. (2011), *Svar til medlem af Folketinget Julie Skovsby på spørgsmål nr. S 2025 til indenrigs- og sundhedsministeren*.
- Hansen, H.F. (2003), *Evaluering i staten*, Frederiksberg: Samfundslitteratur.
- Hansen, H.F. (2005), »Choosing Evaluation Models«, *Evaluation*, 11(4): 447-62.
- Hansen, H.F. (2013), »Systemic Evaluation Governance«, *Scandinavian Journal of Public Administration*, 16(3): 47-64.
- Hansen H.F. og A. Hansen (2000), »Evaluering i Danmark«, *Nordisk Administrativ Tidsskrift*, 81(2): 156-177.
- Hansen, H.F. og T.B. Jørgensen (2009), »Den danske forvaltningsmodel og globaliseringens udfordringer«, i M. Marcussen og K. Ronit, red., *Globaliseringens udfordringer. Politiske og administrative modeller under pres*, København: Hans Reitzels Forlag, pp. 36-64.
- Hansen, M.B., K.N. Breidahl, A. Halvorsen, and J.-E. Furubo. 2015. »Otte opmærksomhedspunkter ved evaluering af politiske reformer« *Økonomi og Politik*: 24-33.
- Indenrigs- og Sundhedsministeriet (2005), *Kommunalreformen – Kort fortalt*, København.
- Kongelig resolusjon (2014), *Oppnevning av et utvalg*

- som skal vurdere Norges helhetlige innsats i Afghanistan i perioden 2001-2014, Oslo, 21.11.14.
- March, J.G. (1995), »Hvordan beslutninger bliver til i organisationer«, i J.G. March red., *Fornuft og forandring. Ledelse i en verden beriget med uklarhed*, København: Samfundslitteratur, pp. 51-70.
- March, J.G. og J.P. Olsen (1989), *Rediscovering Institutions: The Organizational Basis of Politics*, New York: The Free Press.
- Mathison, S. (2005), *Encyclopedia of Evaluation*, London: Sage Publications Inc.
- Pollitt, C. (1995), »Justification by Works or by Faith? Evaluating the New Public Management«, *Evaluation*, 1(2): 133-54.
- Regeringen (2003), *Vækst, Velfærd, Fornyelse II. Supplerende regeringsgrundlag*, København, 27. august.
- Regeringen (2011), *Et Danmark der står sammen*, Regeringsgrundlag, København, oktober.
- Regeringen (2013a), *Bedre kvalitet og samarbejde – Opfølgning på evaluering af kommunalreformen*, København, juni.
- Regeringen (2013b), Aftale mellem regeringen, Venstre, Dansk Folkeparti, Enhedslisten, Liberal Alliance og Konservative, *Rammer for justering af kommunalreformen*, København, 26. januar.
- Regeringen (2014), *Erfaringsopsamling vedr. den samtænkte danske Afghanistanindsats*, København, Politisk aftale, 14. november.
- Roness, P.G. (1997), *Organisasjonsendringar. Teori- ar og strategiar for studiar av endringsprosessar*, Bergen: Fagbokforlaget.
- Schein, E.H. (1985), *Organizational Culture and Leadership*, San Francisco: Jossey-Bass.
- Scott, W.R. og G.F. Davis (2006), *Organizations and Organizing: Rational, Natural and Open Systems*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Stacey, R.D. (2011), *Strategic Management and Organizational Dynamics. The Challenge of Complexity*, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Strukturkommissionen (2004), *Hovedbetænkning. Betænkning nr. 1434*, København.
- Vedung, E. (1998), *Utvärdering i politik och förvaltning*, Lund: Studentlitteratur.
- Wollmann, H. (2003), »Evaluation in public sector reform: Trends, potentials and limits in international perspective«, i H. Wollmann, red., *Evaluation in Public Policy Reform*, Cheltenham/Northampton: Edward Elgar, pp. 231-58
- Weick, K.E. (2001), *Making Sense of Organizations*, Malden, MA: Blackwell Publishers.
- Økonomi- og Indenrigsministeriet (2012a), *Kommunalreformen får et serviceeftersyn*, Pressemeldelse, København, 9. februar.
- Økonomi- og Indenrigsministeriet (2012b), *Kommis-sorium for udvalg om evaluering af kommunalre-formen*, København, 27. januar.
- Økonomi- og Indenrigsministeriet (2013), *Evaluering af kommunalreformen*, København.

Mål- og resultatstyring i kommunerne efter krisen

**Eva Moll Sørensen, postdoc ved Institut for Statskundskab,
Københavns Universitet, ems@ifs.ku.dk**

**Søren Kjær Foged, ph.d. ved Institut for Statskundskab,
Københavns Universitet, skf@ifs.ku.dk**

Denne artikel præsenterer resultaterne af seks kvalitative casestudier af mål- og resultatstyring i danske kommuner efter krisen i 2009. På denne baggrund argumenteres det, at kommunerne siden krisen har arbejdet med at gøre mål- og resultatstyring til et redskab, der kan bruges til at sætte mål for deres strategiske udvikling og skabe grundlag for prioritering af deres indsatser. Det sker bl.a. ved, at kommunernes politiske mål indordnes i hierarkier, og mål- og resultatstyringen kobles til budgetlægningen og regnskabsaflægningen. Kommunerne har samtidig nedprioriteret eller forenklet deres koncernfælles redskaber til kvalitetsstyring af de decentrale institutioner.

Mål, resultatmålinger og evalueringer har længe været en del af hverdagen i de danske kommuner. Dette indebærer dog langt fra, at mål- og resultatstyring er blevet gjort til et princip for kommunernes tværgående koncernstyring. De tidligere undersøgelser har således peget på, at der ofte er beskeden sammenhæng mellem de mål, at der styres efter forskellige steder i koncernen, at der ikke følges systematisk op på formulerede mål, og at der sjældent skabes koblinger mellem faglige og økonomiske resultater (e.g. Rieper et al., 2000; Ejersbo og Greve, 2005; KREVI, 2008).

Kommunerne har dog i en årrække arbejdet med at skabe sammenhængende mål- og resultatstyring på tværs af koncernen, til dels

som svar på en række statslige initiativer. Disse initiativer drejede sig i tiden omkring kommunalreformen særligt om at dokumentere kvaliteten af de kommunale services og skabe råderum og ansvar for resultater på decentralt niveau (Sørensen, 2013). Da krisen ramte i 2009, skiftede fokus for den statslige styring til et krav om bedre økonomisk styring i kommunerne, herunder bedre sammenhæng mellem den økonomiske og faglige styring (Hansen et al., 2014).

Vi ved endnu ikke ret meget om, hvordan kommunernes anvendelse af mål og resultatmålinger har udviklet sig efter den økonomiske krise. Styrer kommunerne i større eller mindre grad efter mål og resultater? Kobler de i højere grad styringen af økonomiske og faglige resultater? Eller får krisen dem måske til at nedprioritere styringen af den faglige kvalitet? Artiklen her kaster lys på disse spørgsmål med udgangspunkt i følgende problemformulering:

Hvordan har kommunernes anvendelse af mål og resultatmålinger i den tværgående koncernstyring udviklet sig efter krisen i 2009?

Spørgsmålet besvares med udgangspunkt i casestudier i seks danske kommuner. Den

teoretiske forventning er, at krisen vil føre til, at mål og resultatmålinger i højere grad bruges til at prioritere udgifterne, mens styringen af ydelsernes indhold og kvalitet decentraliseres til de faglige forvaltninger eller centre. Baggrunden for forventningerne udfoldes nedenfor.

Mål- og resultatstyring i krisetider

Mål- og resultatstyring kan defineres bredt som »anvendelse af mål og resultatmålinger i offentlig styring«. I dette afsnit diskuteres, hvordan mål og resultatmålinger anvendes i offentlig styring, og hvordan vi kan forvente, at anvendelsen udvikler sig i krisetider. Endvidere inddrages den statslige styringspolitik under krisen som baggrund for at formulere forventninger til kommunernes anvendelse af mål og resultatmålinger i deres styring under krisen.

Howdan anvendes mål og resultatmålinger i offentlig styring?

Mål- og resultatstyring beskrives ofte som en styringsform, hvor resultatmålinger bruges til at træffe de beslutninger og fremme den adfærd, der fører til størst mulig opfyldelse af organisationens mål. Det indgår i denne instrumentelle forståelse, at der både formuleres klare mål for organisationen, og at resultatmålinger bruges til at træffe beslutninger og styre efter målene. Mål- og resultatstyring kan således defineres som inkorporering og brug af kvantitative resultatmålinger i beslutningsprocesser (van Dooren et al., 2010) med reference til klare organisatoriske mål. Det er dog også muligt at forstå mål- og resultatstyring ud fra en mere normbaseret og social forståelse af organisationer og adfærd (March og Olsen, 1989; Grøn et al., 2014). I dette tilfælde kan formulering af mål og resultatindikatorer forstås som en måde at skabe en fælles forståelse af den retning organisationen skal arbejde i. Mål og resultatmålinger virker således styrende, fordi der sættes rammer for politiske processer, og organisationens deltagere socialiseres og forpligtes i forhold til

fælles mål. Med en sådan forståelse er der særligt fokus på, hvordan selve målformuleringen virker retningsgivende og dermed styrende og i mindre grad på, hvordan resultatmålinger bruges i beslutningssammenhæng.

I kommunal sammenhæng foregår mål- og resultatstyring både på politisk og administrativt niveau. På det politiske niveau er der tale om, at byrådets politikere formulerer mål for kommunens udvikling og indsatser. Målene kan evt. indgå i et målhierarki fra overordnede visioner til mere opgavespecifikke mål, som der f.eks. følges op på i de politiske fagudvalg. Der kan også formuleres politiske mål i forbindelse med budgetlægningen, som der følges op på i forbindelse med regnskabsaflæggelsen og/eller løbende budgetopfølgninger. De politiske mål kan også udmøntes gennem administrativ mål- og resultatstyring, der indebærer, at der sættes mål for den kommunale organisation og/eller for dens underliggende enheder, f.eks. i form af interne kontrakter eller aftaler. I forbindelse hermed etableres også en fælles forståelse af, hvilke frihedsgrader enhederne har til at nå målene.

Krisers betydning for anvendelsen af mål og resultatmålinger

Der er gode grunde til at forvente, at mål og resultatmålinger kommer til at spille en større rolle i løbet af en krise i takt med, at kravet om besparelser bliver oplevet som mere vedvarende. I litteraturen om 'cutback management', dvs. besparelser i den offentlige sektor, skelnes typisk mellem inkrementalistiske 'grønthøsterbesparelser', produktivitetsforbedringer, og målrettede besparelser baseret på rationel analyse og politisk prioritering (Kickert et al., 2014; Pollitt, 2010; Beck Jørgensen, 1981). Nogle forskere mener, at vedvarende kriser fører til, at grønthøstbesparelser afløses af produktivitetsforbedringer og strategiske prioriteringer (Beck Jørgensen 1981; jf. Kickert et al., 2014). Dette faseforløb finder ikke altid empirisk støtte (Pollitt,

2010; Dunsire og Hood, 1989), men det er, som det argumenteres i det følgende, alligevel en plausibel forventning, at presset for strategisk prioritering stiger i takt med stigende ressourceknaphed.

Grønthøsterbesparelser kræver ikke megen central prioritering og analyse; tværtimod er princippet, at besparelser fordeles jævnt over organisationens opgaveområder, og at udmøntningen af besparelserne delegeres nedad i organisationen. Besparelser gennem produktivetsforbedringer forudsætter derimod, at der findes information om både inputs og outputs. Besparelser gennem strategisk prioritering forudsætter, at der på centralt niveau er klarhed over effekterne af at træffe beslutninger om at nedprioritere bestemte indsatser for at kunne finansiere andre. Nok så vigtigt kræver vedvarende strategisk prioritering af ressourcerne, at der er størst mulig klarhed og enighed om organisationens mål. Dette er en udfordring i en offentlig sektor, som er kendetegnet ved at have mange mål og være underlagt demokratisk styring. I kommunerne er det desuden en udfordring, at deres opgaveområder er underlagt forskellige former for statslig styring, og at de er forbundet med vælger- og brugergrupper med forskellig indflydelse på de lokale politikere (Jørgensen og Mouritzen, 2002). Vi må imidlertid forvente, at der i krisetider gøres anstrengelser for at formulere og udmønte en fælles strategi, som muliggør prioritering af knappe ressourcer.

Der er dog også grund til at forvente, at mål og resultatmålinger kan komme til at spille en mindre rolle. F.eks. har Christopher Pollitt kortlagt resultatstyringens 'alternative' logikker (dvs. de u-intenderede, målforvrængende og selvforstærkende effekter), og han har i den forbindelse overvejet, hvad der sker med mål- og resultatstyring under krisen. Hans primære forventning er, at de politiske beslutningstagere vil ønske at skru ned for omfanget af resultatmålinger – eller i hvert fald offentliggørelse heraf – for ikke at syn-

liggøre det fald i offentlige servicestandarder, der må være resultat af krisen og de økonomiske nedskæringer. Afviklingen vil dog ikke ske pludseligt og åbent men gradvist og begrundet med en retorik om afbureaukratisering (Pollitt, 2013).

Pollitt skriver med reference til en engelsk kontekst, hvor resultatmålinger nok har været emne for større offentlig bevågenhed end i Danmark, og hvor nedskæringerne har været større. Der er dog grund til at fastholde forventningen om, at krisen vil føre til afvikling af mål- og resultatstyring, også i danske kommuner. Man kan f.eks. forvente, at kommunalpolitikeres ønsker om at spare på 'kolde hænder' vil ramme de medarbejdere, der udfører mål- og resultatstyring. Samtidigt er det muligt, at presset på økonomien har flyttet ledelsens fokus; at der så at sige sker en fortrængning af resultatstyring til fordel for budgetstyring. Sidst, men ikke mindst, er der måske områder, hvor man – som Pollitt argumenterer – ikke ønsker at skabe opmærksomhed omkring faldende servicekvalitet.

Statens styringspolitik før og efter krisen

Kommunernes anvendelse af mål og resultatmålinger formes til dels af de krav og signaler, som sendes til dem fra regeringen og KL. Disse signaler har ændret sig i løbet af krisen.

I tiden omkring kommunalreformen viste den daværende regering stor interesse for at fremme mål- og resultatstyring af og i kommunerne. F.eks. tog den initiativ til et 'Dokumentationsprojekt', der skulle skabe et sammenhængende sæt resultatindikatorer på tværs af de kommunale serviceområder, som bl.a. kunne bruges til at følge op på fælles mål i økonomiaftalerne (Sørensen, 2013; 2014a), og den søsatte en 'Kvalitetsreform', som bl.a. skulle skabe synlighed om kvaliteten i de kommunale institutioner. Det førte til, at kommunerne blev forpligtet til at offentliggøre en 'Kvalitetskontrakt' med kvantificerbare resultatmål, og de blev opfordret til at

måle den brugeroplevede og faglige kvalitet på bl.a. ældre- og børneområdet (Sørensen, 2013; 2014a).

Da krisen ramte, skiftede fokus for statens styring af kommunerne. I 2010 indførte VK-regeringen en 'Genopretningsplan', og kommunerne blev underlagt nulvækst og strammere sanktioner for at overtræde statens økonomiske rammer og deres egne budgetter. I 2012 blev sanktionsreglerne indlejret i en Budgetlov, som indebærer, at niveauet for kommunernes serviceudgifter (dvs. ca. 70 pct. af kommunernes udgifter) nu fastsættes af Folketinget i form af 4-årige udgiftslofter (Sørensen, 2014b). Siden indførslen af de nye regler har kommunerne markant forbedret deres budgetoverholdelse og reduceret serviceudgifterne¹ (se f.eks. Søren Kjær Fogeds artikel i dette temanummer). I årene 2008-2012 nedtonedes statens krav om, at kommunerne skulle målsætte og dokumentere kvaliteten af deres services.² F.eks. blev kravet om kvalitetskontrakter afskaffet i 2012 af S-SF-R-regeringen. Mål og resultatmålinger kom dog igen højere på dagsordenen, ikke mindst efter Moderniseringsaftalen i 2012 (Finansministeriet, 2012). F.eks. udgav regeringen og KL et 'Målbillede for God Økonomistyring' i 2013 (Finansministeriet, Økonomi- og Indenrigsministeriet og KL, 2013). Af dette fremgår bl.a., at kommunens budgetlægning bør have sammenhæng til politikker, strategier og mål.

Alt i alt ser det ud til, at krisepolitikken og kravet om besparelser for en periode fortrængte statens ambition om mål- og resultatstyring i kommunerne, men at ambitionen er blevet forstærket igen siden 2012. Dog har 00ernes fokus på at synliggøre kvaliteten i de kommunale institutioner fortonet sig til fordel for et ønske om, at kommunerne skal bruge mål og resultatmålinger til at prioritere deres ressourcer og forbedre produktiviteten.

Forventninger til kommunernes anvendelse af mål og resultatmålinger

Der er med baggrund i ovenstående grund til at forvente, at mål og resultatmålinger som følge af krisen vil komme til at fylde både mere og mindre i kommunernes tværgående koncernstyring.

Med baggrund i litteraturen om cutback management og i regeringen og KL's målbillede for økonomistyring kan vi forvente, at kommunerne i højere grad vil bruge resultatmålinger til at identificere mulige produktivitetsforbedringer, og at de i højere grad vil formulere strategiske mål for kommunen og skabe sammenhæng mellem disse mål og prioriteringen af ressourcer. I denne analyse fokuseres primært på, om kommunerne i højere grad bruger mål og resultatmålinger i forbindelse med strategiske prioriteringer af ressourcerne, da denne form for brug ville repræsentere det største brud med en inkrementalistisk 'før-krise' tilgang til budgetlægning og styring.

Med baggrund i Pollitts argumenter og med udgangspunkt i regeringens tilbagerulning af kravene om tværgående kvalitetsstyring forventes det, at kvalitets- og indholdsstyringen af de kommunale institutioner og deres ydelser i mindre grad vil være en del af den tværgående koncernstyring og i højere grad overlades til kommunens faglige enheder og deres arbejde med sektorspecifikke statslige styringskrav.

Da krisen især har påvirket kommunerne gennem statens strammere rammer og sanktioner for budgetoverskridelser, forventes det at særligt de kommuner, som har haft store problemer med budgetoverholdelsen, vil have foretaget store forandringer af deres styring efter 2010. Analysen tager således udgangspunkt i følgende forventninger:

- Krisen har ført til, at kommunerne i højere grad arbejder med at bruge mål og re-

sultatmålinger til at fremme en strategisk prioritering af ressourcerne i koncernen.

- Krisen har ført til, at kommunerne i mindre grad arbejder med koncernfælles systemer til mål- og resultatstyring af de decentrale institutioner og deres kvalitet.
- At forandringerne vil være størst i de kommuner, som før krisen har haft størst problemer med budgetoverholdelse.

Case studier i seks danske kommuner

Den følgende analyse er baseret på kvalitative case studier i seks danske kommuner. Kommunerne er udvalgt for at sikre variation mht. kommunernes udgiftsniveau før krisen, størrelsen af deres budgetoverskridelser i 2007-2009, og størrelsen af de besparelser, de har gennemført. Variationen bruges til at sandsynliggøre, at der er en sammenhæng mellem kommunernes grad af fokus på at forbedre økonomistyringen og finde besparelser og deres forandringer af mål- og resultatstyringen. Budgetoverholdelsen er opgjort som forskellen på driftsresultatet pr. indbygger i regnskab og budget, gennemsnitligt i årene 2007-2009. Disse tal er hentet fra Danmarks Statistik. Besparelserne er opgjort som den procentvise forskel på serviceudgifterne i 2009 og 2013 uden korrektion for brugertallet, idet det formodes at være vanskeligt at tilpasse udgifterne i takt med et evt. faldende antal brugere. Serviceniveau måles ved kommunens serviceudgifter pr. indbygger divide-

ret med udgiftsbehov. Disse indikatorer findes i de Kommunale Nøgletal. For sammenlignelighedens skyld er case-kommunerne i tabel 1 beskrevet ved, hvorvidt de enten ligger over eller under landsgennemsnittet (middel) på disse indikatorer.

Det er endvidere tilstræbt at sikre spredning mht. geografi, partipolitik og organisationsmodel, da disse faktorer kan have betydning for anvendelsen af mål og resultatmålinger, uden at vi kan sige præcist, i hvilken retning påvirkningen vil gå. Endvidere er fem af de seks kommuner at betegne som 'mellemstore' kommuner med mellem 42.000 og 69.000 indbyggere. Den sidste case-kommune (kommune E) hører til blandt de store bykommuner. Sidst, men ikke mindst, omfatter udvalget af casekommuner fire kommuner, som er resultatet af kommunesammenlægninger i 2007 (kommune A-D), og to fortsætterkommuner (kommune E og F).

Hvert case studie er baseret på semistrukturerede interviews med borgmesteren,³ kommunaldirektøren og de direktører, der har ansvar for de store serviceområder, dvs. børn og unge, socialområdet og ældreområdet. Desuden er der gennemført interviews med økonomichefen, en evt. stabschef for økonomi og styring, og en medarbejder, der arbejder med mål- og resultatstyring på koncernniveau. Interviewene støttes af dokumentarisk

Tabel 1: Case-kommunernes budgetoverholdelse, besparelser og serviceniveau før krisen

	Budgetoverholdelse før krisen (2007-2009)	Besparelser på service 2009-2013	Serviceniveau 2009
Kommune A	Under middel	Over middel	Over middel
Kommune B	Under middel	Under middel	Under middel
Kommune C	Under middel	Over middel	Over middel
Kommune D	Over middel	Under middel	Under middel
Kommune E	Over middel	Over middel	Under middel
Kommune F	Over middel	Under middel	Over middel

Egne beregninger på basis af de Kommunale Nøgletal. Faste priser med udgangspunkt i 2009.

materiale i form af budgetter, kommunernes egne beskrivelser af deres styring, mv. Som det fremgår, er analysen baseret på det syn på styringen, der findes i toppen af organisationen. Det er ikke undersøgt, hvordan styringen opfattes og anvendes på lavere niveauer eller bredt blandt kommunalpolitikkerne.

Mål- og resultatstyring i casekommunerne

Ovenfor argumenteredes det, at det især vil være de kommuner, som tidligere har haft store budgetoverskridelser, der vil være presset til at foretage forandringer af deres styring, herunder deres mål- og resultatstyring. Blandt casekommunerne er det især kommune A, B og C, som har haft store budgetoverskridelser op til krisen. Man må således forvente store forandringer i kommunerne A, B og C og mindre forandringer i kommunerne D, E og F. I det følgende redegøres for mål- og resultatstyringen i alle casekommuner med særligt fokus på de forandringer, der er sket efter krisen i 2009.

Kommuner med store styringsudfordringer

Kommune A har i årtier været udfordret af afvandring og røde sociale nøgletal, og den var fra sin dannelse kendetegnet ved central styring af økonomien og decentral styring af indholdet og kvaliteten af de kommunale services. Den første byrådsperiode var dog præget af en vis optimisme og et ønske om at bruge de styringsredskaber, der var moderne under højkonjunktoren, såsom aftalestyring. Krisen førte imidlertid til en erkendelse af, at kommunen fortsat vil stå overfor afvandring, og i årene efter krisen har man primært arbejdet med styringsredskaber, der skal sikre kapacitetstilpasning, økonomisk styring, og prioritering af beskæftigelse og 'den næste generation'. Informanterne fremhæver især byrådets Plan- og Udviklingsstrategi fra 2012, der indeholder tværgående, kvalitative mål for kommunens udvikling. Strategien er indirekte koblet til budgettildelingen gennem en budgetmodel, der fordeler besparelser med udgangspunkt i aktivitetsudviklingen på

de enkelte områder. Budgettet indeholder dog først fra 2015 eksplicitte politiske mål, og der har hidtil ikke været stort fokus på at følge op på kommunens politiske mål. Aftalesystemet, der blev etableret i kommunens første år, bruges ikke længere.

I kommune B har mål- og resultatstyringen gennemgået tre faser. I den første periode formulerede Byrådet en række servicemål, som var udgangspunkt for budgettildelingen, der gik direkte til de decentrale 'virksomheder'. Under servicemålene formulerede man 'servicekvaliteter' for ydelserne. Dette system blev afskaffet efter en omorganisering i 2010, der skulle skabe bedre økonomisk styring. Budgettet blev i stedet tildelt til politikområder, som blev dækket af 7 brede politikker. Desuden formulerede Byrådet en Vision, hvor sund økonomi står øverst, og en Udviklingsstrategi. Kommunen har også et administrativt aftalestyringssystem, som har overlevet krisen. Der har dog kun været begrænset opfølgning på de ikke-økonomiske resultater i organisationen. Det skyldes både det skarpe fokus på økonomi og mangel på gode data. Man arbejder nu med 3. generation af resultatstyring, som har til formål at skabe fokus på effekter på tværs af de faglige 'siloe' og dermed skaffe politikerne et råderum til prioritering og udvikling.

I kommune C blev der efter kommunalreformen etableret et system med politiske mål, der blev omsat i målbare resultat- og effektmål for de enkelte administrative centre. Disse mål indgik i budgettet og regnskabet frem til 2013, og de blev udmøntet i administrative års- og udviklingsplaner på alle niveauer. Dette system er dog gradvist blevet afløst af et nyt koncept for »Mål- og økonomistyring«, som blev udviklet i 2012. Konceptet indebærer, at der formuleres strategiske og tværgående udviklingsmål i en proces, der involverer hele organisationen. Målene indgår i et målhierarki under kommunens vision og politikker. De indgår også i budgettet, hvor

de udmøntes i 1-årige udviklingsmål for de enkelte centre. De er således en del af kommunens fælles mål- og budgetopfølgning. Ved indførslen af det nye koncept er de tidligere driftsorienterede budgetmål og års- og udviklingsplaner blevet gjort til frivillige redskaber for fagcentrene.

Kommuner med mindre styringsudfordringer

I kommune D blev der i årene efter kommunalreformen formuleret en lang række politikker. I 2010/2011 forsøgte man at skabe orden og koordination af politikkerne, idet man bl.a. vedtog fem overordnede pejlemærker. Øvelsen havde til formål at skabe fokus på kommunens overordnede mål og på effekterne af dens politikker. Den blev også brugt til at omfordele midler i retning af forebyggende indsatser, som skulle knække stigende udgiftskurver på de sociale områder. Samtidig reviderede man alle politikker efter en fælles skabelon. Der følges op på politikkerne i de enkelte fagudvalg. Politikkerne udmøntes gennem et omfattende administrativt mål- og resultatstyringssystem med en strategiplan, dialogbaserede aftaler og fælles redskaber til opfølgning. De politiske mål er p.t. kun løst koblet til budgetlægningen i form af henvisninger i budgettet til kommunens politikker, men der arbejdes med at skabe klarere koblinger mellem de politiske mål, indsatser og budgettildelingen. Samtidig arbejdes der med en forenkling af det administrative mål- og resultatstyringssystem.

Kommune E har siden midten af 1990'erne arbejdet med mål- og resultatstyring, idet der stilles krav om, at de enkelte forvaltninger formulerer politiske mål i forbindelse med budgetlægningen. Over årene er der arbejdet med at gøre målene mere fokuserede på effekt og operationalisere dem i delmål, hvortil der knyttes kvantitative resultatindikatorer. Der følges op på budgetmålene i forbindelse med regnskabsaflæggelsen. Krisen har ikke ført til afgørende forandringer af mål- og resultatstyringen. Der arbejdes dog løbende

med at sanere antallet af mål og skabe sammenhæng mellem sektorpolitikker og budgetmål. Som én af de seneste nyskabelser har byrådet vedtaget et sæt af overordnede mål for kommunen som offentlig organisation og i bredere forstand som bysamfund. Disse mål skal nu indarbejdes i målstyringen, så der skabes sammenhæng til sektorpolitikkerne og budgetmål. Kommunen har ikke en fælles model for administrativ mål- og resultatstyring, men der følges op på de enkelte enheders målopfyldelse vha. fagspecifikke ledelsesinformationssystemer.

Kommune F har en hierarkisk mål- og rammestyringsmodel, som har eksisteret siden 2006. Øverst i hierarkiet er en 4-årig kommunestrategi, der sætter retning for fagudvalgenes politikker og desuden omsættes i politiske effektmål i budgettet. De politiske mål udmøntes i organisationen gennem et administrativt kontraktstyringshierarki. Modellen blev ikke forandret i forbindelse med krisen, men der arbejdes aktuelt med at reformere den. Hvor tidligere kommunestrategier i vidt omfang var sammensat af bidrag fra de enkelte udvalg, har man i den seneste strategi valgt at fokusere på færre målsætninger af mere strategisk og tværgående karakter. Samtidig arbejdes der på at reformere den administrative målstyring, så de hidtidige kontrakter bliver erstattet af slankere dokumenter. Samtidig med, at der altså arbejdes med forenkling og fokusering af den nuværende målstyring, er det også ambitionen, at der skal skabes tættere koblinger mellem effektmål, aktiviteter og budgettildeling. Det er dog endnu uklart for undersøgelsens informanter, hvordan dette skal gøres.

Mere eller mindre mål- og resultatstyring i casekommunerne?

I tabel 2 nedenfor opsummeres analysen med udgangspunkt i de opstillede forventninger. Som forventet arbejder casekommunerne i højere grad med målstyring i forbindelse med deres strategiske prioritering af udgif-

Tabel 2. Opsummering af analysen i forhold til forventningerne

Forventninger	Mere brug af mål og resultatmålinger til prioritering af ressourcerne i koncernen	Mindre koncernfælles brug af mål og resultatmålinger til kvalitets- og indholdsstyring af decentrale enheder
Kommune A	Bekræftet. Stor vægt på strategisk målstyring i form af Strategi- og Udviklingsplan.	Bekræftet. Aftalesystem bruges ikke længere systematisk.
Kommune B	Bekræftet. Tværgående arbejde med effektmål som middel til at skabe råderum i budgetlægningen.	Bekræftet. Afvikling af system med servicemål, servicekvaliteter. Aftalesystem eksisterer dog fortsat.
Kommune C	Bekræftet. Tværgående strategiske udviklingsmål i budgettet.	Bekræftet. Års- og udviklingsplaner gøres frivillige for fagcentrene.
Kommune D	Bekræftet. Målstyring som middel til prioritering af ressourcer på tværs af (og indenfor) opgaveområder.	Delvist bekræftet. Det eksisterende aftalesystem forenkles.
Kommune E	Delvis bekræftet. Eksisterende system med politiske budgetmål for hvert fagområde fortsætter. Indordnes dog under overordnede mål.	Ikke bekræftet. Kommunen har ikke haft et fælles aftalestyringssystem
Kommune F	Delvist bekræftet. Eksisterende kommunestrategi og budgetmål får mere fokus på tværgående mål for kommunen som helhed.	Delvist bekræftet. Den eksisterende kontraktstyring forenkles og gøre mere dialogorienteret.

terne mellem opgaveområderne, hvorimod de har afskaffet eller forenklet deres koncernfælles systemer til styring af de decentrale enheder og deres ydelser. Det gælder i særlig grad de kommuner, som har været nødt til at rette voldsomt op på deres økonomistyring på baggrund af store budgetoverskridelser op til krisen (dvs. kommune A, B og C).

Mål- og resultatstyringen tilpasses behovet for økonomisk styring og prioritering

Som det fremgår af ovenstående, har alle casekommunerne forandret anvendelsen af mål og resultatmålinger i deres tværgående koncernstyring. Det kan dog diskuteres, om forandringerne skyldes krisen og kravet om besparelser. Det kan også tænkes, at casekommunerne har forandret deres styring som følge af de erfaringer, de har gjort med tidligere udgaver, eller fordi strategi og prioritering nu er blevet mere moderne end kvalitets- og indholdsstyring.

Det synes imidlertid klart, at de største forandringer er sket i kommunerne A, B og C, som samtidigt er de kommuner, der har været mest presset til at forbedre deres budgetoverholdelse. I alle tre kommuner har man afskaffet eller decentraliseret styringsredskaber, som havde til formål at udmønte en overordnet kvalitets- og indholdsstyring ned på det decentrale niveau. I kommune A gik man bort fra aftalestyringen, i kommune B afskaffede man systemet med servicemål og servicekvaliteter, og i kommune C har man ladet det være op til de faglige enheder, om de vil have års- og udviklingsplaner. I både kommune B og C er tilbagerulningen af denne del af mål- og resultatstyringen sket som led i større organisationsforandringer med det formål at rette op på økonomien. Kommunerne D, E og F har ikke på samme måde rullet styringsredskaber tilbage, men man har også her arbejdet med at slanke mål- og resultatstyringen af de decentrale enheder.

Kommunerne A, B, C har også arbejdet med nye mål- og resultatstyringsredskaber, der skal skabe tværgående prioritering af udgifterne. I kommune A lægger informanterne især vægt på Plan- og Udviklingsstrategiens bidrag til at sætte en politisk retning for de tværgående prioriteringer i retning af jobskabelse og den næste generation. I kommune B er opfattelsen, at de ændringer, der blev foretaget i 2010, skabte bedre central styring men også var med til at forstærke 'silo-tænkningen', og at man først med den 3. generation af resultatstyring er begyndt at skabe rum for tværgående, strategisk prioritering af udgifterne. I kommune C har målstyringskonceptet fra 2012 til formål at styrke organisationens fokus på tværgående, strategisk udvikling. Målene har dog karakter af udviklingsmål, som udmøntes i konkrete tiltag, frem for effektmål for de brede indsatser. De bruges således ikke til at foretage prioriteringer af udgifterne ud over de relativt begrænsede puljemidler, der er afsat til udvikling. I kommunerne D, E og F har man ikke indført nye mål- og resultatstyringsredskaber med det formål at skabe rum for tværgående strategiske prioriteringer. Kommune D har dog i flere omgange arbejdet med at bruge politisk mål- og resultatstyring til at prioritere udgifterne mellem og inden for opgaveområderne. I kommunerne B, C, E og F har man desuden en politik om løbende effektivisering ved procentvise udgiftsreduktioner. Disse effektiviseringer skal også skabe råderum til politisk prioritering af nye udviklingsområder.

På baggrund af ovenstående sammenligning synes det sandsynligt, at krisen skabte et pres for bedre økonomistyring, som også førte til forandringer af mål- og resultatstyringen. Det ser ud til, at forandringerne havde karakter af, at de mest pressede kommuner i en periode fokuserede på økonomisk styring og rullede dele af mål- og resultatstyringen tilbage. De øvrige kommuner har også arbejdet med at forenkle mål- og resultatstyringen, herunder især den administrative aftale- eller kontrakt-

styring, men de har ikke i samme grad afskaffet styringselementer.

Senere i forløbet efter krisen fokuserede man mere på at bruge mål- og resultatstyring til strategisk prioritering af udgifterne. Målstyringen bruges til strategisk prioritering, idet politiske mål indordnes i målhierarkier fra overordnede visioner og strategier til politikker for de faglige områder og for tværgående temaer. Casekommunerne bruger bl.a. den lovpligtige planstrategi til at sætte retning for udviklingen af kommunen og den offentlige service. Flere af kommunerne har desuden arbejdet med at reducere antallet af politikker og/eller indordne dem i et hierarki under få overordnede politikker. Nederst i målhierarkiet findes i casekommunerne mål om at gennemføre en indsats, evt. i form af et konkret projekt, og/eller mål om hvilke effekter der skal opnås. I nogle af kommunerne tænkes der forandringsteoretisk, så der (i hvert fald i princippet) er sammenhæng mellem deres indsatser og mål vedrørende effekterne.

Målene fungerer som rammesættende for de politiske processer, herunder budgetprocessen, og der indgår i fire ud af seks kommuner politiske mål i budgetterne. Dermed er målene også en del af den politiske budgetopfølgning, som kommunerne gennemfører en gang i kvartalet (eller oftere). I enkelte tilfælde er denne opfølgning baseret på et fælles ledelsesinformationssystem, men det er som regel op til fagområderne at definere målepunkter og resultatindikatorer og evt. gennemføre resultatmålinger til at supplere de indikatorer, der findes i statslige systemer (f.eks. brugertilfredsundersøgelser). Generelt bruges resultatopfølgningen ikke til at tage beslutninger, f.eks. om justeringer af indsatser eller om tildelingen af økonomiske ressourcer, men snarere til at sikre et løbende fokus på organisationens fælles mål. Der er således i højere grad tale om en målstyring baseret på norm-baseret og social handlingslogik end om en

målstyring, der indebærer en instrumentel anvendelse af resultatmålinger og evalueringer til at justere indsatsen og skabe bedre målopfyldelse. Der arbejdes dog i flere af casekommunerne med at skabe tættere, instrumentelle koblinger mellem strategier, effektmål, indsatser og budgettildeling. Det diskuteres dog stadig, hvordan man kan forene princippet om rammestyring med princippet om at målrette tildelingen til indsatser, som man forventer vil skabe bedre effekter, og det er også uklart, hvordan der vil blive fulgt op på, om en mere effektorienteret tildeling skaber de forventede effekter.

Konklusion: Fra kontraktstyring af institutionerne til strategisk målstyring af koncernen

At dømme ud fra vore case studier i 6 kommuner er der efter krisen sket en forandring af kommunernes anvendelse af mål og resultatmålinger i deres styring.

Kommunerne har som forventet nedtonet deres centrale kvalitets- og indholdsstyring af den service, der leveres på de kommunale institutioner. De har forenklet deres administrative mål- og resultatstyring, hvad enten denne har taget form af interne kontrakter, aftaler eller års- og udviklingsplaner, eller de har ladet det være op til deres faglige enheder, om de vil bruge sådanne styringsredskaber. Koncernstyringen omfatter heller ikke kommunikation af mål for den kommunale service til borgerne på tværs af opgaveområder, og anvendelsen af fælles kvalitetsmodeller, evalueringspolitikker el.lign. er også til at overse. Denne nedtoning er i overensstemmelse med udviklingen i den statslige politikformulering for kommunerne, hvor 00'ernes begejstring for kvalitetskontrakter med borgerne og kontraktlignende forhold mellem kommunerne og deres institutioner er afløst af en opfordring til at sikre stram økonomisk styring. Dette skal ikke forstås sådan, at hverken staten eller kommunerne er holdt op med at interessere sig for kvaliteten på de enkelte

kommunale opgaveområder, men det kunne se ud til, at kvalitets- og indholdsstyringen i mindre grad baseres på koncernfælles redskaber.

Dette indebærer dog langt fra, at kommunerne har opgivet at gøre mål og resultatmålinger til en integreret del af deres koncernstyring. Kommunerne arbejder, som forventet, i høj grad med at gøre mål- og resultatstyring til et redskab, der kan bruges til at sætte mål for deres strategiske udvikling og skabe grundlag for prioritering af deres indsatser. Det gælder, som forventet, ikke mindst i de kommuner, som har været nødt til at forbedre deres økonomistyring. Mål- og resultatstyringen opfattes således som et redskab, der skal hjælpe politikerne med at træffe store strategiske beslutninger frem for blot at få enderne til at mødes.

Uanset graden af økonomisk pres går tendensen i casekommunerne i retning af, at kommunens politiske mål indordnes i hierarkier, og at målformuleringen kobles til budgetlægningen. Resultatopfølgningen bliver derved et led i budgetopfølgningen, og kommunerne har således ikke sluppet styringen af de ydelser, der leveres, eller deres effekter. Dette er også i overensstemmelse med signalerne fra Slotsholmen, som går på at integrere politikker, strategier og resultatmålinger i økonomistyringen. Det skal dog fremhæves, at brugen af mål- og resultatstyring i forbindelse med budgetlægningen og regnskabsaflæggelsen er på udviklingsstadiet i de fleste af casekommunerne, og at koblingen mellem målformulering og budgettildeling er baseret på, at de politiske mål inddrages som rammesættende for budgetprocessen snarere end ved, at effektmål kobles direkte til budgettildelingen. Endvidere er der i kommunerne begrænset fokus på at følge op på de politiske mål vha. resultatmålinger og bruge opfølgningen til at tage beslutninger.

Noter

1. Serviceudgifterne er defineret som kommunens samlede nettodriftsudgifter fratrukket udgifter til overførsler og udgifter til de brugerfinansierede områder. Serviceudgifterne er blandt de vigtigste genstande for styring i de årlige aftaler mellem regeringen og KL.
2. Som illustration aftog brugen af ordet 'dokumentation' i økonomiaftalerne mellem staten og KL fra 21 gange i aftalen fra 2005 til 6 gange i 2008 og 0 gange i 2011 og 2012.
3. I den store kommune udestår interviewet med borgmesteren.

Referencer

- Dunsire, Andrew og Christopher Hood (1989), *Cut-back Management in Public Bureaucracies: popular theories and observed outcomes in Whitehall*, Cambridge: Cambridge University Press.
- Ejersbo, Niels og Carsten Greve (2005), *Contracts as reinvented institutions in the public sector*, Westport: Praeger Publishers.
- Finansministeriet (2012), *Aftaler om den kommunale og regionale økonomi for 2013*.
- Finansministeriet, Økonomi- og Indenrigsministeriet og KL (2013), *God økonomistyring i kommunerne*.
- Grøn, Caroline Howard, Hanne Foss Hansen og Mads Bøge Kristiansen, red. (2014), *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag.
- Hansen, Hanne Foss, Mads Bøge Kristiansen og Eva Moll Sørensen (2014), »Resultat- og evalueringstyring i en krisekontekst«, i Grøn, Caroline Howard, Hanne Foss Hansen og Mads Bøge Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag.
- Jørgensen, Torben Beck (1981), *Når staten skal spare*, København: Nyt fra Samfundsvidenskaberne.
- Jørgensen, Torben Beck og Poul Erik Mouritzen (2002), *Udgiftspolitik og Budgetlægning*, Århus: Systime.
- Kickert, Walter, Tiina Randma-Liiv og Rin Saavi (2014), »Fiscal Consolidation in Europe: A Comparative Analysis«, Cocops: www.cocops.eu.
- Koch, Kasper Teglgaard (2014), »Organisationskultur og resultatstyring«, i Mads Bøge Kristiansen, red., *Resultatstyring i den offentlige sektor*, København: Jurist- og Økonomforbundets Forlag.
- KREVI (2008), *Kommunale kontrakter i overblik. En kortlægning af intern kontraktstyringspraksis i kommunerne*, Århus: KREVI.
- March, James og Johan P. Olsen (1989), *Rediscovering Institutions, The Organisational Basis of Politics*, New York et al.: The Free Press.
- Pollitt, Christopher (2010), »Cuts and reforms – public services as we move into a new era«, *Society and Economy*, 32(1): 17-31.
- Pollitt, Christopher (2013), »The logics of performance management«, *Evaluation*, 19(4): 346-63.
- Rieper, Olaf, Niels Ejersbo og Peter Hansen (2000), »Målstyring i kommunerne«, i Marianne Antonsen og Torben Beck Jørgensen, red., *Forandringer i teori og praksis – skiftende billeder fra den offentlige sektor*, København: Jurist- og Økonomiforbundets Forlag.
- Siverbo, Sven og Tobias Johansson (2006), »Relative Performance Evaluation in Swedish Local Government«, *Financial Accountability & Management*, 22(3): 271-90.
- Sørensen, Eva Moll (2013), *A Window for Centralisation? The Danish »Structural Reform« and the measurement and steering of municipal service performance*, Ph.D. afhandling. København: Institut for Statskundskab.
- Sørensen, Eva Moll (2014a), »Statslig resultatmåling af kommunerne«, i Mads Bøge Kristiansen, red., *Resultatstyring i den offentlige sektor*, København: Jurist- og Økonomforbundets Forlag.
- Sørensen, Eva Moll (2014b), »Økonomisk styring af kommunerne i en krisetid«, i Caroline Grøn, Hanne Foss Hansen, og Mads Bøge Kristiansen, red., *Offentlig styring. Forandringer i krisetider*, København: Hans Reitzels Forlag.
- Wittrup, Jesper et al. (2013), *Notat om kommunal benchmarking: Service og produktivitet i 94 kommuner*, København: KORA

En effektevaluering af sanktioner på danske kommuners regnskaber, 2010-2013¹

Søren Kjær Foged, ph.d.-studerende ved Institut for Statskundskab, Københavns Universitet, skf@ifs.ku.dk

Denne artikel evaluerer de lovbaserede sanktioner på de kommunale regnskaber, der blev indført af regeringen i 2011. Evalueringen fokuserer på effekten i forhold til en række evalueringskriterier, der danner baggrund for de opstillede hypoteser. Hypoteserne vedrører kommunernes overholdelse af de budgetterede udgifter, hvor præcist kommunerne rammer deres budgetter og om effekterne varierer på tværs af kommunernes økonomiske situation og sektorer. Hypoteserne testes primært via multiple regressionsmodeller med data fra 2010 til 2013. Evalueringen finder, at kommunerne har forbedret budgetoverholdelsen i tråd med sanktionernes formål, men at kommunerne ikke har øget præcisionen af deres regnskaber. Derudover viser evalueringen, at sanktionerne har påvirket velstående kommuner mest, ligesom skole, dagtilbud og det specialiserede socialområde er de mest påvirkede sektorer.

Regnskabssanktionerne kalder på en evaluering

Formålet med artiklen er at levere en foreløbig evaluering af de statsligt indførte, lovbaserede sanktioner på de kommunale regnskaber. Disse sanktioner, som har været permanent gældende siden 2011, indebærer, at kommunerne modregnes i deres bloktilskud fra staten, hvis de for kommunerne under ét overskrider budgetterne. Sanktionerne har både et kollektivt og individuelt islæt, idet 60 pct. af overskridelsen modregnes individuelt, mens de resterende 40 pct. fordeles

kollektivt på de 98 kommuner på baggrund af indbyggertal. Relevansen af en evaluering af regnskabssanktionerne er for det første, at sanktionseffekterne, som led i den generelt tættere styring af kommunernes økonomi, har været heftigt debatteret i dansk forvaltningspolitik (Christensen og Lyndrup, 2013; KL, 2012; Økonomi- og Indenrigsministeriet, 2014; Olsen, 2013; Produktivitetskommissionen, 2013). Dertil kommer, at sanktionseffekter er interessant i et teoretisk perspektiv, da en analyse heraf bidrager til forståelsen af skatte- og udgiftslofter for lokalregeringer (Blom-Hansen, Bækgaard, og Serritzlew, 2014). Effekten af sådanne sanktionerede lofter er interessant i en dansk kontekst, hvor kommunerne står for cirka halvdelen af de offentlige forbrugsudgifter (Foged, Houlberg og Petersen, 2014: 102), og hvor både budgetunderskridelser eller -overskridelser derfor har store velfærdspolitiske konsekvenser. Eksempelvis beregnede Arbejderbevægelsens Erhvervsråd, at de offentlige budgetunderskridelser i 2013 beløb sig til 7 mia. DKK svarende til omtrent 16.000 offentlige ansatte, et tabt vækstpotentiale på 0,4 pct. af BNP og dermed mindre offentlig service. Af underforbruget på 7 mia. DKK stod kommunerne for de 5 mia. DKK (Arbejderbevægelsens Erhvervsråd, 2014). Med andre ord har sanktionerne i den grad mulige velfærds-mæs-

sige konsekvenser, om end disse konsekvenser nok lever en relativt skjult tilværelse i den enkelte borgers bevidsthed.

Netop nu synes udviklingen i forvaltningspolitikken at kalde på en evaluering af regnskabssanktionerne. Produktivitetskommissionen pegede således på, at sanktionerne kunne have en række uhensigtsmæssige effekter for kommunernes ressourceudnyttelse samt incitamenter til opsparing og investeringer (Produktivitetskommissionen, 2013). Denne opfordring om en diskussion af sanktionernes effekter blev dog affejet af regeringen (Økonomi- og Indenrigsministeriet, 2014), og heller ikke kommunernes interesseorganisation, KL, synes at have haft en ændring af sanktionsloven som en mærkesag på trods af, at mange individuelle kommuner har ytret modstand mod loven. Dette skyldes muligvis, at sanktionerne placerer både regering og KL mere centralt i styringen af kommunernes økonomi,² hvorimod de individuelle kommuner, som mærker potentielle sideeffekter af loven, kan have en svagere stemme i den offentlige debat. Den efterspurgte, men indtil nu fraværende, vurdering af regnskabssanktionerne gennemføres derfor her som en forskerdrevet evaluering.

Evalueringen fokuserer på effekten af regnskabssanktionerne, og evaluerer ikke andre aspekter af den strammere statslige styring af kommunernes økonomi, som har kunnet iagttages siden 2009. Gældende fra 2009 indførte regeringens således de første permanente lovfæstede sanktioner, da der blev indført kollektive budgetsanktioner, hvorved en samlet overskridelse af økonomiaftalen i kommunernes budgetlægning medførte et reduceret bloktilskud. Også gældende for 2009 indførtes en blanding af kollektive og individuelle sanktioner, hvis kommunerne under ét øgede skatterne. Disse sanktioner på budget og skat skærpedes i 2011, hvor det individuelle element i sanktionerne kom til at stå stærkere, og hvor også regnskabssanktioner blev en

del af styringsregimet. Det var således ikke længere nok, at kommunernes budgetlægning var inden for økonomiaftalen – regnskaberne skulle også overholde budgetterne! Med vedtagelsen af budgetloven i 2012, gældende fra 2014, blev den statslige styring strammet yderligere, idet udgiftslofter for kommunerne herefter skulle vedtages i Folketinget og reelt erstatte de årlige økonomiaftaler (Finansministeriet, 2012).

På relativt få år er der med andre ord sket en betydelig stramning af den statslige styring af kommunernes samlede skatte- og udgiftsniveau. Denne artikel evaluerer effekterne af den tættere styring af kommunernes udgiftsniveau ved at se på det hjørne, der vedrører effekten af regnskabssanktioner. Det skal tilføjes, at der i 2008 også var sanktioner på kommunernes regnskaber, men at disse var midlertidige og begrænset til det pågældende år.

Evalueringen ser på andet end budgetoverholdelse

Valg af evalueringsmodel og kriterier

Inden for evalueringsteori kan der skelnes en række evalueringsmodeller, som har betydning for undersøgelsesspørgsmål, evalueringskriterier og evalueringsdesign (Hansen, 1989, 2011). Valg af evalueringsmodel kan afhænge af formål, problemets karakter, karakteristika ved organisationen og organisationsfeltet samt om evalueringen har et forandrende eller støttende sigte (Hansen, 2011). Evalueringsmodel er her valgt ud fra formål samt datatilgængelighed. Formålet med evalueringen er således primært at *undersøge effekterne* af regnskabssanktionerne, herunder om disse har levet op til de politiske mål, og om sideeffekter kan iagttages. Sekundært vil artiklen ideelt *forklare de observerede effekter*. Hvor en undersøgelse af effekter lægger op til anvendelse af såkaldte resultatmodeller, vil en forklarende analyse ofte trække på procesmodeller og programteoretiske modeller, der kan belyse mekanismerne fra sanktionernes indførelse til de observerede effekter,

og eksempelvis identificere problemer i de foreslåede mekanismer eller sanktionernes implementering decentralt i kommunerne (Dahler-Larsen, 2009). Givet det tilgængelige datagrundlag fokuserer artiklen på effekterne, og den valgte evalueringsmodel er derfor en resultatmodel.

Under kategorien resultatmodeller kan der sondres mellem målopfyldelsesevaluering, der ser på om eksternt definerede mål er indfrie, og effektevaluering, der anvender et bredere sæt af kriterier til at vurdere sanktionerne. Artiklen gør i det følgende brug af begge former, da der både ses på effekter i forhold til de i loven opstillede mål samt en række øvrige kriterier, udvalgt af evaluatoren. Man kan kritisere effektevalueringer for, at parametre uden for det officielle formål, kan blive tilfældige, da der sjældent findes klare retningslinjer for kriterievalg. Denne udfordring håndteres her ved at fokusere på kriterier, som nævnes som vigtige af aktørerne på området (Hansen, 2011: 406) samt dimensioner, der i litteraturen beskrives som centrale.

Den første effektdimension er, om regnskabs-sanktionerne har modvirket kommunale budgetoverskridelser, hvilket var den politiske begrundelse for loven. Som led i henstillingen fra EU og den økonomiske genopretning blev det af regeringen således opfattet som afgørende, at de aftalte rammer blev overholdt (Retsinformation, 2011). For det andet undersøges, om effekter for kommunernes budgetoverholdelse varierer for forskellige kommuners økonomiske situation og sektorer. I loven var der nemlig kun fokus på kommunernes samlede budgetoverholdelse, men ikke overvejelser om, hvordan sanktionerne ville påvirke forskellige kommuner og opgaveområder. Dette spørgsmål undersøges her, da nogle kommuner og sektorer muligvis har forbedret budgetoverholdelsen mere end andre. For det tredje undersøges også, om kommunerne rammer tættere på deres budgetter efter sanktionernes indførelse sammenlignet med perio-

den før sanktionerne. Et regnskab, der rammer relativt tæt på budgettet, kan opfattes som ønskværdigt, da det dels udtrykker det politisk definerede serviceniveau i kommunerne, og dels vedrører præcisionen af statens makroøkonomiske styring via kommunerne (Blom-Hansen, 1998). Som tidligere omtalt kan eksempelvis et underforbrug have relativt store velfærdspolitiske implikationer for eksempelvis antal kommunale arbejdspladser og omfanget af service til borgerne. Endelig diskuteres effekterne for kommunernes interne organisering og styring, herunder kommunernes incitamenter til effektiv ressourceudnyttelse samt konsekvenser for kommunale aktørers relative indflydelse på udgiftspolitikken. Da datamaterialet ikke tillader en egentlig test af sidstnævnte dimension, opstilles dog ikke en egentlig hypotese herfor, og dimensionen behandles i stedet i diskussionsafsnittet.

Hypoteser

De følgende hypoteser er udledt på baggrund af evalueringskriterierne. Det første kriterium vedrører det politiske formål med loven om at undgå samlede budgetoverskridelser for kommunerne. Regnskabssanktionerne i 2011 medførte således to hovedændringer: For det første blev regnskabssanktioner lovbase-rede, hvor det tidligere ville bero på en konkret forhandling mellem regering og kommunerne, repræsenteret ved KL, om mulige reaktioner på en budgetoverskridelse (Juul, 2006: 205-206). Ud over at sanktionerne blev lovfæstede frem for forhandlingsbaserede, blev sanktionerne for det andet gjort delvist individuelle. Samlet forventes dette, i tråd med lovens formål, at øge den enkelte kommunes tilskyndelse til budgetoverholdelse, da udløsningen af sanktioner bliver gjort mere automatisk samtidig med at incitamenter til at spekulere i budgetoverskridelser mindskes via individuelle sanktioner. Hypotesen bliver dermed:

H₁: Lovbaserede og delvist individuelle regnskabssanktioner forbedrer kommunernes

budgetoverholdelse sammenlignet med et forhandlingsbaseret system.

Det andet evalueringskriterium vedrører ligeledes kommunernes budgetoverholdelse, men undersøger nu heterogene effekter. For heterogene effekter på tværs af kommunetyper forventes det, at regnskabssanktioner ikke mindst påvirker velstående kommuner, som kan have finansiering til at bruge flere penge end budgetteret. Hvor disse velstående kommuner førhen ville være tilbøjelige til at overskride budgetterne, fordi dette ikke ville gå nævneværdigt ud over kommunens økonomiske balance, betyder regnskabssanktionerne, at disse kommuner nu tvinges til budgetoverholdelse. Omvendt vil økonomisk pressede kommuner være mindre påvirkede af sanktionerne, da disse kommuner alligevel havde dårligere råd til at budgetoverskride også før sanktionernes indførelse, da dette vil bringe kommunen i økonomisk ubalance. Hypotesen bliver derfor:

H_{1a}: Lovbaserede og delvist individuelle regnskabssanktioner forbedrer kommunernes budgetoverholdelse, især for økonomisk velstående kommuner.

Angående heterogene effekter på tværs af sektorer forventes det, at regnskabssanktioner særligt vil påvirke områder med stærke bruger- og producentinteresser (Foged, Houlberg og Petersen, 2014; Serritzlew, 2005) samt sektorer, hvor høje enhedsomkostninger samt en stor og svært styrbar efterspørgsel kan give et opadgående udgiftspres. Årsagen til dette er, at uden troværdige sanktioner vil områder med stærke bruger- og personaleinteresser have held til at presse kommunalpolitikere til tillægsbevillinger over budgetåret. Sådanne områder er ikke mindst skole- og dagtilbudsområdet, og til dels ældreområdet (Serritzlew, 2005: 417). Ligeledes forventes det, at det specialiserede socialområde, som er kendetegnet ved en svært forudsigelig, ofte stigende, efterspørgsel og høje enhedsom-

kostninger, især vil forbedre budgetoverholdelsen, da budgetoverskridelser ikke længere er mulige i samme omfang under et sanktionsregime. Hypotesen bliver derfor:

H_{1b}: Lovbaserede og delvist individuelle regnskabssanktioner forbedrer kommunernes budgetoverholdelse, især for sektorer med stærke bruger- og personaleinteresser samt sektorer med høje enhedsomkostninger og en stor og svært styrbar efterspørgsel.

Det fjerde kriterium vedrører præcisionen i kommunernes regnskaber i forhold til budgetterne. Her forventes det, at kommunerne godt nok vil reducere udgifterne i tråd med lovgivningsformålet, men at de vil gøre dette i en sådan udstrækning, at de faktisk rammer længere fra budgetterne end under det forhandlingsbaserede system. Dette skyldes, at regnskabssanktionerne tilskynder til forsigtighed hos kommunerne, da disse vil være sikre på, at de ikke straffes økonomisk i det efterfølgende budgetår. Når alle kommuner tænker sådan, er resultatet en samlet underskridelse, som forventes længere fra det vedtagne budget end under det forhandlingsbaserede system. Hypotesen bliver dermed:

H₂: Lovbaserede og delvist individuelle regnskabssanktioner reducerer kommunernes budgetpræcision sammenlignet med et forhandlingsbaseret system.

Det er metodisk udfordrende at isolere effekten af sanktionerne

Identifikationsstrategi

Det er metodisk udfordrende at isolere effekten af regnskabssanktionerne. Dette skyldes, at sanktionerne ramte alle kommuner samtidigt, og vi derfor ikke kan adskille en stimuli- og kontrolgruppe. Ideelt set var sanktionerne således implementeret tilfældigt på den ene halvdel af kommunerne, og den kausale effekt ville herefter være forskellen i udviklingen i budgetoverholdelse mellem de to grupper (Angrist og Pischke, 2009; Hariri, 2012).

Da dette ikke kan lade sig gøre, anvendes en fremgangsmåde, hvor den relevante kontrolgruppe er de samme 98 kommuner identificeret før sanktionerne trådte i kraft. Den afgørende forudsætning for at estimere retvisende sanktionseffekter er herefter, om der kontrolleres for ændringer over tid, der er korreleret både med sanktionernes indførelse og kommunernes budgetoverholdelse. Den foretrukne model, der estimeres med OLS på data fra 2010 til 2013, kan skrives som:

$$\text{budgetafvigelse}_{it} = \beta_0 + \beta_1 2011_t + \beta_2 2012_t + \beta_3 2013_t + \beta_4 \text{Kontrol}_{it} + \varepsilon_{it}$$

Ligningen viser, at budgetafvigelsen er forskellen mellem den enkelte kommunes budget og regnskab i procent, mens de inkluderede årsummies viser forskellen i budgetafvigelsen mellem det angivne år og sammenligningsåret 2010, hvor sanktionerne endnu ikke var indført. Der fås således tre testresultater for sammenhængen mellem regnskabssanktioner og budgetoverholdelse i henholdsvis 2011, 2012 og 2013. Disse estimater er retvisende, hvis vektoren af kontrolvariable, opfanger alle relevante ændringer over tid, der både er korreleret med sanktioner og budgetoverholdelse. Dette er en relativt streng antagelse, da en række vigtige faktorer kan ændre sig over tid, herunder den økonomiske og politiske situation i kommunen, krisebevidstheden og sideløbende ændringer i økonomistyringen, der er uafhængige af sanktionerne³ (Foged og Sørensen, 2014).

Den økonomiske krise er her håndteret ved inklusion af kontrolvariable samt valg af sammenligningsåret 2010, hvor krisen allerede mærkes i de kommunale budgetter. Den politiske situation er ligeledes håndteret ved valg af sammenligningsåret 2010, således det samme byråd sidder i hele den observerede periode. Ved at anvende 2010 som reference er tanken ligeledes at få et konservativt sammenligningsgrundlag, da forskning har vist, at kommunerne »cykler« på deres budget-

overskridelser, således at de har tendens til at underskride deres budget de to første år i en byrådsperiode og overskride de to sidste år (Houlberg, 1999). Holder denne tendens, vil estimaterne undervurdere sanktionseffekten. Dette opvejes dog af, at det ikke har været muligt at finde dækkende kontrolvariable for mulige eksogene ændringer i kommunernes økonomistyring eller krisebevidsthed, hvorved der er risiko for, at effekten fra disse uobserverede faktorer opfanges af årsummierne og medfører en overvurdering af sanktionseffekterne. Modellen estimeres med klyngekorrigerede standardfejl med den enkelte kommune som klynge for i beregningen af usikkerheden omkring estimaterne at tage hensyn til, at observationer for den samme kommune ikke er uafhængige over tid.

Data: Registerbaserede kvantitative data støttet af interviews

Analysen baserer sig primært på data fra kommunale regnskaber og andre registre. Derudover er hypoteser og fortolkninger i den kvantitative analyse informeret af interviews med centrale aktører i 2013 og 2014, herunder borgmester, direktion og økonomichefer fra 6 kommuner, centrale embedsmænd i Finansministeriet og Økonomi- og Indenrigsministeriet samt repræsentanter fra KL.

Den afhængige variabel er operationaliseret som forskellen mellem budgetterede og regnskabsførte nettoserviceudgifter. Dette adskiller sig fra tidligere undersøgelser af danske kommuner, der har undersøgt kommunernes budgetoverskridelser ved det skattefinansierede driftsresultat, hvorved der både tages hensyn til eventuelle budgetafvigelser på indtægts- og udgiftssiden (Dalsgaard og Kristensen, 2009; Houlberg, 1999; Houlberg og Jensen, 2010; Krevi, 2008; Serritzlew, 2005). Da regnskabssanktionerne imidlertid kun gælder udgiftssiden, er nettoserviceudgifter det relevante mål i dette tilfælde. Der er dog grund til at tro, at resultaterne også er overførbare til

det samlede driftsresultat, da budgetafvigelse typisk skyldes ændringer på udgiftssiden (Houlberg og Jeppesen, 2006).

De inkluderede kontrolvariable vedrører for det første *den økonomiske situation*, som potentielt kan forbedre kommunernes budgetoverholdelse, fordi kommunerne har dårligere råd til budgetoverskridelser, når en krise rammer. Den økonomiske situation er operationaliseret ved det budgetterede driftsresultat og den årlige udvikling i budgetternes størrelse. Driftsresultatet er et udtryk for kommunens økonomiske situation, da driftsresultat gerne skal finansiere nettoanlægsudgifter og eventuelle rentebetalinger (Panduro og Lauriden, 2014). Et lavt driftsresultat kan tyde på, at kommunen er økonomisk presset, og fx må bruge af kassebeholdning eller lånefinansiere. Driftsresultatet måles som driftsresultat i budgettet, da driftsresultat i regnskabet påvirkes af eventuelle budgetafvigelser. Den økonomiske situation er også målt ved den årlige udvikling i budgettets størrelse sammenlignet med sidste års regnskab, hvor et reduceret budget alt andet lige vanskeliggør udgiftsoverholdelse. For det andet er der medtaget variable, der skal kontrollere for uventede *stød til de kommunale indtægter og udgiftsbehov samt »budgetpynt«*, hvor kommunerne kan manipulere konteringen af deres serviceudgifter for at undgå sanktioner. Afvigelser fra det forventede indtægtsniveau er målt ved forskellen mellem budgetterede og realiserede indtægter samt den årlige befolkningsudvikling, hvor ændringer afspejler sig i bloktilskuddets størrelse. Den årlige befolkningsudvikling kan desuden være en proxy for ændrede behov. For at undgå, at kommuner pynter på serviceudgifternes størrelse ved at kontere driftsudgifter som anlægsudgifter, inkluderes også en kontrolvariabel for budgetafvigelsen for nettoanlægsudgifterne. For det tredje kontrolleres for *politiske og administrative faktorer* i form af kommunalpolitikernes partifarve og antallet af administrationsprofessionelle. Kommunalpolitikernes

partifarve inkluderes som andelen af venstreorienterede byrådsmandater, operationaliseret som politikere fra Socialdemokraterne, Socialistisk Folkeparti og Enhedslisten, hvilket opfanger en eventuel partipolitisk forskel i tilbøjeligheden til budgetafvigelser. Derudover kontrolleres for antallet af administrationsprofessionelle pr. 10.000 indbyggere i kommunen, hvilket ideelt set opfanger ændringer i økonomistyringen i kommunen, som ikke er afledt af regnskabssanktionerne. Administrationsprofessionelle måles som medlemmer af Dansk Jurist- og Økonomforbund (DJØF).

Til sidst er der også kontrolleret for en række *strukturelle variable* af generel interesse i forhold til at forklare niveauet af budgetoverholdelse på tværs af kommuner. Disse variable er konstante over tid med undtagelse af kommunens økonomiske rammevilkår, som måles ved forholdet mellem kommunens udgiftsbehov og indtægtsgrundlag, og derved også kontrollerer for ændringer over tid i udgiftsbehov og indtægtsbase. De øvrige strukturvariable er indbyggertal og en dummyvariabel for henholdsvis små ø-kommuner og kommuner, der sammenlagdes i 2007. For at teste, om effekten af sanktionerne varierer på tværs af sektorområder, er der opgjort 9 sektorer, som er defineret ved de 8 sektorområder,⁴ som Økonomi- og Indenrigsministeriet, Finansministeriet og KL er enige om, samt kulturområdet. En oversigt over inkluderede variable, operationalisering og deskriptiv statistik findes i tabel 3 i appendikset.

Resultater: Sanktionerne har reduceret udgifterne, omfordelt penge mellem kommuner og sektorer, men ikke ført til større præcision

Ved første øjekast tyder noget på, at regnskabssanktionerne har haft en effekt på kommunernes budgetoverholdelse. Figur 1 viser således, at de tre år med regnskabssanktioner fra 2011 til 2013 har været fulgt af årlige budgetoverskud. Budgetoverskuddet i 2013 er

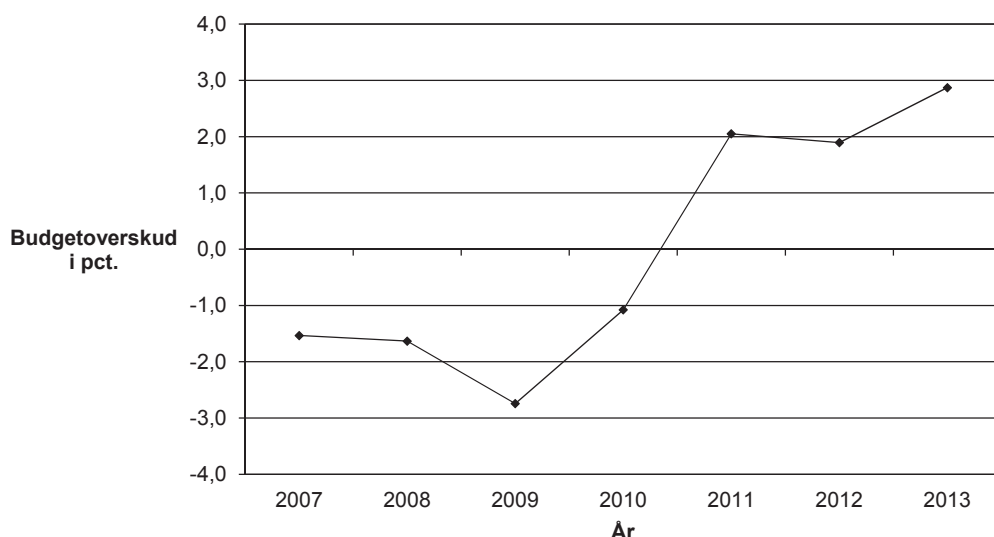
dog formentligt overvurderet, da tallet ikke er korrigeret for, at KL og regeringen i starten af 2013 aftalte at konvertere 2 mia. DKK fra service til anlæg samtidig med at lærerkonflikten også bidrog til lavere udgifter (KL, 2014). Tilsammen vurderes disse forbehold at udgøre omtrent halvdelen af budgetoverskuddet i 2013, hvilket leder til den observation, at der i sanktionsårene har været et årligt budgetoverskud i størrelsesordenen 1,5-2 pct. Ligeledes fremgår det, at budgetafvigelsen i sammenligningsåret 2010 på 1 pct. er tættere på budgettet end sanktionsårene. Figur 1 støtter således umiddelbart hypotese 1 og 2 før der er gennemført statistisk kontrol for andre mulige faktorer. I det følgende gennemføres disse analyser ved først at teste effekten på graden af budgetoverholdelse (H_1) og budgetpræcision (H_2), samt hvorvidt effekten varierer for velstående og mindre velstående kommuner (H_{1a}). Herefter testes, om effekten

af sanktionerne varierer på tværs af sektorer (H_{1b}).

Sanktionerne har forbedret budgetoverholdelsen, og mest for velstående kommuner

I tabel 1 model 1-5 testes sanktionernes effekt på kommunernes budgetoverholdelse ved løbende inklusion af kontrolvariable. I alle modeller er sanktionsårene korreleret med en signifikant forbedring af kommunernes budgetoverholdelse, og efter kontrol for den økonomiske krise i model 2 ændrer estimaterne sig kun marginalt. I den endelige model 5 fremgår det, at sammenlignet med 2010 er regnskabssanktioner associeret med en forbedring af budgetoverholdelsen på 3,2 procentpoint i 2011, 0,9 procentpoint i 2012 og 5,7 procentpoint i 2013. Igen skal der dog tages forbehold for den mulige overvurdering af 2013-estimatet. I forhold til kontrolvariablene viser model 5, at den økonomiske

Figur 1: Gennemsnitligt budgetoverskud for nettoserviceudgifter, 2007-2013 (N=98)



Note: Uvægtet gennemsnit. Budgetoverskuddet i 2013 er formentlig overvurderet, da tallet ikke er korrigeret for, at KL og regeringen i 2013 aftalte at konvertere 2 mia. DKK fra service til anlæg, samtidig med at en lockout af lærerne også bidrog til sparede udgifter. Dette forklarer skønsmæssigt halvdelen af kommunernes samlede budgetunderskridelse på 4,9 mia. DKK i 2013.

krise har den forventede effekt på budgetoverholdelsen, idet et relativt lavt driftsresultat reducerer kommunens udgifter i forhold til budgettet, mens det samtidig er sværere at overholde budgetterne, hvis disse tilpasses nedad i forhold til året før (udtrykt ved den årlige budgetudvikling).

Derudover er lavere indtægter end budgetteret korreleret med bedre budgetoverholdelse, hvilket muligvis skyldes, at kommuner, som er på vej til at overskride deres serviceudgifter (og også ender med en relativt dårlig budgetoverholdelse), formår at øge deres indtægter mere end budgetteret, hvorved stigende indtægter derved kompenserer for de overskredne budgetter i årsresultatet. Endelig er budgetoverholdelse for anlægsudgifter positivt korreleret med budgetoverholdelse for serviceudgifterne, hvilket enten kan skyldes, at den enkelte kommune enten agerer relativt ensartet for både service og anlæg eller, at kommuner »overvælter« en del af en budgetoverskridelse fra drift til anlæg. De resterende kontrolvariable er alle insignifikante.

I model 6 gennemføres en robusthedstest med fixed effects, der som forventet giver lignende, positive og signifikante estimater for sammenhængen mellem regnskabssanktioner og budgetoverholdelse. Også for hovedparten af kontrolvariablene viser modellen samme effekt, selvom der nu kun anvendes variation over tid og ikke, som i OLS-modellerne, både variation på tværs af kommuner og over tid. Dog estimerer fixed effects modellen en signifikant, negativ sammenhæng mellem antallet af administrationsprofessionelle og budgetoverholdelse, hvilket formentlig skal tolkes som omvendt kausalitet, hvor udgiftsoverskridelser fører til ansættelse af flere administrationsprofessionelle. Opsummerende støtter model 1-6 hypotese H_1 om at regnskabssanktioner forbedrer kommunernes budgetoverholdelse sammenlignet med et forhandlingsbaseret system.

Hypotese H_{1a} , der forudsagde, at effekten af sanktionerne var størst for velstående kommuner, er testet i model 7-8 ved en interaktion mellem årsummierne og driftsresultatet. I model 7 er hypotesen testet enkeltvist for sanktionsårene, mens model 8 gennemfører en samlet test for sanktionsårene. Model 7 viser, at effekten af sanktionerne som forventet stiger for kommuner med et højere driftsresultat, men at effekten dog kun er signifikant i 2012. Når sanktionsårene slås sammen og sammenlignes med 2010 i model 8, er sammenhængen som forventet positiv og signifikant. Denne effekt i model 8 kan omskrives til, at kommunen med et driftsresultat på 373 DKK pr. indbygger (svarende til det 10. percentil) forventes at øge sin budgetoverholdelse med 2,9 procentpoint, mens kommunen med et driftsresultat på 2.772 DKK pr. indbygger (svarende til det 90. percentil) forventes at øge budgetoverholdelsen med 3,8 procentpoint. Opsummerende støtter model 7-8 dermed i overvejende grad hypotese H_{1a} , der siger, at regnskabssanktioner især forbedrer budgetoverholdelsen for økonomisk velstående kommuner. Forskellen på velstående og økonomisk udfordrede kommuner indikerer i øvrigt også, at der er en selvstændig sanktionseffekt uafhængig af den økonomiske krise, da de kommuner, som er mindst økonomisk udfordrede er mest påvirkede af sanktionerne.

I tabel 1 model 9 og 10 findes sluttelig en test af kommunernes budgetpræcision. Budgetpræcision måles ved at udregne den absolute værdi af kommunernes budgetafvigelse, således variablen varierer fra 0 til 6,3 pct., hvor 0 indikerer fuldstændig overensstemmelse mellem budget og regnskab. Et negativt estimat for årsummierne indikerer nu, at kommunerne rammer tættere på budgettet sammenlignet med referenceåret. I model 9 er referenceåret 2010, og det fremgår heraf, at kommunerne var signifikant dårligere til at ramme budgettet i 2011 og 2013 sammenlignet med 2010, hvorimod budgetpræcisionen

Tabel 1: Afvigelse mellem budget og regnskab for nettoserviceudgifter i pct., 2010-2013 (N=98)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Andel af totale nettodriftsudgifter (pct.)	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.	Budgetafvigelse i pct.
Ardsammes										
2011	3,129*** (16,18)	3,379*** (15,64)	3,273*** (13,08)	3,263*** (13,02)	3,210*** (12,23)	3,430*** (13,81)	2,997*** (6,31)	-0,499*** (-2,97)	0,641*** (2,90)	0,209 (0,78)
2012	2,671*** (13,22)	0,943*** (2,97)	0,825** (2,59)	0,819** (2,58)	0,929*** (2,72)	2,506*** (5,79)	0,486 (1,28)	-0,00332 (-0,22)	-0,714** (-2,56)	-0,619** (-2,05)
2013	3,948*** (18,85)	5,304*** (19,31)	5,819*** (22,03)	5,808*** (22,10)	5,653*** (18,29)	5,218*** (14,59)	5,239*** (10,16)	2,361*** (7,42)	1,839*** (6,49)	1,839*** (6,49)
Økonomisk situation										
Driftsresultat 1.000 DKK (budget)	-0,0800 (-1,02)	-0,0800 (-1,02)	-0,153** (-2,10)	-0,158** (-2,09)	-0,170** (-2,28)	-0,261** (-2,52)	-0,346** (-2,21)	-0,499*** (-2,97)	0,0328 (0,44)	0,0662 (0,90)
Arlig budgetudvikling i pct.	0,320*** (7,19)	0,320*** (7,19)	0,365*** (8,51)	0,364*** (8,47)	0,340*** (6,88)	0,130** (2,29)	0,334*** (6,38)	0,206*** (5,01)	0,133*** (3,76)	0,133*** (3,76)
Stød til indtægter og behov										
Indtægtsafvigelse i pct.	0,206** (2,21)	0,206** (2,21)	0,212** (2,26)	0,212** (2,26)	0,201** (2,11)	0,138 (1,15)	0,197** (2,05)	-0,0983 (-1,46)	0,122 (1,39)	0,191 (1,61)
Arlig betjningsudvikling i pct.	0,351*** (2,93)	0,351*** (2,93)	0,330*** (2,39)	0,330*** (2,39)	0,125 (0,65)	-0,206 (-0,84)	0,118 (0,64)	-0,159 (-0,65)	0,323** (2,21)	0,383*** (2,83)
Budgetbynt										
Budgetafvigelse for anlægsgudgifter i pct.	0,00063*** (6,40)	0,00063*** (6,40)	0,000571*** (5,81)	0,000571*** (5,81)	0,000386*** (2,95)	0,000417*** (3,48)	0,000381*** (2,55)	0,000311 (1,40)	0,000425** (2,24)	0,000463** (2,47)
Politiske og administrative faktorer										
Andel venstreorienterede mandater	0,00831 (0,39)	0,00831 (0,39)	0,00129 (0,18)	0,00129 (0,18)	-0,00241 (-0,31)	-0,248** (-2,32)	-0,00218 (-0,28)	-0,00465 (-0,47)	-0,00367 (-0,61)	-0,00861 (-1,48)
Administrationsprofessionelle pr. 10.000 indb.	-0,00175 (-0,07)	-0,00175 (-0,07)	-0,00175 (-0,07)	-0,00175 (-0,07)	-0,00175 (-0,07)	-0,00195 (-0,08)	-0,00195 (-0,08)	-0,00176 (0,05)	-0,00170 (-0,64)	-0,0208 (-0,63)
Strukturelle vilkår										
Indbyggertal (ln)	0,0835 (0,54)	0,0835 (0,54)	0,0835 (0,54)	0,0835 (0,54)	0,0835 (0,54)	0,0748 (0,50)	0,0748 (0,50)	0,0634 (0,34)	0,0646 (0,36)	-0,0561 (-0,93)
Sammenlagt (dummy)	-0,268 (-1,12)	-0,268 (-1,12)	-0,268 (-1,12)	-0,268 (-1,12)	-0,268 (-1,12)	-0,280 (-1,16)	-0,280 (-1,16)	-0,381 (-1,23)	-0,0688 (-0,35)	0,0170 (0,08)
Lille ø-kommune (dummy)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	-0,997 (-1,50)	0,650 (1,15)
Økonomiske ramnevilkår	-2,499 (-0,90)	-2,499 (-0,90)	-2,499 (-0,90)	-2,499 (-0,90)	-2,499 (-0,90)	15,62* (1,75)	-1,867 (-0,68)	-3,246 (-1,04)	0,761 (0,39)	2,272 (1,04)
Interaktioner mellem årdsammes og økonomisk situation										
2011 X Driftsresultat i 1.000 DKK (budget)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)	0,143 (0,62)
2012 X Driftsresultat i 1.000 DKK (budget)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)	0,310* (1,77)
2013 X Driftsresultat i 1.000 DKK (budget)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)	1,27 (1,27)
Alle senklionsår (dummy for 2011-2013) X Driftsresultat i 1.000 DKK (budget)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)	2,754*** (7,27)
Konstant	-1,077*** (-5,39)	-0,585*** (-2,95)	-0,600*** (-3,30)	-0,722* (-1,74)	-0,296 (-0,20)	-3,487** (-1,13)	-0,131 (-0,09)	0,685 (0,35)	1,099 (0,60)	2,276 (1,37)
Estimationsmetode										
2010-2013	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)	392 (5,39)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62)	0,682 (19,69)	0,614 (16,62)	0,506 (12,27)	0,208 (5,39)	0,115 (3,76)
2010-2013	0,482 (12,27)	0,569 (15,94)	0,598 (16,62)	0,569 (15,94)	0,611 (16,62					

var signifikant bedre i 2012. For at give kommunerne de bedste betingelser for at ramme lige så præcist som i referenceårene, består referencekategorien i model 10 både af 2009 og 2010, da 2009 var kendetegnet ved store budgetoverskridelser (se figur 1). Resultatet af denne test viser, at der ikke synes at være en entydig sammenhæng mellem sanktioner og budgetpræcision, da kommunerne kom signifikant tættere på budgettet i 2012, men omvendt signifikant længere fra budgettet i 2013, mens der ikke var en signifikant sammenhæng i 2011. Opsummerende kan det indtil videre hverken konkluderes, at regnskabssanktionerne har gjort kommunerne væsentligt bedre eller dårligere til at ramme budgetterne. Sikkert er det imidlertid, at hvor kommunerne førhen typisk overskred deres budgetter, er normalen nu en underskridelse, hvilket som tidligere nævnt har betydelige velfærdspolitiske implikationer.

Sanktionerne har påvirket skole, dagtilbud og det specialiserede socialområde mest

Angående den sidste hypotese H_{1b} var forventningen, at sektorer med stærke bruger- og producentinteresser samt sektorer med høje enhedsomkostninger og en stor og svært styrbar efterspørgsel ville forbedre deres budgetoverholdelse mere end andre sektorer. Hypotesen er testet på 9 kommunale sektorer, der tilsammen udgør 88 pct. af serviceudgifterne. Det var således forventet, at områder med stærke bruger- og personaleinteresser, skole, dagtilbud og ældre samt det svært styrbare specialiserede socialområde ville forbedre budgetoverholdelsen mere end sundhed, kultur, det tekniske område og administrationsområdet. Denne forventning er undersøgt ved at specificere 9 modeller med budgetafvigelsen på sektorområdet som afhængig variabel. Resultaterne herfra er vist i tabel 2. For det tekniske område er der specificeret to modeller, hvoraf den foretrukne anvender 2009 frem for 2010 som sammenligningssår, da budgetafvigelsen i 2010 på det

tekniske område synes at være en outlier (se figur 3 i appendikset).

Overordnet støtter resultaterne de opstillede hypoteser, da skole, dagtilbud, udsatte børn og unge samt udsatte voksne og handicappede alle viser en signifikant forbedring af budgetoverholdelsen i alle sanktionsår, 2011-2013. Ydermere er sammenhængen mellem sanktioner og budgetoverholdelse som forventet mindre tydelig for sektorerne, sundhed, kultur, det tekniske område (med 2009 som referenceår) og administrationsområdet. For sundhed observeres insignifikante effekter bortset fra i 2013, hvor kommunernes budgetoverholdelse faktisk er dårligere sammenlignet med 2010, mens der for kultur, det tekniske område (med 2009 som referenceår) og administration kun observeres signifikante sanktionseffekter i 2011. Hypotese H_{1b} nyder således opbakning, når der ses bort fra ældreområdet, hvor der mod forventning ikke identificeres signifikante ændringer i nogle af årene.

Figur 2 i appendikset indikerer, at de observerede sektorforskelle i sanktionseffekterne i tabel 2 dækker over en konvergens i budgetafvigelse, således sektorer, der tidligere overskred deres budgetter mere end øvrige områder, nu overholder budgetterne i samme omfang som andre sektorer. En formel signifikantest af, hvorvidt sanktionseffekten varierer på tværs af sektorer, bekræfter, at effekten har været størst for udsatte børn og unge, efterfulgt af skole, dagtilbud, administration samt socialt udsatte voksne og handicappede (tabel 4 i appendikset). Effekten af administration drives dog alene af effekten i 2011 jf. tabel 2.

Opsummerende støtter tabel 2 således hypotese H_{1b} , når der ses bort fra ældreområdet, idet den største effekt af regnskabssanktionerne har været for områder med stærke bruger- og personaleinteresser samt områder med høje enhedsomkostninger og en stigende og svært styrbar efterspørgsel. Dette viser, at de velfærdspolitiske implikationer ikke kun

Tabel 2: Sektorfordelt afvigelse mellem budget og regnskab for netroudgifter i pct., 2010-2013 (N=98)

	(1) Skole	(2) Dagtilbud	(3) Ældre	(4) Sundhed	(5) Udsatte børn og unge	(6) Udsatte voksne	(7) Kultur	(8) Teknisk område	(9) Teknisk område	(10) Administration
Andel af totale netoserviceudgifter (88.0 pct.)	10.9	17.1	15.3	2.2	6.9	12.3	2.8	6.1	6.1	14.6
Årsdømmes										
2010	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
2011	2.235*** (4.56)	2.011*** (1.65)	1.051 (1.65)	1.477 (1.46)	9.263*** (5.13)	1.095*** (4.41)	1.016* (1.66)	8.535*** (5.63)	-5.346*** (4.18)	4.375*** (5.66)
2012	1.031** (2.39)	1.398*** (3.04)	-0.132 (2.39)	-1.286 (-1.06)	4.157*** (2.88)	0.585*** (2.89)	-0.251 (-0.44)	3.167** (2.61)	-2.242 (-1.66)	0.965 (1.32)
2013	3.262*** (6.48)	1.446*** (2.75)	0.225 (0.32)	-1.756* (-1.89)	3.182* (1.95)	1.361*** (5.29)	0.245 (0.35)	3.401*** (2.87)	-2.064 (-1.46)	0.359 (0.45)
Økonomisk situation										
Driftsresultat i 1.000 DKK (budget)	-0.0929 (-0.88)	-0.146 (-0.82)	-0.208 (-1.09)	0.215 (0.70)	0.473 (0.83)	-0.135 (-1.59)	0.205 (0.97)	-1.007** (-2.38)	-0.899** (-2.12)	0.237 (1.07)
Årlig budgetudvikling i sektor i pct.	0.555*** (9.89)	0.577*** (11.63)	0.450** (8.60)	0.524*** (5.29)	0.795*** (12.48)	0.00165 (0.09)	0.176*** (3.30)	0.319*** (4.81)	0.324*** (5.48)	0.707*** (22.16)
Sted til indtægter og behov										
Indtægtsafvigelse i pct.	0.134 (0.72)	0.386** (2.10)	0.903*** (3.68)	-1.008* (-2.48)	0.593 (0.70)	0.0592 (0.67)	0.207 (0.72)	-0.0169 (-0.03)	0.0374 (0.12)	0.187 (0.57)
Årlig befolkningsudvikling i pct.										
Årlig elevudvikling i pct. (antal 7-16 årige)										
Årlig børneudvikling i pct. (antal 0-6 årige)	0.0383 (0.29)									
Årlig ældreudvikling i pct. (antal 80+)		-0.380*** (-5.61)								
Årlig børne- og ungeudvikling i pct. (antal 0-17 årige)			-1.910 (-0.47)		0.389 (0.74)					
Budgetbyrnt										
Budgetafvigelse for anlægsudgifter i pct.	-0.000417 (-1.22)	0.000389 (1.21)	-0.000121 (-0.35)	0.00238* (1.98)	0.00472 (0.77)	-0.000215 (-0.37)	0.00305*** (6.01)	0.000588 (0.34)	0.000751 (0.45)	0.00198*** (4.58)
Politiske og administrative faktorer										
Andel af vesterorienterede mandater	0.00981 (0.72)	0.0251 (1.55)	-0.00894 (-0.55)	-0.0166 (-0.66)	0.0112 (0.30)	-0.0120 (-1.52)	-0.0120 (0.64)	0.0384 (0.78)	0.0271 (0.60)	0.00512 (0.27)
Administrationsprofessionelle pr. 10.000 indb.	0.00131 (0.04)	0.0482 (0.68)	0.0200 (0.31)	-0.179 (-1.39)	-0.167 (-1.01)	-0.0231 (-0.67)	0.0870 (0.78)	0.184 (1.32)	0.142 (1.10)	0.0291 (0.42)
Strukturelle vilkår										
Økonomiske rammevilkår	-2.452 (-0.54)	-9.528* (-1.98)	-5.075 (-0.75)	7.777 (0.66)	-11.54 (-0.61)	-2.550 (-0.81)	0.407 (0.06)	-24.31 (-1.42)	-33.39* (-1.95)	-5.995 (-0.70)
Indbyrdes (ln)	-0.197 (-0.96)	0.199 (0.67)	-0.378 (-0.93)	0.273 (0.44)	0.278 (0.30)	-0.171 (-1.03)	-0.405 (-0.95)	-0.467 (-1.04)	-0.467 (-0.53)	1.492*** (4.37)
Sammenlagt (dummy)	0.0695 (0.17)	0.485 (0.86)	-1.533*** (-3.23)	-0.367 (-0.40)	-0.839 (-0.77)	-0.463* (-1.66)	-0.281 (-0.34)	1.476 (1.11)	1.336 (1.12)	0.741 (1.34)
Lille økonomi (dummy)	-0.893 (-0.96)	0.652 (0.66)	1.012 (0.95)	-1.551 (-0.78)	-5.035 (-0.78)	-0.0397 (-0.08)	-0.626 (-0.37)	-3.876 (-1.35)	-3.239 (-1.26)	2.930 (1.64)
Konstant	2.807 (1.08)	-0.564 (-0.15)	7.171* (1.74)	-2.698 (-0.40)	-2.427 (-0.25)	3.686** (2.18)	2.739 (0.60)	9.904 (1.22)	14.68* (1.77)	-16.93*** (-4.27)
N	392	392	392	392	392	392	392	392	487	392
F	0.505	0.473	0.307	0.296	0.434	0.158	0.0816	0.280	0.268	0.585
E	26.11	18.10	14.49	4.973	35.05	6.115	11.32	10.22	8.700	41.39
Estimationsmetode	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Indspilte	2010-2013	2010-2013	2010-2013	2010-2013	2010-2013	2010-2013	2010-2013	2010-2013	2008-2013	2010-2013

Note: * signifikansniveau 0.10, ** signifikansniveau 0.05, *** signifikansniveau 0.01. Klyngekorrigerede standardfej med kommune som klynge. t-test i parentes, 2013-priser.

er markante ved en overgang fra en situation med generelle budgetoverskridelser til generelle underskridelser, men at disse effekter ydermere er fordelt ujævnt, således det specialiserede socialområde, skole og dagtilbud har været mest påvirket med større reduktioner i ansatte og den leverede service, som en formodet følge, sammenlignet med en situation uden sanktioner. Omvendt er budgetafvigelseernes størrelse generelt blevet mere ens på tværs af sektorer, hvor det førhen var de samme sektorer, som overskred budgetterne år efter år (Serritzlew, 2005).

I forhold til de inkluderede kontrolvariable i tabel 2 fremgår det, at ikke mindst udviklingen i budgettets størrelse sammenlignet med sidste års resultat er en vigtig prædikator for om budgettet overskrides eller underskrides. På alle sektorer undtagen det specialiserede voksenområde er det således tilfældet, at en reduktion i budgettet i forhold til sidste år også gør det sværere at holde sig inden for disse rammer. Derudover fremgår det, at det især er for dagtilbud og ældreområdet, at der er en sammenhæng mellem ekstraintægter og budgetoverskridelser, hvilket muligvis skyldes, at kommunerne for disse to sektorer øger indtægterne, hvis der er optræk til en udgiftsoverskridelse. Budgetpynt forstået som korrelation mellem budgetafvigelser på drift og anlæg, synes primært at finde sted for sundhed, kultur og administration, hvor det måske er nemmest at konvertere udgifterne mellem drift og anlæg. Endelig er store kommuner væsentligt bedre til at overholde deres budgetterede administrationsudgifter, mens kommuner der sammenlagdes i 2007, har haft vanskeligere ved at overholde deres ældrebudgetter sammenlignet med fortsætterkommunerne, hvilket kan skyldes, at der efter strukturreformen er taget flere politisk-geografiske hensyn i sammenlægningskommuner (Kjær, 2012). Eksempelvis angiver flere interviewpersoner fra sammenlægningskommunerne, at der ofte er stærke politiske interesser forbundet med ældretilbud i kom-

munens yderområder, hvilket politisk kan gøre det sværere at forhindre tillægsbevillinger.

Diskussion: Hvad med effekter for organisering og styring?

Den sidste effektdimension vedrører kommunernes organisering og styring. Effekterne for kommunernes organisering og styring kan både belyses fra et perspektiv, der lægger vægt på effektiv ressourceudnyttelse og en synsvinkel, hvor kommunale aktørers relative indflydelse er i centrum. Med effektiv ressourceudnyttelse forstås her, hvorvidt regnskabssanktioner har påvirket målsætningen om at opnå de bedste faglige resultater med færrest mulige ressourcer (Produktivitetskommissionen, 2013).

En sådan ambition kræver først og fremmest viden om de anvendte ressourcer, hvilket regnskabssanktionerne muligvis har bidraget til via en oprustning af kommunernes økonomistyring (Finansministeriet, Økonomi- og Indenrigsministeriet og KL, 2013; Foged og Sørensen, 2014; Krevi, 2012). Hvor direkte sammenhængen er mellem sanktioner og økonomistyring kan naturligvis debatteres, men interviewpersonerne synes at bekræfte, at sanktionerne sammen med en forværret økonomisk situation har bidraget til en opprioritering af økonomistyringen, bl.a. via hyppigere budgetopfølgninger og et øget fokus på prissætning af aktiviteter. Omvendt giver flere informanter udtryk for, at den store opmærksomhed på økonomistyringen, til tider har været fulgt af en nedprioritering af styringen efter kvantificerbare mål for servicekvalitet (Sørensen og Foged, 2014). Opfattes viden om resultater, foruden viden om medgående ressourcer, som en forudsætning for bedre ressourceudnyttelse, har sanktionerne således ikke på dette punkt bidraget til bedre ressourcestyring.

En central sideeffekt af regnskabssanktionerne er desuden, at sanktionerne muligvis kan have medført en mindre hensigtsmæssig ressource-

ceudnyttelse. Blandt andet kan sanktionerne have reduceret incitamentet til opsparing og investeringer, da flere kommuner har indført buffere i budgettet samt indskrænket decentrale institutioners overførselsadgang, fx ved kvantitative begrænsninger på overførslers størrelse eller indefrysning af opsparing i enkeltår. Dette kan give anledning til såkaldt »benzinafbrænding«, hvor kommunen ikke anvender ressourcer på den opgave eller det tidspunkt, hvor behovet/effekten er størst, men derimod anvender sine ressourcer med sigte på at undgå, at det fremtidige budget beskæres (Produktivitetskommissionen, 2013). Eksempelvis kan kommunen indlægge buffere i budgettet, som så »brændes af« sidst på året, når man er sikker på, at man ikke overskrider sit budget og tilsvarende kan decentrale institutioner, der har fået begrænset sin mulighed for at overføre forbrugte midler mellem budgetår, miste tilskyndelsen til opsparing og derimod spekulere i at anvende alle ressourcer inden budgetårets udløb. Ligeledes nævner flere informanter, at sanktionerne kan have bidraget til en »egernmentalitet« hos de enkelte budgetholdere, hvor kravet om budgetoverholdelse er blevet så dominerende, at evnen til at flytte midler til trængte områder inden for budgetåret begrænses.

Hvor nettoeffekten for kommunens ressourcestyring således er omtvistet, er informanterne mere entydige angående kommunale aktørers indflydelse på udgiftspolitikken, hvor sanktionerne vurderes at have bidraget til at styrke »udgiftsvogterne« på bekostning af »udgiftsadvokaterne« (Christiansen, 2008). Informanterne beskriver således, at sanktionerne generelt har styrket borgmesteren og økonomiudvalgets mandat ved at øge troværdigheden af de udmeldte rammer over for fagudvalgene. Også på dette område kan evalueringen dog kun foreslå mekanismer mellem sanktioner og de observerede effekter, og en videre evaluering kunne fokusere på disse mulige mekanismer samt eventuelt trække på en praksisorienteret evalueringstil-

gang (Schwandt, 2005). Ved en sådan tilgang kunne evaluator facilitere refleksion hos budgetholderne om, hvad god økonomistyring er, og hvordan god økonomistyring følges inden for regnskabssanktionernes rammer.

Konklusion: Sanktionerne har reduceret kommunernes udgifter og omfordelt penge på tværs af kommuner og sektorer

Denne artikel har gennemført en effektvurdering af regnskabssanktionerne for danske kommuner, der trådte i kraft i 2011. Evalueringen har fokuseret på opfyldelsen af lovens officielle formål om at undgå samlede budgetoverskridelser samt en række mulige sideeffekter af sanktionerne, herunder budgetpræcision, effekter for kommuner med forskellig økonomisk situation, effekter på forskellige sektorer og endelig effekter for organisering og styring. Evalueringen konkluderer, at regnskabssanktionerne har forbedret kommunernes overholdelse af de budgetterede serviceudgifter med det forbehold, at evalueringsdesignet ikke kontrollerer for ændringer i fx den generelle krisebevidsthed. Evalueringen finder desuden, at velstående kommuner har forbedret budgetoverholdelsen mere end kommuner, der er udfordret på driftsresultatet. Ligeledes har sanktionerne haft den største effekt på sektorer med stærke bruger- og personaleinteresser i form af skole og dagtilbud samt det specialiserede socialområde, hvor enhedsomkostningerne er høje og efterspørgslen stor og svært styrbar. Endelig finder evalueringen ikke nogen konsistent ændring i kommunernes budgetpræcision før og efter sanktionernes indførelse. Før regnskabssanktionerne overskred kommunerne typisk deres udgiftsbudget, mens dette nu er omvendt, men uden at den absolutte afvigelse fra budgettet er signifikant anderledes end før sanktionerne. De velfærdsmæssige implikationer af at gå fra en situation med budgetoverskridelser til en situation med underskridelser er dog potentielt særdeles markante, da en sådan udgiftsreduktion påvirker antallet af kommunale arbejdspladser og om-

fanget af den udførte service til borgerne. Angående effekterne for organisering og styring indikerer evalueringen, at sanktionerne har styrket kommunens budgetvognere i form af borgmester og økonomiudvalg på bekostning

af budgetadvokaterne i skikkelse af de politiske fagudvalg, ved at gøre budgetrammerne mere troværdige. Der lægges op til yderligere forskning i sanktionernes betydning for en hensigtsmæssig ressourceudnyttelse.

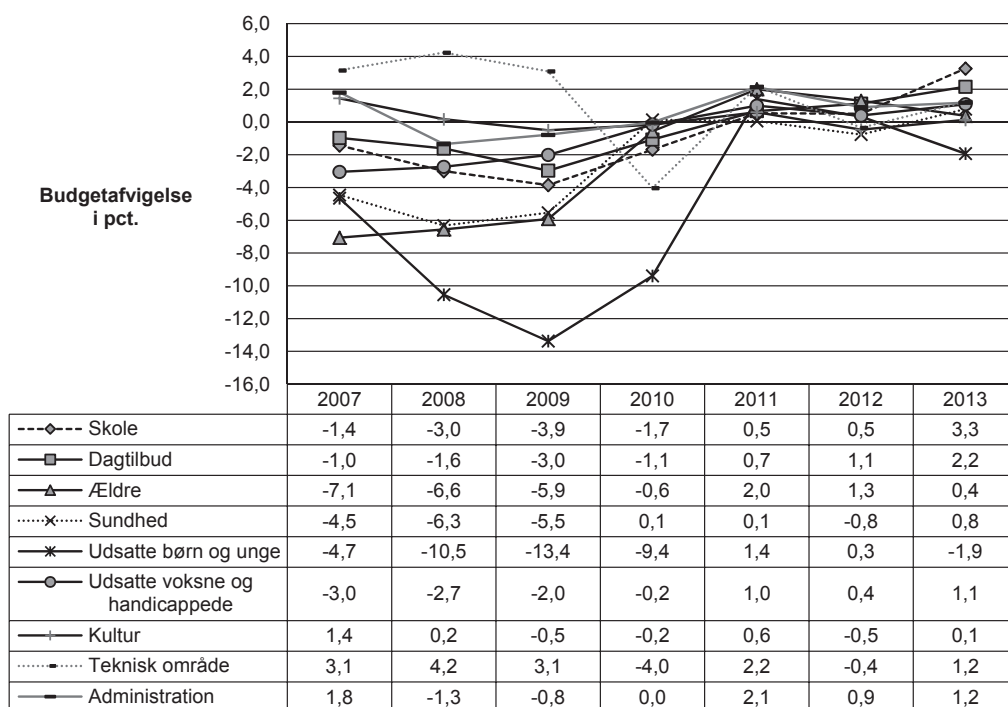
Appendiks

Tabel 3: Deskriptiv statistik og definitioner af variable, 2010-2013 (N=98)

Variabel	Observationer	Middel	Std.afv.	Min	Max	Definition	Kilde
Budgetafvigelse for service i pct.	392	1.4	2.2	-5.7	6.3	Budget fratrukket regnskab for nettoserviceudgifter jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Nøgletal for kommunernes økonomiske styring (KREVI).
Budgetafvigelse for skole i pct.	392	0.7	3.6	-24.1	14.2	Budget fratrukket regnskab for nettoudgifter til skole jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for dagtilbud i pct.	392	0.7	4.0	-10.4	18.2	Budget fratrukket regnskab for nettoudgifter til dagtilbud jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for ældre i pct.	392	0.8	4.7	-18.5	23.2	Budget fratrukket regnskab for nettoudgifter til ældre jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for sundhed i pct.	392	0.1	7.6	-61.7	27.9	Budget fratrukket regnskab for nettoudgifter til sundhed jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for udsatte børn og unge i pct.	392	-2.4	12.3	-85.7	44.6	Budget fratrukket regnskab for nettoudgifter til udsatte børn og unge jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for udsatte voksne og handicappede i pct.	392	0.6	1.6	-4.5	8.5	Budget fratrukket regnskab for nettoudgifter til udsatte voksne og handicappede jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for kultur i pct.	392	0.0	4.4	-26.9	18.4	Budget fratrukket regnskab for nettoudgifter til kultur defineret ved funktion 3.32.50, 3.35.60-3.35.64, 3.38.70-3.38.76 og 3.38.78 i hele perioden.	Danmarks Statistik
Budgetafvigelse for teknisk område i pct.	392	-0.2	8.9	-46.4	32.2	Budget fratrukket regnskab for nettoudgifter til teknisk område jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Budgetafvigelse for administration i pct.	392	1.0	6.7	-22.2	27.4	Budget fratrukket regnskab for nettoudgifter til administration jf. KL's definition af serviceområder i 2013 anvendt i hele perioden.	Danmarks Statistik
Driftsresultat i DKK (budget)	392	1649	1053	-1668	7018	Budgetteret skattefinansieret driftsresultat, dvs. budgetterede indtægter fra skatter, tilskud og nettoerenter fratrukket nettodriftsudgifter.	Nøgletal for kommunernes økonomiske styring (KREVI).
Årlig budgetudvikling for serviceudgifter i pct.	392	-0.8	4.3	-9.9	9.9	Budget i året fratrukket sidste års regnskab på område.	Nøgletal for kommunernes økonomiske styring (KREVI).
Årlig budgetudvikling for skole i pct.	392	-2.3	3.6	-12.5	14.0	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for dagtilbud i pct.	392	-2.7	4.3	-15.3	9.7	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for ældre i pct.	392	0.5	5.1	-15.3	22.8	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for sundhed i pct.	392	1.5	7.1	-19.9	60.4	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for udsatte børn og unge i pct.	392	-2.6	9.0	-34.7	29.6	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for udsatte voksne og handicappede i pct.	392	1.3	6.1	-26.8	19.9	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik

Variabel	Observationer	Middel	Std.afv.	Min	Max	Definition	Kilde
Årlig budgetudvikling for kultur i pct.	392	-1.1	4.8	-23.1	25.8	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for teknisk område i pct.	392	1.0	11.3	-35.6	53.1	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Årlig budgetudvikling for administration i pct.	392	1.7	6.9	-20.5	41.5	Budget i året fratrukket sidste års regnskab på område.	Danmarks Statistik
Indtægtsafvigelse i pct.	392	0.5	1.1	-3.2	4.3	Budget fratrukket regnskab for skatter og tilskud.	Nøgletal for kommunernes økonomiske styring (KREVI).
Årlig befolkningsudvikling i pct.	392	0.1	0.8	-3.1	5.1	Indbyggertal d. 1. januar i året fratrukket sidste års antal d. 1. januar.	De kommunale nøgletal
Årlig elevudvikling i pct. (antal 7-16 årige)	392	-1.1	1.6	-10.9	5.1	Antal 7-16 årige d. 1. januar i året fratrukket sidste års antal d. 1. januar.	De kommunale nøgletal
Årlig børneudvikling i pct. (antal 0-6 årige)	392	-1.5	2.5	-16.0	8.4	Antal 0-6 årige d. 1. januar i året fratrukket sidste års antal d. 1. januar.	De kommunale nøgletal
Årlig børne- og ungeudvikling i pct. (antal 0-17 årige)	392	-1.1	1.6	-9.3	7.3	Antal 0-17 årige d. 1. januar i året fratrukket sidste års antal d. 1. januar.	De kommunale nøgletal
Årlig ældreudvikling i pct. (antal 80+)	392	0.1	0.1	0.0	0.8	Antal 80+ årige i 1. kvartal i året fratrukket sidste års antal i 1. kvartal.	Danmarks Statistik
Budgetafvigelse for anlægsudgifter i pct.	392	-46.1	276.7	-3711.0	480.3	Budget fratrukket regnskab for skattefinansierede nettoanlægsudgifter.	Nøgletal for kommunernes økonomiske styring (KREVI).
Andel venstreorienterede mandater	392	45.9	12.5	10.5	78.9	Byrådsmedlemmer fra Socialdemokraterne (A), Socialistisk Folkeparti (F) og Enhedslisten (Ø) som procent af alle byrådsmedlemmer.	KMD's valgarkiv
Administrationsprofessionelle pr. 10.000 indb.	392	9.1	4.1	2.0	28.1	Antal medlemmer af Dansk Jurist- og Økonomforbund (DJØF) pr. 10.000 indbyggere d. 1. juli i året. For enkelte observationer oplyser DJØF mellem 0 og 6 DJØF'ere grundet persondatabeskyttelse. I disse tilfælde er observationen omskrevet til middelværdien 3.	Dansk Jurist- og Økonomforbund (DJØF)
Økonomiske rammevilkår	392	0.31	0.04	0.18	0.44	Udgiftsbehov divideret med summen af beskatningsgrundlag, udligning og tilskud samt nettorenter pr. indbygger.	De kommunale nøgletal
Indbyggertal	392	56832	64341	1839	559440	Indbyggertal pr. 1. januar i året.	De kommunale nøgletal
Sammenlagt (dummy)	392	0.67	0.47	0	1	Dummy for kommuner, der sammenlagdes i 2007. Ærø medtaget trods sammenlægning året før i 2006.	Egen kodning
Lille ø-kommune (dummy)	392	0.05	0.22	0	1	Dummy for kommuner beliggende på ø med mindre end 10.000 indbyggere og uden fast forbindelse. Langeland er medtaget.	Egen kodning

Figur 2: Gennemsnitlig budgetafvigelse for sektorer, 2007-2013
(uvægtede gennemsnit)



Note: Nettodriftsudgifter, dvs. bruttoudgifter fratrukket brugerbetaling og statsrefusion.

Tabel 4: Test af om effekten af sanktioner varierer på tværs af sektorer (N=98)

	Skole	Dagtilbud	Ældre	Sundhed	Udsatte børn og unge	Udsatte voksne	Kultur	Teknisk område	Administration
Sanktionsår (dummy 2011-13)	2.129*** (6.10)	1.652*** (4.69)	0.471 (0.96)	-0.389 (-0.57)	5.971*** (4.73)	0.982*** (5.59)	0.364 (0.80)	-0.262 (-0.21)	2.045*** (3.18)
Kontrolvariable	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Dagtilbud	Ikke signifikant								
Ældre	Signifikant***	Signifikant**							
Sundhed	Signifikant****	Signifikant***	Ikke signifikant						
Udsatte børn og unge	Signifikant***	Signifikant****	Signifikant****	Signifikant****					
Udsatte voksne	Signifikant***	Signifikant*	Ikke signifikant	Signifikant*	Signifikant****				
Kultur	Signifikant****	Signifikant**	Ikke signifikant	Ikke signifikant	Signifikant****	Ikke signifikant			
Teknisk område	Signifikant*	Ikke signifikant	Ikke signifikant	Ikke signifikant	Signifikant****	Ikke signifikant	Ikke signifikant		
Administration	Ikke signifikant	Ikke signifikant	Signifikant*	Signifikant***	Signifikant****	Ikke signifikant	Signifikant**	Signifikant*	

Note: * signifikant ved 0,10-niveau, ** signifikant ved 0,05-niveau, *** signifikant ved 0,01-niveau, **** signifikant ved 0,001-niveau. Test af om effekten af sanktioner varierer på tværs af sektorer. Estimeret via *seemingly unrelated estimation* på baggrund af tabel 2 model 1-7 og 9-10. Sanktionsårene 2011-2013 er slået sammen til en samlet dummy. Klyngekorregerede standardfejl med kommune som klynge. t-test i parentes.

Noter

1. Forfatteren takker alle interviewede informanter samt kollegaer, der har kommenteret artiklen, herunder post.doc. Eva Moll Sørensen, med hvem baggrundsinterviews med 6 kommuner er gennemført. En særlig tak skal lyde til deltagerne ved »Nordiske Kommuneforskningskonference« 27.-29. november 2014, forsknings- og praktiker-netværket »Forum Forvaltning« samt repræsentanter fra KL. Forfatteren er alene ansvarlig for indholdet.
2. Sanktionerne tilfredsstiller således Regeringens interesse i at undgå budgetoverskridelser, mens

KL generelt har fået en større opgave i at koordinere kommunernes økonomi for at gardere sig mod sanktioner.

3. Hvis ændringer i økonomistyringen imidlertid er et resultat af regnskabssanktioner, kan disse ændringer opfattes som mellemliggende variable, dvs. mekanismer, hvorigennem sanktioner virker, og man vil ikke skulle kontrollere for effekten af økonomistyring.
4. Skole, dagtilbud, ældre, sundhed, udsatte børn og unge, udsatte voksne og handicappede, teknisk område og administration.

Litteratur

- Angrist, J.D. og Pischke, J.S. (2009), *Mostly harmless econometrics. An empiricist's companion*, Princeton, New Jersey: Princeton University Press.
- Arbejderbevægelsens Erhvervsråd (2014), »Offentligt underforbrug koster tusindvis af job«, www.ae.dk/sites/www.ae.dk/files/dokumenter/analyse/ae_offentligt-underforbrug-koster-tusindvis-af-job_0.pdf.
- Blom-Hansen, J. (1998), *Studier i statens styring af den kommunale sektors økonomi*, Århus: Forlaget Politica.
- Blom-Hansen, J., Bækgaard, M. og Serritzlew, S. (2014), »Tax Limitations and Revenue Shifting Strategies in Local Government«, *Public Budgeting & Finance*, 34(1): 64-84.
- Christensen, J.G. og Lyndrup, M. (2013), »Interview med departementschef i Finansministeriet David Helleman: Budgetloven er et paradigmeskifte!«, *Administrativ Debat*, 1: 3-9.
- Christiansen, P.M. (2008), »De politiske budgetter«, i P. Munk Christiansen, red., *Budgetlægning og offentlige udgifter*, Århus: Academica, pp. 9-36.
- Dahler-Larsen, P. (2009), »Opskrift på virkningsevaluering«, i P. Dahler-Larsen og H.K. Krogstrup, red., *Nye veje i evaluering*, Viborg: Akademia, pp. 51-79.
- Dahler-Larsen, P. (2014), »Evaluation and NPM – as seen from the perspective of evaluation«, *LEGES*, (1): 27-38.
- Dalsgaard, C. og Kristensen, S.R. (2009), »Struktur i styringen? – Strukturelle forklaringer på kommunernes økonomistyring 1996-2005«, *Økonomistyring & Informatik*, 25(2): 137-61.
- Finansministeriet (2012), »Budgetlov«, www.fm.dk/nyheder/presmeddelelser/2012/03/regeringen-indgaar-aftale-om-ny-budgetlov/~media/Files/Nyheder/Presmeddelelser/2012/03/Budgetlov/budgetlov_handouts.ashx [hentet 2. januar 2014].
- Finansministeriet, Økonomi- og Indenrigsministeriet og KL (2013), »God økonomistyring i kommunerne. Sammenfatning«, www.fm.dk/nyheder/rss/rssfeed/~media/Files/Nyheder/Presmeddelelser/2013/06/KL%20aftale/Bilag%201.God%20%C3%B8konomistyring%20i%20kommunerne.pdf.
- Foged, S.K., Houlberg, K. og Petersen, O.H. (2014), »Kommunal forvaltning: Hvordan forklares kommunernes brug af private leverandører?«, i M. Boss, C.H. Grøn, S.T. Kristjansen og S. Rømer, red., *Den offentlige sektor – cases fra maskinrummet*, Hans Reitzels Forlag, pp. 101-124.
- Foged, S.K. og Sørensen, E.M. (2014), »Organiseringen af kommunernes økonomistyring i en krisetid«, 46. årsmøde i Dansk Selskab for Statskundskab – konference d. 23.-24. oktober, panel »Forvaltningspolitik og offentlig styring under krisen«, Hotel Vejlefjord, Vejle, Denmark.
- Hansen, H.F. (1989), »Moderniseringens effektivitet«, *Nordisk Administrativ Tidsskrift*, (2): 189-210.
- Hansen, H.F. (2011), »Evaluerings-teori«, i A. Berg-Sørensen, C.H. Grøn and H.F. Hansen, red., *Organiseringen af den offentlige sektor*, Hans Reitzels Forlag, pp. 397-420.
- Hariri, J.G. (2012), »Kausal inferens i statskundskaben«, *Politica*, 44(2): 184-201.
- Houlberg, K. (1999), »Budgetoverskridelsernes anatomi – en analyse af forskelle i danske kommuners overholdelse af budgetterne«, *Nordisk Administrativ Tidsskrift*, 80(3): 200-32.
- Houlberg, K. og Jensen, K.B. (2010), *Kommunernes økonomiske situation og udgiftspolitiske prioriteringer*, København: AKF, Anvendt Kommunal Forskning.
- Houlberg, K. og Jeppesen, T. (2006), »De drilske kommunale budgetter«, www.kora.dk/media/275432/Analyse_drilske_kommunale_budgetter_311006_0.pdf [hentet 4. november 2014].
- Juul, T. (2006), *Håndbog i kommunernes og regionernes nye økonomi*, København: Jurist- og Økonomiforbundets Forlag.
- Kjær, U. (2012), »Geografiske skillelinjer på flere niveauer«, *Økonomi & Politik*, 85(1): 3-13.
- KL (2012), »KL-høringssvar vedrørende forslag til budgetlov«, København.
- KL (2014), »Ingen overraskelser i de kommunale regnskabstal«, www.kl.dk/Okonomi-og-dokumentation/Ingen-overraskelser-i-de-kommunale-regnskabstal-id155838/?utm_source=kl.dk&utm_medium=email&utm_content=advisering&utm_campaign=email-advisering-15-05-2014&utm_id=f3770999-4049-4288-9aaf-17ea886fafd8 [hentet 30. december 2014].
- Krevi (2008), »Udfordringer for kommunal økonomistyring«, Århus.
- Krevi (2012), »God kommunal økonomistyring – en undersøgelse af seks kommuners praksis«, www.kora.dk/media/276536/God_Kommunal_Okonomistyring.pdf.
- Økonomi- og Indenrigsministeriet (2014), »Besvarelse af samrådsspørgsmål J om Produktivitetskom-

- missionens forslag om ændring af budgetloven», <http://www.ft.dk/samling/20131/almdel/kou/spm/87/svar/1133869/1365715.pdf>.
- Olsen, J. (2013), »Fremtiden for aftalesystemet«, *Administrativ Debat*, 1: 39-42.
- Panduro, B. og Lauriden, J.M. (2014), »Kommunernes regnskaber 2013. Overskud for første gang i 13 år«, KORA.
- Produktivitetskommissionen (2013), »Styring, ledelse og motivation i den offentlige sektor. Analyserapport 3«, København, Produktivitetskommissionen.
- Retsinformation (2011), »Forslag til Lov om nedsættelse af statstilskuddet til kommunerne i 2012 ved overskridelse af det aftalte udgiftsniveau for 2011«, www.retsinformation.dk/Forms/R0710.aspx?id=135459 [hentet 21. december 2014].
- Schwandt, T.A. (2005), »The Centrality of Practice to Evaluation«, *American Journal of Evaluation*, 26(1): 95-105.
- Serritzlew, S. (2005), »Breaking Budgets: An Empirical Examination of Danish Municipalities«, *Financial Accountability & Management*, 21(4): 413-35.
- Sørensen, E.M. og Foged, S.K. (2014), »Mål- og resultatstyring i danske kommuner efter krisen«, 46. årsmøde i Dansk Selskab for Statskundskab – konference d. 23.-24. oktober, panel »Forvaltningspolitik og offentlig styring under krisen«, Hotel Vejlefjord, Vejle, Denmark.

Giver resultatmåling klarhed for medarbejderne?¹

**Peter Dahler-Larsen, professor, Institut for Statskundskab,
Københavns Universitet, pdl@ifs.ku.dk**

Et af de fremmeste formål med resultatmåling er at give medarbejderne et klart billede af, hvad der skal gøres i organisationen. Imidlertid findes signifikant empirisk variation i, hvorvidt to sammenlignelige grupper af offentligt ansatte (gymnasielærere og sproglærere) oplever, at resultatmålinger giver klarhed. Kvalitative data om resultatmålingssystemerne og de organisatoriske kontekster, de fungerer i, hjælper os til at forstå kilderne til klarhed/uklarhed. Afslutningsvis diskuteres fordele og ulemper ved klar resultatmåling i et offentligt styringsperspektiv og et demokratisk perspektiv.

Performanceindikatorer og resultatmåling er ofte begrundet i løfter om øget gennemsigelighed og transparens i den offentlige sektor, om bedre grundlag for ledelsesbeslutninger og ultimativt højere effektivitet (Talbot, 2005). Hvis målene er klare, operationelle og realistiske, er der plads til at lempe klassisk regelstyring og i stedet give medarbejderne frihed til at vælge metoder og tilgange i det daglige arbejde. Klare krav til arbejdets retning og omfang er også fornuftige ud fra såvel motivationsteori som stressteori. Et fælles omdrejningspunkt for alle disse ræsonnementer er, at medarbejderne oplever, at resultatmålinger medvirker til at give klarhed over, hvad man skal nå.

Imidlertid må de gode virkninger af resultatmåling betragtes som hypoteser snarere end

som kendsgerninger (Kristiansen, 2014: 24). Det gælder også graden af klarhed.

Resultatmålinger kan f.eks. tilsammen være uklare, hvis der er flere af dem, hvis de er ustabile og indbyrdes modstridende. Der kan også være uklarhed som følge af usikkerhed om, hvordan resultatmålinger anvendes i beslutninger, idet selv sofistikeret resultatmåling ikke udgør nogen garanti for kompetent anvendelse (Vakkuri og Meklin, 2006: 235). I det hele taget er brugen af information under tvetydighed et klassisk tema i organisations-teori (Feldman og March, 1981; March og Olsen, 1976). Endelig kan man forestille sig, at mange kriterier indgår i medarbejdernes fortolkning og vurdering af resultatmålinger, såsom validitet, retfærdighed og nytte. Filosofisk set lurder tvetydighed altid i grænsefladen mellem det målbare og det ikke-målbare (Best, 2008).

Der er derfor grund til at undersøge empirisk, hvorvidt medarbejdere oplever, at resultatmåling giver klarhed. Formålet med denne artikel er at beskrive de empiriske forskelle heri for to udvalgte medarbejdergruppers vedkommende, at finde kilderne til disse forskelle og afslutningsvis sætte resultaterne i et større velfærdspolitisk og demokratisk perspektiv.

Undersøgelse af gymnasielærere og sprog-lærere

De resultater, som gengives nedenstående, stammer fra en undersøgelse af sammenhængen mellem resultatmålinger og stress finansieret af Arbejdsmiljøforskningsfonden (Dahler-Larsen og Pihl-Thingvad, 2014). I det følgende fokuseres alene på klarheden i resultatmålinger som oplevet af to medarbejdergrupper nemlig sproglærere (for indvandrere i dansk) og gymnasielærere.

Fra de to grupper foreligger surveydata med svarprocenter på hhv. 61,4 fra sproglærerne og 43,1 fra gymnasielærerne. I surveyen blev resultatmålinger forklaret således: »Ved resultatmålinger forstår vi generelt tal, der handler om, hvor meget man når på arbejdspladsen. Det er tal, som sendes opad til ledelsen. Vi kalder det resultatmåling, uanset om det er dig eller andre, der indberetter, dokumenterer og tæller sammen.« Endvidere blev nævnt konkrete eksempler på resultatmålinger for hver enkelt gruppe.

Hos sproglærerne benyttes et nationalt testsystem i samtlige undervisningsmoduler, og sprogskolernes økonomi afhænger mestendels af antal beståede modultest. Gymnasiernes resultatmåling omfatter karakterer, frafaldsprocenter og elevtilfredshedsdata samt en række andre dokumentationspraksisser. Gymnasiernes økonomi er taxameterbaseret og afhænger dermed af søgning og gennemførelse.

Sammenligning af grupperne er meningsfuld. Begge varetager undervisning og deler dermed nogle af de fundamentale vanskeligheder ved resultatmåling inden for denne sektor. Sproglærerne er typisk magistre og folkeskolelærere med efteruddannelse i sprogundervisning, mens gymnasielærerne helt overvejende består af magistre. Der er ikke ved første øjekast særlig grund til at antage, at en af grupperne på forhånd skulle have meget stærkere positive eller negative holdninger til resultatmåling end den anden.

Men deres resultatmålingssystemer er ganske forskellige. Det betyder ikke, at alle forskelle i klarhed i resultatmålinger kan henføres til selve resultatmålingssystemet. Klarhed er et spørgsmål om nogens tolkning af noget, det vil sige en relation. Gruppernes arbejds-kontekst, specifikke faglighed og fortolknings-systemer i det hele taget må med ind for at forstå, hvorvidt der findes »klarhed« i resultatmålinger. For at komme nærmere på disse kontekstuelle forhold beskrives de to gruppers situation (i efterfølgende afsnit) ved kvalitative casestudier baseret på interviews, virksomhedsbesøg og dokumentstudier.

Skaber resultatmålinger klarhed?

Men lad os først se på medarbejdernes oplevelse af, hvorvidt resultatmålinger skaber klarhed som de fremgår af spørgeskema-data. Tabel 1 viser de to gruppers vurderinger.

Tabel 1: Holdninger hos sproglærere og gymnasielærere

	Sproglærere	Gymnasielærere
Resultatmålinger skaber klarhed over, hvad vi skal nå	3,33	2,88
Hvor ofte tænker du på resultatmålinger	3,29	2,84
Resultatmålinger er belastende	2,72	2,58
Resultatmålinger fører til ensidigt fokus	3,76	3,48
Jeg er villig til at arbejde lidt mere, end jeg bliver betalt for, hvis det giver høj faglig kvalitet	3,55	3,73

Note: gennemsnitlig vurdering på klassisk fem-punkts Likert-skala, »helt uenig«, »aldrig« og »slet ikke« svarer til 1, mens »helt enig«, »altid« og »i meget høj grad« svarer til 5.²

Resultatmålinger opleves af sprog-lærere at give klarhed over, hvad man skal nå, i højere grad end tilfældet er hos gymnasie-lærerne. Der er ikke automatik i forholdet mellem resultatmåling og klarhed.

Tabel 1 rummer også andre aspekter af resultatmåling, som medarbejderne oplever dem. Sproglærere tænker mere på resultatmåling, de oplever dem som mere belastende, og de oplever i højere grad end gymnasie-lærerne, at resultatmålinger fører til ensidigt fokus i arbejdet. Sidstnævnte nævnes ofte som et hovedproblem ved resultatmåling, især når enkle resultatindikatorer sættes ind i en of-fentlig opgavekontekst præget af komplek-sitet, flertydighed og dynamik. Medmindre opgavekonteksten hos sprog-lærerne er mere kompliceret end i gymnasiet, motiveres man kun yderligere til at lede efter forklaringer på sprog-lærernes høje score i selve opbygningen af resultatmålingssystemet.

Sidste emne i tabel 1 vedrører det profes-sionelle engagement. Gymnasie-lærerne har et (endnu) højere professionelt engagement end sprog-lærere, i hvert fald målt på viljen til at arbejde ekstra for at opnå høj kvalitet i arbej-det.

I tabellen er termen »klarhed« baseret på respondenternes egen oplevelse af, hvorvidt resultatmålingerne giver klarhed. Vi vil nu grave i de kvalitative data for at finde ud af, hvordan en sådan klarhed eventuelt kan ska-bes, og hvad kilderne til uklarhed på den an-den side kan være.

Klarhed i resultatmålinger blandt sprog-lærere

På området »undervisning i dansk for indvan-drere« domineres indikatorlandskabet en en-kelt type af talstørrelse, nemlig antal beståede modultest. Hele danskundervisningen er op-delt i moduler. Bestemte resultater i bestemte modultest er afgørende for statsborgerskab

og opholdstillade, og antallet af beståede mo-dultest er afgørende for skolens økonomi.

Da der er løbende optag på holdene, gøres hver lærer fra administrationens side bekendt med det normalt forventede testtidspunkt for den enkelte elev i det enkelte modul. Efter lærerens og kursistens bedømmelse kan den faktiske test foretages før eller efter det nor-merede tidspunkt.

Den administrative ledelse siger, at den let kan måle den enkelte lærers bidrag til skolens økonomi i kraft af elevernes beståede mo-dultest. Men det ville være »ubehageligt« og »urimeligt«, siger flere lærere, fordi testene ikke er lige lette på alle moduler og for alle elever. Eleverne er meget uens med hensyn kulturel og boglig baggrund. Men tallene »bliver nævnt og bliver ikke nævnt«, som en lærer siger, også selv om man har stoppet en tidligere praksis med at cirkulere månedlige testtal for hele skolen.

Lærerne har meget blandede synspunkter om testene. Kritiske røster mener, testen er for mekanisk, at den tester »papegøjessprog«, men ikke i tilstrækkelig grad omfatter ordfor-råd og udtale. Den er for »skoleorienteret«. Der er simpelthen ikke tid til at undervise i handlingspræget og praktisk dansk. På den anden side giver testen en håndfast melding til kursisterne om undervisningens mål. Kur-sisterne får »en klar fornemmelse« af, hvad de skal nå. Selv om testen har svagheder, så er den »ikke fuldstændig åndssvag«, som en lærer siger. Ligeledes sikrer testen oprykning af kursister til næste modul på et nogenlunde objektivt grundlag. I »gamle dage« kunne kursisterne studere dansk »for evigt«. Det var ifølge en lærer »cowboyland«, hvor der var løst definerede mål og ingen opgørelse af, hvad kursisterne lærte undervejs.

Men testen har medvirket til et vist enheds-præg og en vis professionalisering. Målsæt-ninger, emner og materialer er blevet mere

strukturerede og »vi er blevet dygtigere«, siger en lærer. I ordet »professionalisering« lægger flere lærere også evnen til at skabe resultater i undervisningen uden at blive overmåde følelsesmæssigt involveret i kursisterne, deres problemer og deres skæbne. »Hvis du er en god lærer, hvis du er egnet, så kan du få det til at virke«, som en siger. »Hvis det går godt i testen, så er læreren god, og hvis det sker for mange gange, at kursisterne dumper, så har det nok et eller andet at gøre med læreren.« »Jeg tror, jeg er en god lærer. Jeg gør det, så godt jeg kan, så jeg er ikke søvnløs om natten.« Den samme siger: »Man udvikler en vis professionalisme med tiden. Når man er professionel, siger man ikke fra starten, at ledelsen er dumme. Man skal ligesom have flere øjne og ører samtidig.« Med andre ord ser flere lærere virkeligheden med test som en virkelighed, den gode lærer må håndtere.

Testen er blevet til en kendsgerning. »Det er bare sådan, det er«, siger en tester. Testen kobler tid og penge og knytter sig derved til to store institutionelt anerkendte målestokke i vores samfund. Flere specifikke mekanismer hjælper formodentligt med at skabe et institutionelt »lock-in« (Osterloh og Frey, 2014), altså en situation, hvor forestillinger, belønninger og sanktioner hjælper med til at holde en bestemt virkelighed på plads. De mest utilfredse lærere siges at have forladt skolen efter frugtesløse debatter om testsystemet.

En lærer vil ikke længere være tester. »Det er simpelthen for hårdt at erklære nogen mennesker for anden-rangsmennesker, når de dumper«, siger vedkommende. De lærere, der bliver på posten, ser det som en del af deres professionalisme, at de kan håndtere de udfordringer, som testpresset giver. Samtidig udvikler de fortællinger om, hvilke personlighedstræk der har gjort, at netop de er gode til det.

Selv om der normativt er både plusser og minusser ved modultests systemet, så er dets in-

stitutionelle rolle meget klar. Efter en vis tid er således indtrådt en vis accept af testen om ikke normativt, så i hvert fald faktisk.

Uklarhed i resultatmålinger blandt gymnasielærere

Hos gymnasielærerne finder vi ikke i de kvalitative data en tilsvarende klarhed. En første vigtig kilde til uklarhed er derimod antallet af resultatindikatorer og forholdet mellem dem. Når både karakterer, gennemførelsesprocenter og elevtilfredshedstal gøres gældende, oplever lærerne hurtigt et spændingsforhold mellem disse. Hvor skrap skal man for eksempel være, hvad angår sanktioner for ikke at aflevere opgaver til tiden? Skal man presse de grupper af elever, der måske ikke burde være på gymnasiet, så de alligevel kan gennemføre? Og hvilke konsekvenser vil et sådant pres have på tilfredshedsmålingssiden? »Hvis vi prøver at få frafaldet ned er det måske på bekostning af en højere score på elevtilfredsheden«, siger en lærer.

Ydermere er der betragtelig tvivl om flere af indikatorernes validitet og mening. Er tilfredshedsmålinger egentlig relevante på en institution, der tilbyder almindelse og studieforberedelse?

Forældres socioøkonomiske og etniske baggrund udgør en væsentlig forklaringsfaktor bag elevernes resultater på enhver uddannelsesinstitution. Hvornår elevsammensætningen skal tælles med, og hvornår den ikke skal, er ikke entydigt. På den anden side, hvis den tælles med, »så er det ikke vores skyld«, som en lærer siger, men så kan elevsammensætningen også bruges som en »undskyldning« for dårlige resultater.

Hvis et gymnasium beslutter sig for at konkurrere på et helt åbent marked, kan det sætte andre faglige mål, end hvis det først og fremmest ønsker at være en institution for områdets unge, uanset deres sociale baggrund. Det bliver derfor svært at vurdere validiteten af

karakterniveauet som udtryk for gymnasiets kvalitet uden at tage gymnasiets mission i betragtning.

De forskellige faglige paradigmer mellem natur-, samfunds-, og humanvidenskaber skaber ligeledes forskellige forudsætninger for at acceptere, at resultater af undervisning kan beskrives ved tal. Inden for områderne natur og teknik, for ikke at nævne erhvervsøkonomi, er der større modtagelighed for grundideen. Karakterer er særlig vanskelige at tolke uden at tage faglighed i betragtning: »Der er fem gange større sandsynlighed for at dumpe i skriftlig fysik A end i skriftlig samfundsfag A«, som en lærer siger.

Med til fortolkningsmæssig tvetydighed er også det forhold at mange resultatmålinger opgøres på skole- og klasseniveau. Det gør det vanskeligt at henføre dem til den enkelte lærer. Det er derfor uklart, i hvilket omfang en lærer skal føle sig ramt af en dårlig elevtilfredshedsmåling på klasseniveau.

Endelig er det meget uklart ifølge vore kvalitative data, hvem der bruger resultatmålingerne til hvad. Der er spredte formodninger om, hvad et eventuelt kvalitetsudvalg vil bruge hvilke data til. Ledelsens rolle er heller ikke indlysende, selv om flere lærere fortæller, at det på temadage er taget op, hvordan man giver karakter, »og det er nok ikke tilfældigt«. Men meget ligger i nogle udvalg og nogle grupper, og meget ser ud til at være et resultat af diskussioner snarere end tydelige beslutninger. »Vi mangler en klar melding«, siger en. »Det kunne være godt med en strategi.«

Uklarheden om, hvad tallene betyder, medvirker til en uklarhed om, hvorvidt tallene i det hele taget kan bruges. Men, som en siger, »når så mange har været med (i en tilfredshedsundersøgelse)«, så skal der sørme også komme noget ud af det.«

For at få optimalt blik for det uklare i resultat-

målinger og for diskussionens rolle i tolkning af resultatmålinger, specielt elevtilfredsheds-målinger vil vi nu besøge et gymnasium, hvor man skal diskutere en elevtilfredsheds-måling på en temadag. Gymnasiet har som 55 andre deltaget i en ElevTrivselsUndersøgelse (ETU) udført af et konsulentfirma.

Der er store forbehold i lærerkredse mod at sammenligne gymnasier, fordi elevgrundlaget er forskelligt. Men, oplyser vicerektor, hvis det kommer til arbejdsmiljøundersøgelser, så har lærerne lige pludselig ingen indvendinger mod at sammenligne gymnasier, også selv om lærerne er forskellige fra gymnasium til gymnasium.

På temadagen lægger en lærer og vicerektor resultaterne frem for de ca. 60 lærere. Gymnasiet ligger under gennemsnittet på flere punkter. Det bliver diskuteret, om forskellige elever fra forskellige gymnasier er lige dygtige, motiverede og interesserede, og om man på den baggrund kan sammenligne elevtilfredshed lige over mellem forskellige gymnasier. Der stilles også spørgsmålstejn ved den anvendte skala og rimeligheden af rangliste over skoler.

En lærer stiller spørgsmålstejn ved undersøgelsens kriterier. Hvad er egentlig vores værdier her på gymnasiet? Skulle vi ikke have drøftet vores værdigrundlag først, inden der kommer en sådan undersøgelse? Vicerektor svarer hertil, at »jamen, det er så det vi har temadag om nu«!

Drøftelsen strejfer også et dilemma i forbindelse med elevtrivsel. Hvis man stiller højere krav til faglighed og tilstedeværelse, kan det gå ud over elevtrivslen. Hvis man ønsker høj elevtrivsel, kan det opnås ved en slappere praksis i forhold til fravær, manglende aflevering af opgaver osv.

Fire diskussioner ser ud til at være på spil samtidig: En sandhedsdiskussion, en værdi-

diskussion, en retfærdighedsdiskussion og en brugbarhedsdiskussion. Sandhedsdiskussionen handler om, hvorvidt ETU giver et korrekt og troværdigt billede, og om metoderne hertil er gangbare. Værdidiskussionen handler om, hvorvidt gymnasiets værdier er blevet respekteret i ETU-målingen. Retfærdighedsdiskussionen går både på, om sammenligningen mellem gymnasier er rimelig, og om undersøgelsen er tilrettelagt på den rigtige måde i forhold til inddragelse af medarbejderne. Sidst, men ikke mindst, går brugbarhedsdiskussionen på, hvad resultaterne skal bruges til.

De fire diskussioner er forbundne. For eksempel hænger sandhedsdiskussionen og værdidiskussionen sammen, fordi undersøgelsen ikke giver et »sandt« billede, hvis den måler på de »forkerte« værdier. Og brugbarhedsdiskussionen afhænger meget af de tre andre diskussioner. Spørgsmålene fra flere af de tilstedeværende lærere kan ses som forsøg på at rejse sandheds-, værdi- og retfærdighedsdiskussioner, for derved at blokere for en brugbarhedsdiskussion under antagelsen: Hvis undersøgelsen ikke er god (sand, retfærdig osv.), kan den ikke bruges. Ledelsen, derimod, forsøger at pakke de tre første diskussioner godt ned, således at man kan bruge undersøgelsen som springbræt til at skabe udvikling i de pædagogiske virkemidler blandt lærerne.

Uklarheden er produktiv. I hvert fald kommer der i løbet af debatten forslag frem om, hvordan fremtiden kan eller bør formes. Der blev for eksempel formuleret et ønske om, at man fremover også leverer elevtilfredshedsdata på individuelt lærerniveau, således at den enkelte lærer kan handle mere direkte som følge af data. Nogen ønsker, at ledelsen får klarere redskaber, således at man mere direkte griber ind over for lærere med dårlige resultater. Der ønskes også »en klar melding« og en tydelig »strategi« for brugen af resultatomålinger.

Selv om lærerne er skeptiske over for validiteten i for eksempel karakterer eller tilfredshedsmålinger, så ser man muligheden for at bruge et godt tal i markedsføring af gymnasiets kvaliteter. »Det skal bruges i salgsøjemed.« »Vi holder op med at være pæne om 4 år cirka« (for der kan man forudsige en mangel på elever). Selv om man er irriteret over at blive hængt op på tal, hvor der for eksempel ikke er taget højde for ændringer i elevsammensætningen, så er fristelsen på den anden side stor til selv at bruge tal, der er lige så lidt valide, hvis de til gengæld udstråler noget mere positivt. På den måde sender resultatomålinger gymnasielærerne ud i et tveetydigt moralsk terræn. Hvis man ikke kan modvirke den ukomfortable brug af et ikke-validt dårligt tal, må man så gerne fremme den komfortable brug af et godt tal, der er lige så lidt validt?

Det imaginære arbejder på højtryk, måske netop på grund af den megen uklarhed omkring resultatomålingerne. De mange forslag om, hvordan uklarheden kan reduceres i fremtiden, skaber mere uklarhed, for eksempel når der er konflikt mellem hensynene til validitet og til brugbarhed.

Klarhed har en pris

Empirisk har vi fundet signifikant forskellige opfattelser blandt to grupper af lærere af, i hvilket omfang resultatomåling giver klarhed.

På sprogskolen er disse normative diskussioner blevet reduceret, efter at de mest kritiske lærere har forladt skolen, og testsystemet som fakticitet er etableret. De lærere, der er kritiske over for testens validitet og sociale retfærdighed giver udtryk herfor, men har institutionelt ikke mange andre steder at gå hen end at forlade skolen eller at ophøre med de opgaver, man finder mest ubehagelige. Diskussionerne er ikke længere åbenlyse, eksplícite og manifeste.

Medvirkende til at fastholde testen som en

helt central del af sprogskolens resultatindikatormiljø er, at den indgår som en afgørende faktor i skolens økonomi, og at der kun er én resultatindikator af økonomisk betydning. Måske som konsekvens af den relativt etablerede institutionelle klarhed har resultatmålingssystemet hos sproglærerne succes med at påvirke undervisningen i retning af et entydigt fokus på testresultater. Og måske derfor oplever sproglærerne resultatmåling som belastende.

Modsat er det på gymnasierne, hvor både ETU, karakterer, frafaldsundersøgelser og variationer af lokalt evalueringsarbejde indgår i resultatmålingslandskabet. Kriterierne er præget af fortolkningsmæssig tvetydighed og opleves i et vist omfang som indbyrdes modstridende. Der er stor uklarhed om, hvem der eventuelt bruger resultatmålingerne til hvad. Den meget involverende arbejdsstil på temadagen er formodentlig et tegn på, at ledelsen alene ikke magter at formulere et sæt resultatmålinger med en entydig tolkning. Samtidig kan de mange forsøg og gruppearbejder og løbende drøftelser også være med til at vedligeholde en uklarhed.

I gymnasiets kontekst indgår en mangfoldighed af politiske målsætninger, som genspejles i resultatmålingernes sammensathed. Omvendt på sprogskolerne, hvor hverken kursister eller deres forældre eller erhvervslivet organiserer sig og artikulere varierende politiske krav. Den politiske linje er entydig: Indvandrere skal lære dansk, og det kan kun gå for langsomt med at bestå disse test.

Hos sproglærerne ser vi en langt større klarhed, men også at resultatmålingerne skaber større skævvridning af det pædagogiske arbejde i retning af »teaching-to-the-test«, og resultatmålinger bliver, måske i kraft af netop deres klarhed, til en større belastning for medarbejderne. Vi ser også hos sproglærerne en mindre villighed til at arbejde ekstra uden betaling, som var vores indikator for professi-

onelt engagement. Det er forventeligt, at man trækker engagementet, når man udsættes for klar resultatmåling, man finder ensidig og belastende. Omvendt kan en vis mængde uklarhed og inkonsistens i resultatmålinger hos gymnasielærerne udgøre et bolværk imod at tænke alt for meget på resultatmålinger og imod alt for stærke indvirkninger på praksis. Her er det professionelle engagement større.

Perspektiverer vi disse to sæt resultater ved at se dem som mulige udfaldsrum på et styringsmæssigt kontinuum, bliver dilemmaet tydeligt. Klarhed i resultatmålinger medvirker til at strukturere et felt, til at stoppe utallige diskussioner og til at give tydelige meldinger til medarbejdere og kursister om undervisningens mål. Men medarbejderne tænker mere på resultatmålinger og finder dem mere belastende og mere styrende i undervisningen. På den anden side: Ved de uklare resultatmålinger finder man færre negative virkninger af resultatmåling og et større professionelt engagement. Hvor resultatmålingerne er mere klare, er de negative virkninger af resultatmålinger stærkere og det professionelle engagement mindre. Måske kan prisen for klarhed kan blive for høj. I det mindste antyder undersøgelsens resultater et vanskeligt styringsdilemma, såvel som et velfærdspolitisk og demokratisk problemområde snarere end blot et enkeltstående ideal om, at klarhed altid er et gode i sig selv.

Noter

1. Artiklen er en stærkt omarbejdet udgave af kapitel 12 i Dahler-Larsen og Thingvad (2014).
2. De fem forskelle mellem grupperne er fundet signifikante ved t-test ($p < 0.05$).

Litteratur

- Best, J. (2008), »Ambiguity, Uncertainty, and Risk: Rethinking Indeterminacy«, *International Political Sociology*, 2: 355-74.
- Dahler-Larsen, P. og S. Pihl-Thingvad (2014), *Resultatmåling og Stress*, Syddansk Universitetsforlag.
- Feldman, Martha S. og James G. March (1981), »Information in Organizations as signal and sym-

- bol«, *Administrative Science Quarterly*, 26(2): 171-86.
- Kristiansen, M.B. (2014), »Introduktion til resultatstyring«, i M.B. Kristiansen, *Resultatstyring i den offentlige sektor*, København: Jurist- og økonomiforbundets forlag, pp. 13-34.
- March, J. og J.P. Olsen (1976), *Ambiguity and Choice in Organizations*, Universitetsforlaget.
- Osterloh, M. og B. Frey (2014), »Academic Rankings between 'the Republic of Science' and 'new Public Management'«, i A. Lanteri og J. Vromen, red., *The Economics of Economists*. Cambridge: Cambridge University Press.
- Talbot, C. (2005), »Performance Management«, i E. Ferlie, L.E. Lynn og C. Pollitt, red., *The Oxford Handbook of Public Management*, Oxford university Press, pp. 491-517.
- Vakkuri, J. og P. Meklin (2006), »Ambiguity in Performance Measurement: A theoretical approach to organisational uses of performance measurement«, *Financial Accountability and Management*, 22(3): 235-50.

Review: Produktivitetskommissionen – en analyse. Del 1: Generel del

Christen Sørensen, fhv. overvismand
cs020345@gmail.com

Produktivitetskommissionen skulle analysere årsager til en utilfredsstillende produktivitetsudvikling i Danmark siden midten af 1990'erne og følge op med konkrete forslag til produktivitsfremme. Kommissionen fik og accepterede vel også et meget krævende kommissorium såvel som en kort tidsfrist på ca. 2 år til at udføre de stillede opgaver.

Desværre oplyses det ikke, hvilke generelle overvejelser kommissionen gjorde sig i lyset heraf. Det heraf følgende ekstra tidspres har præget kommissionens arbejde. Planlagte undersøgelser blev helt eller delvist opgivet, relevante problemstillinger blev ikke analyseret, og på centrale områder – bl.a. uddannelsesområdet – blev analyserne ikke afsluttet med konkrete forslag til, hvordan identificerede svagheder kunne afhjælpes.

Næppe nogen kommission kunne have løst de i kommissoriet stillede opgaver – og da slet ikke inden for en tidsfrist på ca. 2 år. Men kommissionen burde straks have gjort eksplicit opmærksom på, at de stillede opgaver ikke kunne løses tilfredsstillende inden for en tidsfrist på ca. 2 år.

Kommissoriet: krævende og alligevel med meget kort tidsfrist

Den efterfølgende analyse af Produktivitetskommissionens arbejde er primært baseret på kommissionens 6 analyserapporter og Slutrapporten.

Kommissoriet for Produktivitetskommissionen blev desværre formuleret på en sådan

måde, at en tilfredsstillende opgaveløsning nærmest på forhånd var umuliggjort.

For det første skulle der gennemføres en meget bred og omfattende analyse af årsagerne til den utilfredsstillende produktivitetsudvikling i Danmark siden midten af 1990'erne. Årsagerne hertil skulle identificeres for såvel den private som den offentlige sektor samt helt ned i en række underliggende erhverv indenfor industri, byggeri og service. Og i tilknytning hertil skulle det kortlægges, hvordan viden og uddannelse i sammenhæng hermed udnyttedes af virksomhederne.

I løsningen af denne opgave skulle ikke blot eksisterende viden, men også viden fra nye analyser, inddrages. På basis heraf skulle kommissionen opstille *konkrete forslag* til, hvordan produktiviteten i såvel private erhverv som i det offentlige kunne øges, så produktiviteten i Danmark igen kunne nærme sig topniveauet blandt OECD-landene. Dette var en kolossal opgave ikke mindst grundet de meget generelle og vage formuleringer i kommissoriet. Næppe nogen kommission ville have kunnet løfte denne opgave tilfredsstillende.

For det andet fik Produktivitetskommissionen en meget kort tidsfrist til at gennemføre

det ønskede omfattende og krævende analysearbejde, som kommissoriet fra marts 2012 lagde op til. Oprindeligt skulle arbejdet være afsluttet ved udgangen af 2013, men fristen blev udsat til ultimo marts 2014.

Kombinationen af den meget omfattende og bredt formulerede opgave for Produktivitetskommissionen og den ret korte tidsfrist, der også skal ses på baggrund af, at kommissionen selv skulle etablere et sekretariat, har haft afgørende betydning for opgaveløsningen. Dette fremgår måske mest tydeligt, når de planlagte undersøgelser sammenholdes med de realiserede undersøgelser.

Analysen af Produktivitetskommissionens arbejde er opdelt i hhv. en generel del, hvor kommissionens arbejde belyses under et, og en speciel del, hvor der dels fokuseres på det talmæssige udgangspunkt for kommissionens arbejde, dvs. de anvendte produktivitetsmål, og dels på uddannelsesområdet. Mens det ikke er nødvendigt at udbygge, hvorfor de anvendte produktivitetsmål er centrale i en analyse af produktivetsproblemet, kan det være relevant at begrunde, at uddannelsesområdet sættes i fokus udover dets betydning for arbejdskraftens effektivitet el. produktivitet. Kommissionen undervurderede tilsyneladende i det meste af dens arbejdsperiode uddannelsesområdets betydning for den utilfredsstillende produktivitetsudvikling. Dette har givetvis også medført, at kommissionen end ikke har forsøgt at analysere konkrete forslag til nye reformer på uddannelsesområdet. På dette helt centrale område har kommissionen begrænset sig til at diagnosticere problemer i vort uddannelsessystem, selvom der i henhold til kommissoriet skulle fremlægges konkrete forslag.

Den generelle del er opdelt i fire hovedafsnit og en konklusion: 1) Planlagte undersøgelser, der helt eller delvist blev opgivet, 2) Problemstillinger eller områder, der ikke er analyseret, 3) Kommissionens forslag til yderli-

gere undersøgelser, 4) Problemstillinger og områder, der i realiteten blev negligeret samt 5) Generel vurdering af opgaveløsningen.

Planlagte undersøgelser, der helt eller delvist blev opgivet

I skema 1 er der en oversigt over de planlagte syv undersøgelser, der ikke eller kun i meget sporadisk form er blevet gennemført. Som det fremgår, hidrører disse syv eksempler fra de to første analyserapporter, hvilket kunne indikere, at kommissionen, som deadline nærmede sig, indså, at der var en række undersøgelser, som det ikke var muligt at gennemføre indenfor den givne tidsfrist. Det er da også konsistent hermed, at det først er i senere rapporter – fra og med Analyserapport 4 – at der eksplicit gøres opmærksom på, at en række ønskelige undersøgelser ikke gennemføres, jf. skema 2. I slutrapporten er det dog i nogle tilfælde forsøgt at rette op herpå, jf. det følgende.

Sammenligninger af kommunernes produktivitet

I Slutrapporten er der side 104-05 et kort referat af resultaterne af den i Analyserapport 1 annoncerede analyse af de effektivitetsforbedringer, som hypotetisk kunne opnås, hvis effektiviteten i den kommunale opgaveløsning i alle kommuner blev løftet til effektivitetsniveauet i den mest effektive kommune. Denne analyse er på Produktivitetskommissionens foranledning blev udført af KORA (Det Nationale Institut for Kommuners og Regioners Analyse og Forskning). Der er kun kort anført resultater for folkeskoler, daginstitutioner og ældrepleje, hvor potentialet er opgjort til omkring 10 pct. i effektivitetsforbedringer. Den tilsigtede ambition med analysen er derfor næppe opnået.

Branchevise problemstillinger vedrørende beregning og fortolkning af produktivitet

Betragtningerne herom for så vidt angår de private serviceerhverv er i Analyserapport 1 samlet i kapitel 8. Betragtningerne er foku-

Skema 1: Planlagte undersøgelser, der helt eller delvist blev opgivet

Rapport ^a	Side	Emnet for undersøgelsen
1	14 58	Sammenligninger af kommunernes produktivitet: »Sådanne sammenligninger vil være en del af Kommissionens fremadrettede arbejde.« »Kommissionen har bl.a. igangsat et projekt, der på baggrund af detaljerede regnskabstal skal skabe ny viden om kommunernes effektivitet. Projektet vil indgå i Produktivitetskommissionens vurdering af potentialet for at øge produktivitet, effektivitet og kvalitet i den offentlige sektor.«
1	29	»I del 2 redegøres for branchespecifikke problemstillinger vedrørende beregning og fortolkning af produktivitet.«
1	66	»Produktivitetskommissionen vil derfor være særligt opmærksom på, om et bedre samspil mellem forsyningsvirksomheder og andre virksomheder kan bidrage til øget produktivitetsvækst i Danmark som helhed.»
1	92	»Produktivitetskommissionen vil i sit kommende arbejde se nærmere på, om Danmark er god nok til at kanalisere investeringerne derhen, hvor de gavner produktiviteten mest.«
2	97	Selskabsskatten: »... emnet vil blive behandlet i en senere rapport, hvor også skattens betydning for udenlandske selskabers tilskyndelse til at investere i Danmark indgår.»
2	102	Forskerskatteordningen: »... et tema, som Kommissionen vender tilbage til ... i forbindelse med en bredere analyse af muligheden for at anvende skattepolitikken til at styrke produktivitetsudviklingen.»
2	133	El-, gas- og vandforsyning: »... i en senere rapport om infrastrukturens betydning for produktiviteten vil Produktivitetskommissionen vende tilbage til forholdene inden for el-, gas- og vandforsyning.«

a. Analyserapporter er i kolonne 1 i skema 1-3 anført ved løbenummer.

seret på de problemer, der måtte følge af, at der ikke er anvendt A-metoder, men B- eller C-metoder i deflateringen, dvs. i omregningen fra løbende til faste priser. Opdelingen af deflationsmetoder i A-, B- eller C-metoder hidrører fra EU's statistikenhed, Eurostat. Ved A-metoder tages kvalitetsændringer i høj grad i betragtning i opgørelsen af kvalitetsjusterede priser, dvs. priser korrigeret for kvalitetsændringer. Såfremt det ikke er muligt at opstille kvalitetsjusterede priser, siges deflateringen at være gennemført efter en B-metode. Det vil ved kvalitetsforbedringer føre til en overvurdering af prisudviklingen og derfor til en undervurdering af mængdeudviklingen. Såfremt der på et område ikke findes prisserier, som f.eks. på en række offentlige områder, siges deflateringen at følge en C-metode.

Om dagligvarebranchen anføres (Analyserapport 2: 117) eksempelvis i fortsættelse

heraf: »B-metoder kan kun anvendes med forbehold ved produktivitsanalyser, da de ikke tager højde for, at en del af prisstigningen i en branche kan skyldes en højere kvalitet af produkterne. Det kan derfor ikke udelukkes, at de officielle statistikker giver et for negativt billede af produktivitsudviklingen i dagligvaresegmentet«.

I betragtning af, at kommissionen lægger til grund, at arbejdstimeproduktiviteten skulle være faldet med i gennemsnit 1,4 pct. om året i 12-årsperioden 1995-2007 for dagligvarebranchen, er det anførte såvel i Analyserapport 1 som 2 helt utilstrækkeligt.

Produktivitet går normalt kun én vej, nemlig opad. I begyndelsen af en nedgangskonjunktur kan der dog indtræffe et produktivitsfald, idet arbejdsstyrken (hoarding) ofte ikke umiddelbart afstemmes hermed, fordi der kan herske usikkerhed om, hvordan konjunkt-

rudviklingen udvikler sig. Men i det konkrete tilfælde drejer det sig om en 12 årig periode, der i øvrigt afsluttes med en voldsom overop-hedning af dansk økonomi i 2006 og 2007, hvorfor hoarding-argumentet ikke bidrager til at forklare den anførte negative produkti-vitetsudvikling.

Den utilstrækkelige behandling af produktivi-tetsparadokset i dagligvarehandlen understre-ges af, at det intet sted er anført, at produkti-onsværdien i handelserhvervene, herunder i dagligvarebranchen, opgøres på en specifik måde i nationalregnskabsstatistikken: nemlig ved bruttoavancen. Denne opgørelsesmåde (Sørensen, 2014) kan bl.a. i perioder med sti-gende konkurrencepres føre til en undervur-dering af produktivitetsudviklingen.

For forretningsservice, som der også fo-kuseres på i Analyserapport 2, lægger kommissionen til grund, at der endog har været et gennemsnitligt fald i arbejdstime-produktiviteten på 2,0 pct. om året i 15-års-perioden 1995-2010. Med udgangspunkt heri (Analyserapport 2: 109-110) fokuseres især på de mulige negative produktivitetseffekter af ejerskabsrestriktioner i en række erhverv under forretningsservice. Men dels kan det næppe være ejerskabsrestriktioner, der har bevirket dette meget markante produktivi-tetsfald, da disse restriktioner har været gæl-dende i hele undersøgelsesperioden, og dels kan det ikke være retvisende, at der er sket et så dramatisk produktivitetsfald i denne sektor. Nationalregnskabet må enten give et forkert billede af produktiviteten i denne sektor, jf. anvendelsen af også B-metoder til deflatering, eller også må arbejdstimeproduk-tiviteten give et meget misvisende udtryk for produktiviteten i denne sektor, se herom nær-mere i del 2.

Det er meget bemærkelsesværdigt, at Pro-duktivitetskommissionen alt for ukritisk an-vender de anførte tal for dagligvarebranchen og forretningsservice. Det er måske endnu

mere bemærkelsesværdigt, at det især er dag-ligvarebranchen, som kommissionen fokuse-rer på, idet forretningsservice målt på brut-toværditilvæksten er ca. dobbelt så stor som detailhandel og omkring 7 gange større end dagligvarebranchen!

Samspil mellem forsyningsvirksomheder og andre virksomheder

Som det fremgår af skema 2, har kommis-sionen, jf. Analyserapport 5, ikke analyseret betydningen af forsyningsområdet (energi-og vandforsyning mv.) for produktivitetsud-viklingen. Det følger heraf også, at der ikke, som ellers annonceret i Analyserapport 1, er analyseret, om et bedre samspil mellem for-syningsvirksomheder og andre virksomheder kan bidrage til øget produktivitetsvækst i Danmark som helhed.

På dette område er der i slutrapporten tilføjet et 12-siders afsnit om »En omkostningseffek-tiv energiproduktion og -distribution«. Dette afsnit er især koncentreret om konsekvenser af den ambitiøse danske klimapolitik med tilhørende grøn energiomstilling for omkost-ninger og forsyningssikkerhed i en verden med ret så skiftende relative energipriser.¹ I denne sammenhæng er det endvidere ikke mindst PSO-afgiften, der er i fokus. Det er i denne sammenhæng interessant, at det net-op er PSO-afgiften, der er den klart største monetære post i regeringens efterfølgende vækstudspil »Danmark helt ud af krisen« fra maj 2014. I henhold til dette udspil skal PSO-afgiften nedsættes med 5,8 mia. kr. frem til 2020, hvilket er planens klart største post in-den for rammen på 10 mia. kr.²

Kanaliserings af investeringer

Det ses ikke, at dette emne er eksplicit be-handlet i nogen af de seks analyserapporter. Det anføres blot i Slutrapporten, at beskatnin-gen af erhvervsindkomst og kapitalafkast bør indrettes, så det samfundsøkonomiske inve-steringsafkast bliver størst.

Selskabsskatten og skattens betydning for udenlandske selskabers tilskyndelse til at investere i Danmark

Dette emne er behandlet i kapitel 6 i Slutrapporten. Hovedanbefalingen er, som det også anføres i sammenfatningen af denne rapport side 18-19, at der sker en skatteomlægning fra selskabs- og personskatter til skatter på forbrug og ejendom. I tilknytning hertil anbefales, at der nedsættes et sagkyndigt udvalg, som især skal udrede, hvordan der kan skabes et mere neutralt og symmetrisk regelsæt for beskattning af erhvervsindkomst og kapitalafkast, så kapitalen ledes derhen, hvor det samfundsøkonomiske afkast er størst. Det anbefales i tilknytning hertil, at den effektive selskabsbeskatning i Danmark ikke bør være højere end i vore nabolande.

Forskerskatteordningen analyseres i forbindelse med en bredere analyse af muligheden for at anvende skattepolitikken til at styrke produktivitetsudviklingen

Dette emne ses ikke at være analyseret. Forskerskatteordningen er udover annonceringen i Analyserapport 2 kun nævnt side 129 i Slutrapporten.

Betydningen af el-, gas- og vandforsyningen for produktiviteten

Der henvises til det anførte under *Samspil mellem forsyningsvirksomheder og andre virksomheder*, ligesom der henvises til det

anførte i skema 2 vedrørende Analyserapport 5. Her anføres direkte, at forholdene indenfor energisektoren er så komplicerede, at Produktivitetskommissionen ikke har haft mulighed for en grundling behandling af denne sektor indenfor den givne tidsramme. En tilsvarende formulering indgår også i Slutrapporten, se side 62.

Problemstillinger eller områder, der ikke er analyseret

Mens det forrige afsnit vedrørte undersøgelser, der var planlagt, men som næppe kan vurderes at være gennemført som tiltænkt, selv om Produktionskommissionen ikke forholder sig hertil, er der også problemstillinger eller områder, som ikke er analyseret, og som kommissionen eksplicit gør opmærksom på. Bortset fra de ovenfor nævnte områder indenfor infrastruktur er det forhold indenfor undervisningsområdet, som ikke er analyseret, jf. skema 2.

Kommissionens forslag til yderligere undersøgelser

Som anført fik kommissionen såvel en meget bred og omfattende opgave som en relativ kort tidsfrist til løsningen af opgaven. Dette har, som det fremgår af det ovenfor anførte, betydet, at en række planlagte analyseopgaver næppe er blevet gennemført som tilsigtet, jf. skema 1, ligesom en række opgaver direkte er blevet valgt fra, jf. skema 2.

Skema 2: Problemstillinger eller områder, der ikke er analyseret

Rapport	Side	
4	7	Efteruddannelse og andre uddannelses tilbud (udover fra grundskoler, ungdomsuddannelser via gymnasier og erhvervsskoler samt universiteter).
4	95	Erhvervsskoler: ikke detaljerede reformforslag, da regeringen i oktober 2013 kom med et udspil på området, som der arbejdes på.
4	124	Finansiering og styring af de højere læreanstalter: ingen konkret model. Der foreslås en radikal gentænkning af den måde, institutionerne er styret og finansieret på.
5	7-8, 14	Infrastrukturen på forsyningsområdet (energi- og vandforsyning mv.) diskuteres ikke. Forholdene indenfor især energisektoren er imidlertid så komplicerede, at Produktivitetskommissionen ikke har haft mulighed for en grundlig behandling af denne sektor inden for den givne tidsramme.

Men dette er ikke det eneste, der viser, hvor presset kommissionen har været. I skema 3 er det søgt sammenfattet, hvilke ekstra undersøgelser kommissionen anbefaler, at an-

dre gennemfører med henblik på at forbedre produktiviteten.

Oversigten i skema 3 viser, at Produktivitets-

Skema 3: Undersøgelser anbefalet af Produktivitetskommissionen

Rapport	Side	Art ^a	Emne
1	63	Udvalg	Bedre statistikgrundlag for bygge- og anlægsbranchen med repræsentanter fra bl.a. Danmarks Statistik og branchen. ^b
2	13, 54, 58	Udvalg	Forslag til styrkelse af de institutionelle rammer for konkurrence- og forbrugermyndighederne.
2	13, 67, 79	Udvalg	Forslag til en mere effektiv organisering og koordinering af den samlede danske eksportfremmeindsats.
2	91, 96, 105	?	Gennemgang af den branchespecifikke regulering med henblik på forenkling og tilpasning, herunder afskaffelse af ejerskabsrestriktioner.
2	114	Undersøgelse	Internationalt udbud af områder til hypermarkeder i tilknytning til lempelser af Planloven.
2	134, 140	Undersøgelse	Oprettelse af et uafhængigt selskab til administration og salg af postadresser.
3	78	Undersøgelse	Administrative fællesskaber mellem offentlige enheder.
3	100	Undersøgelse	Udvikling af praktisk ledelseskompetence i den offentlige sektor.
4	20, 83	Undersøgelse	Hvordan intensiveres børns indlæring i dagtilbuddene?
4	22, 118, 123	Undersøgelse	Om antallet af forskellige uddannelsesretninger på de videregående uddannelser med fordel kan begrænses.
4	22, 128	Undersøgelse	Om de videregående uddannelser er afstemt efter arbejdsmarkedets (samfundets) behov.
4	22, 163	Udvalg	Om vidensamarbejde og -overførsel fra universiteter til erhvervslivet kan forbedres.
4	83, 88	Undersøgelse	Er kompetencemålene i læreruddannelsen tilstrækkelige?
4	83, 88	Undersøgelse	Er mange små lærerseminarier hensigtsmæssige?
4	88	Undersøgelse	Er folkeskolelæreruddannelsen blevet mere forskningsbaseret?
4	107	Undersøgelse	Om den nuværende model med selvejende institutioner på ungdomsuddannelsesområdet er hensigtsmæssig.
4	125	Undersøgelse	Om et indkomsttaxameter kan gennemføres i praksis.
6	19, 20, 71, 79	Uafhængig central instans	Mulighed for udbud af tilkøbsydelse fra offentlige myndigheder. Erfaringer fra frikommuneforsøg inddrages.
6	20, 89	?	Betydning af frit valg for bl.a. den offentlige sektors produktivitet og effektivitet.
6	20, 110	Undersøgelse	Hvordan gøre offentlige indkøb via indkøbscentraler mere effektive?
6	41, 46	Undersøgelse	Fordele/ulempes ved fjernelse af lovbestemte barrierer for udlicitering på »bløde« områder (folkeskoler, gymnasier, erhvervsskoler).
Slut ^c	21, 64	Undersøgelse	Om teknologineutralitet kan opnås i forhold til støtte af vedvarende energi. Om en vis energiproduktion baseret på gas skal fastholdes. Om den værende PSO-støtteordning med fordel kan omlægges.
Slut ^c	21, 65	Undersøgelse	Styreformen for selskaberne i energisektoren (herunder om el-netselskaberne har de rette incitamenter).
Slut ^c	19, 114	Sagkyndigt Udvalg	Hvordan skabes et mere neutralt og symmetrisk regelsæt for beskatning af erhvervsindkomst og kapitalafkast.

- I de fleste tilfælde er det kun anført, at en undersøgelse anbefales gennemført. I nogle få tilfælde er det dog anført, at undersøgelsen foreslås gennemført af et udvalg. I et par tilfælde er der hverken anvendt betegnelser som undersøgelser eller udvalg. I disse tilfælde er der anført et ? i denne kolonne.
- Også anført side 148 i Analyserapport 2.
- Betegnelse for Slutrapporten. Der er i skema 3 kun anført undersøgelser m.v. fra Slutrapporten, hvis de ikke allerede er nævnt i de forudgående Analyserapporter.

kommissionen i hvert fald har anbefalet 24 supplerende undersøgelser. Og det kan ikke udelukkes, at der i virkeligheden er foreslået flere, da det i en række tilfælde er anført, at evalueringer af bestemte områder er ønskelig. Det er ikke klart, hvorfor kommissionen i nogle tilfælde bruger udtrykket evalueringer fremfor undersøgelser. Men evalueringer er ikke inddraget i skema 3.

Det fremgår endvidere af skema 3, at 9 ud af de 24 undersøgelser mv. vedrører uddannelsesområdet. Dette har, som anført i indledningen, været medvirkende til, at der fokuseres på uddannelsesområdet i del 2 af denne analyse.

Problemstillinger og områder, der i realiteten blev negligeret

Når et arbejde skal vurderes, er et af de mest afgørende og vanskeligste punkter at vurdere, hvad der burde være inddraget, men som ikke blev det eller kun i yderst sporadisk omfang blev inddraget. Her skal anføres to eksempler herpå, dels et område, der ikke blev inddraget: it-systemerne i den offentlige serviceproduktion og dels et område, der kun meget sporadisk blev inddraget: produktiviteten i forretningsservice.

It-systemerne i den offentlige serviceproduktion

I Produktivitetskommissionens analyser anføres gang på gang, hvor vigtige it-systemer er blevet for produktivitsudviklingen. På denne baggrund er det meget overraskende, at der ikke er overvejelser over, hvad der er baggrunden for de næsten systematiske fejlslagne større offentlige it-udviklingsprojekter. Derfor er der naturligvis heller ingen forslag til, hvordan det kan opnås, at der fremover opnås langt bedre resultater.

Kommissionen anfører (Analyserapport 6: 126) dog: »Erfaringer fra it-projekter illustrerer risikoen ved innovation i den offentlige sektor. I en analyse af statslige it-projekter

konkluderer Finansministeriet (2010) i en arbejdsgrupperapport, at vellykkede projekter snarere er undtagelsen end reglen. Der har været store overskridelser af budgetter og tidsplaner.«

Der kommer jo stadig det ene eksempel efter det andet på, at vellykkede projekter snarere er undtagelsen end reglen. Et af de seneste eksempler herpå er det fælles inddrivelsessystem, som SKAT satte i drift i september 2013, men som er forbundet med så store problemer, at kun indsættelse af meget betydelige ekstra personressourcer fra andre afdelinger i SKAT hindrer et egentligt sammenbrud.

Af ovenstående årsager burde it-projektproblemer være langt dybere undersøgt af kommissionen. Her ligger et afgørende produktivitsproblem.

Produktiviteten i forretningsservice

I Analyserapport 2 lægger Produktivitetskommissionen til grund, at timearbejdskraftproduktiviteten i branchen forretningsservice er faldet med 2 pct. om året i femtenårsperioden 1995-2010.

På denne baggrund er det dybt underligt, at kommissionen helt undlader en egentlig analyse af årsagerne til produktivitsudviklingen i forretningsservice. Der anføres (Analyserapport 2: 109-110) kun nogle overordnede betragtninger om ejerskabsrestriktioner. Interessen fokuseres i stedet på ikke mindst udviklingen i detailhervet, og herunder særligt på dagligvarebranchen. I den anvendte tolvårsperiode 1995-2007 faldt produktiviteten i detailbranchen/dagligvarebranchen med i gennemsnit 0,4 pct./1,4 pct. i henhold til nationalregnskabstallene. Altså mindre end i forretningsservice.

Negligeringen kunne være begrundet med, at forretningsservice havde langt mindre betydning i en samfundsøkonomisk sammenhæng

end detailbranchen. Men det er ikke tilfældet – tværtimod. Forretningsservice er som anført aktivitetsmæssigt (målt ved bruttoværditilvæksten) dobbelt så tungvejende som detailbranchen og syv gange så tungvejende som dagligvarebranchen. Man efterlades derfor med et indtryk af, at kommissionen nærmest har glemt forretningsservice i iveren efter at sætte lup på dagligvarebranchen. I hvert fald er den mærkelige vægtlægnings ikke forklaret.

Generel vurdering af opgaveløsningen

Produktivitetskommissionen fik som anført en meget omfattende opgave, der endvidere skulle forsøges løst inden for en relativt kort tidsperiode. Dette har, jf. ovenstående, efter min vurdering medført: at planlagte undersøgelser/analyser ikke eller kun yderst sporadisk blev gennemført, at undersøgelser/analyser eksplicit blev opgivet og at undersøgelser/analyser, som burde være gennemført i relation til hovedopgaven, blev overset eller blot opgivet. I denne sammenhæng er det også relevant at anføre den række af yderligere undersøgelser, som kommissionen anbefaler efterfølgende bliver gennemført, jf. skema 3.

Når det overordnet skal vurderes, hvad Produktivitetskommissionen har analyseret og ikke analyseret, er det vanskeligt at se bort fra, at der i analyserapporterne og Slutrapporten ikke er gjort overvejelser over, hvordan kommissionen skulle løse den stillede opgave. Det er således ikke forsøgt at identificere de afgørende årsager til den utilfredsstillende produktivitetsudvikling i Danmark siden midten af 1990'erne.

Der er derimod taget udgangspunkt i, at produktiviteten i en række hjemmemarkedsorienterede erhverv har været ringe eller ligefrem negativ bedømt ud fra tallene fra nationalregnskabet. Som det imidlertid vil fremgå af del 2, kan de anvendte produktivitetstal herfra i en række tilfælde ikke betragtes som retvisende. Men med udgangspunkt heri fo-

kuserer kommissionen på konkurrence og internationalisering som den røde tråd gennem kommissionens arbejde. Og med den brede tilgang som kommissionen har fulgt i sit arbejde fremsættes 25 hovedforslag og mere end 100 konkrete forslag. Men hvad der er væsentligt eller mindre væsentligt for produktivetsproblemet, fremgår ikke tydeligt. Tværtimod anføres, at der skal sættes ind over en bred kam.

Noget tyder dog på, at kommissionen i arbejds løb har fået øjnene op for undervisningsområdet potentielle betydning for den utilfredsstillende produktivitetsudvikling. I Slutrapporten anføres således: »... der (er) meget væsentlige svagheder ved det danske uddannelsessystem, som uden tvivl hæmmer produktiviteten. ... Skal produktivetsproblemet løses, er der behov for gennemgribende reformer i store dele af uddannelsessystemet«, side 13. Og videre: »... Produktivitetskommissionen vurderer, at det er en væsentlig årsag til Danmarks produktivetsproblem, og at der er behov for vidtgående reformer«, side 81.

I henhold til Jyllands-Posten af 29. marts 2014 blev kommissionen ligefrem overrasket over, hvor dybe problemerne var på uddannelsesområdet. At denne erkendelse har manifesteret sig relativt sent, fremgår vel også af, at kommissionen stort set indskrænker sig til at opstille indikationer for fejludviklinger – altså begrænser sig til diagnosticering – inden for undervisningsområdet. Dette er nærmere uddybet i del 2 af denne analyse.

Noter

1. Afsnittet om »En omkostningseffektiv energiproduktion og -distribution« er endvidere sammenfattet side 20-21 i Slutrapporten.
2. Det er svært at abstrahere fra muligheden af, at kommissionen har efterkommet et politisk pres, jf. at forsyningsområdet først eksplicit blev udelukket ved Analyserapport 5 fra dec. 2013 for så kort tid herefter – og med fokus på PSO-afgiften, der var hovedposten i regeringens vækststud-

spil i »Danmark helt ud af krisen«, der blev publiceret kort efter Slutrapporten – alligevel at blive inddraget i Slutrapporten fra marts 2014.

Litteraturhenvisninger

Finansministeriet (2010), *Professionalisering af arbejdet med it-projekter i staten. Afrapportering fra arbejdsgruppen vedrørende bedre statslige it-projekter*, København.

Produktivitetskommissionen (2013a), »Danmarks produktivitet – hvor er problemerne? *Analyserapport 1*, København, april.

Produktivitetskommissionen (2013b), »Konkurrence, internationalisering og regulering«, *Analyserapport 2*, København, maj.

Produktivitetskommissionen (2013c), »Styring, ledelse og motivation i den offentlige sektor«, *Analyserapport 3*, København, september.

Produktivitetskommissionen (2013d), »Uddannelse og innovation«, *Analyserapport 4*, København, dec. 2013.

Produktivitetskommissionen (2014a), »Infrastruktur«, *Analyserapport 5*, København, januar.

Produktivitetskommissionen (2014b), »Offentlig-privat samarbejde«, *Analyserapport 6*, København, februar.

Produktivitetskommissionen (2014c), »Det handler om velstand og velfærd«, *Slutrapport*, København, marts.

Regeringen (2014), *Danmark helt ud af krisen*, København, maj.

Sørensen, Christen (2014), *Produktivitetskommissionens analyse af produktivitetsudviklingen i detailhandel*, marts (Notat udarbejdet til De Samvirkende Købmænd. Se hjemmesiden www.envis-mand.dk under fanebladet Netop Nu).

Review:

Produktivitetskommissionen – en analyse.

Del 2: Produktivitetsmål og uddannelse

Christen Sørensen, fhv. overvismand
cs020345@gmail.com

Produktivitetskommissionen anvendelse af produktivitetstal og gennemgang af uddannelsesområdet er emnet for denne analyse.

Kommissionen var bl.a. bemandet med 6 professorer i økonomi og havde et sekretariat med 8 økonomer, hvorfor der burde være tilstrækkelig viden om, at det især for en række serviceerhverv kan være store fejlvisninger i de produktivitetstal, der kan udledes af nationalregnskabsstallene. Alligevel blev disse tal anvendt alt for ukritisk. Dette dokumenteres først og fremmest for dagligvarebranchen, hvor kommissionen baserede anbefalinger på, at der i den undersøgte 12-årsperiode, 1995-2007, havde været et årligt produktivitetsfald på 1½ pct. Det vises, at dette tal kan ændres til ca. + 1 pct., såfremt der korrigeres for åbenbare fejkilder og ikke medtagne forhold.

Uddannelsesområdet er afgørende – også i relation til produktivitet. Kommissionen giver en nærmest sønderlemmende kritik af uddannelsesområdet. Men diagnosen af svaghederne følges desværre ikke op med konkrete løsningsforslag. Tidspresset har givetvis været medvirkende hertil.

Produktivitetsmål og uddannelse

I del 1 af denne analyse blev Produktivitetskommissionens arbejde generelt analyseret. Analysen i denne del er koncentreret om to hovedemner: produktivitetsmål og uddannelse.

Produktivitetskommissionen anvendte af empiriske grunde et skævt produktivitetsmål: time-arbejdskraftproduktiviteten. Dette kombineret med, at Nationalregnskabet for en række erhverv eller brancher ikke giver retvisende mål for produktionsudviklingen, har resulteret i åbenbare problemer. Trods dette baserer kommissionen sig i alt for høj grad på, at der kunne konkluderes på talgrundlaget fra Nationalregnskabet. Dette gælder f.eks. i relation til dagligvarebranchen, hvor kommissionen i realiteten lægger til grund, at der i tolvårsperioden 1995-2007, som anvendes i deres analyse, har været et fald i time-arbejdskraftproduktiviteten på næsten 1½ pct. om året! Sammenlignet hermed er det underligt, at kommissionen stort set undlader at undersøge forholdene i forretningsservice, hvor der i en endda 15-årig periode, 1995-2010, skulle have været et årligt fald i time-arbejdskraftproduktiviteten på 2 pct. Dette skal endvidere ses i lyset af, at aktiviteten målt ved bruttoværditilvæksten er omkring 7 gange større i forretningsservice end i dagligvarebranchen.

Produktivitetskommissionens analyse af uddannelsesområdet er koncentreret i Analyse-

rapport 4. Kommissionen leverer en nærmest sønderlemmende kritik af uddannelsesområdet, hvor et enormt ressourceforbrug er kombineret med meget dårlige resultater. Selv om kommissionen i henhold til dets kommissorium skulle fremlægge konkrete forslag, er dette ikke sket i relation til uddannelsesområdet. Det er i øvrigt heller ikke undersøgt, hvad der ligger bag det store ressourceforbrug. Det kan ikke være til en intensivering af uddannelsesindsatsen, idet det på mange måder påvises, at der heri er sket et markant fald.

Produktivitet og talgrundlaget

Produktivitetskommissionen tog som anført i del 1 overordnet udgangspunkt i de produktivitetstal, der kan udledes fra Nationalregnskabet. Produktivitet kan imidlertid måles på flere måder. Tages der hensyn til udviklingen i indsatsen af alle produktionsfaktorer, er det mest retvisende mål den såkaldte totalfaktorproduktivitet (TFP). Når produktivitet måles med udgangspunkt i bruttoværditilvæksten (BVT), skal der korrigeres for den værditilvækst, der alene er en følge af mængdemæssige ændringer i indsatsen af arbejdskraft og realkapital, når restbidraget – totalfaktorproduktiviteten – skal beregnes.

Beregninger af den mængdemæssige indsats af såvel arbejdskraft som kapital er såvel teoretisk som praktisk udfordrende, fyldt som opgørelser heraf er med mange antagelser og tilnærmelser, som det fremgår af den lange litteratur herom. Der hugges såvel tæer som hæle i den proces, især når det drejer sig om opgørelser af realkapitalen. Her skal alene henvises til de bidrag (Andersen, 1964a og 1964 b), som ADAM-modellens grundlægger, Ellen Andersen, tidligt udgav.

Dette forhold har naturligvis også præget Produktivitetskommissionens arbejde, idet kommissionen i de anvendte produktivetsberegninger dels helt har set bort fra indsatsen af realkapital og dels har approksimeret udviklingen i indsatsen af arbejdskraft med

udviklingen i antallet af præsterede arbejdstimer. Når udviklingen i indsatsen af arbejdskraft måles på den måde, ses der bort fra det bidrag fra arbejdskraftindsatsen, der hidrører fra ændringer i arbejdskraftens sammensætning i retning af f.eks. et højere uddannelsesniveau.

Produktivitetskommissionen (Analyserapport 1: 21-23) begrundet sit valg af produktivitetsmål – arbejdskraftens timeproduktivitet – på følgende måde:

»TFP er teoretisk set det mest korrekte produktivitetsmål, men der er store udfordringer forbundet med at anvende det i praksis. ...

De empiriske problemer med TFP er vanskelige at løse, og et alternativ er at benytte et simple mål. Simple produktivitetsmål er dog samtidig mindre fyldestgørende set fra et teoretisk perspektiv. ...

Arbejdskraft er den ressource, som er nemmest at måle. De fleste lande har ret præcise opgørelser over, hvor mange i befolkningen der er i beskæftigelse. Et empirisk robust produktivitetsmål er derfor BVT per beskæftiget. BVT per beskæftiget er dog robust, netop fordi det er simpelt. Det tager fx ikke højde for, hvor mange timer de beskæftigede arbejder på et år. Det er u hensigtsmæssigt, da fritid er et gode på linje med de forbrugsgoder, en beskæftiget kan købe for sin arbejds løn. ...

Et mere interessant produktivitetsmål er derfor timeproduktiviteten, det vil sige BVT per arbejdet time. Timeproduktiviteten er et bredere begreb end TFP. Benyttes mere kapital, f.eks. maskiner, i produktionen, øger det timeproduktiviteten. TFP er derimod uændret. Et højere uddannelsesniveau for arbejdsstyrken virker på samme måde. Det er både en styrke og en svaghed ved timeproduktiviteten i forhold til TFP. Det er en styrke, fordi det afspejles i timeproduktiviteten, hvis vi i Danmark har været for dårlige til at uddanne vores arbejdsstyrke, eller hvis virksomheder har svært ved at låne til nye investeringer. Omvendt er det en svaghed, fordi omkostningerne til investeringer i både kapital og uddannelse ikke tages i betragtning. Overdrevne investeringer kan hæve timeproduktiviteten, men det kan ske på bekostning af samfundets samlede velfærd.

Produktivitetskommissionen vurderer dog, at timeproduktiviteten har den bedste balance mellem at

være teoretisk relevant og empirisk robust. Når det er sagt, er der forsat en række måleproblemer. ...«

Produktivitetskommissionen overvurderer dog klart fordelene ved at vælge arbejdskraftens timeproduktivitet som mål for produktivitsudviklingen. Det er således ret så kunstigt at anføre, at det er en styrke ved dette mål, at det afspejles heri, »hvis vi i Danmark har været for dårlige til at uddanne vores arbejdsstyrke, eller hvis virksomheder har svært ved at låne til nye investeringer«. Dette fremgår jo direkte meget bedre af udviklingen i arbejdskraftens uddannelsesniveau og af investeringsudviklingen. Det eneste holdbare argument for at vælge arbejdskraftens timeproduktivitet som produktivitetsmål fremfor totalfaktorproduktiviteten er, at dette mål er enklere at beregne. Hertil kunne tilføjes, at usikkerheden på arbejdskraftens timeproduktivitet er mindre end usikkerheden på totalfaktorproduktiviteten, men det kan næppe være afgørende, da arbejdskraftens timeproduktivitet i sig selv er et skævt mål.

Alligevel skal dette valg ikke kritiseres, selv om dette produktivitetsmål på forhånd kan være skævvredet af to forhold. Dette vil for det første være tilfældet, såfremt indsatsen af realkapital og arbejdskraft ikke udvikler sig parallelt. Og det vil for det andet være tilfældet, såfremt arbejdskraftens sammensætning ændres. Sædvanligvis forudsætter økonomer, og også Produktivitetskommissionen (Analyserapport 1: kapitel 11), at lønforskelle afspejler arbejdstagernes produktivitsforskelle. Hvis f.eks. personalesammensætningen i en branche er skiftet i retning af anvendelse af relativt flere ikke-faglærte, dvs. til arbejdskraft med en lavere økonomisk effektivitet, undervurderes produktivitsudviklingen ved Produktivitetskommissionens opgørelsesmetode.

Det, der her kritiseres, er derimod kommissionens alt for ukritiske fortolkning og anvendelse af de fra nationalregnskabstallene

udledte produktivitetsmål. De to mest markante eksempler herpå i Produktivitetskommissionens arbejde vedrører underbranchen dagligvarehandel inden for detailhandlen og forretningsservice. I begge de anførte erhverv har der i henhold til kommissionen (Analyserapport 2: kapitel 11-14) været en meget utilfredsstillende produktivitsudvikling.

For detailbranchen har der således kun været en stigning i timearbejdskraftproduktiviteten på 0,4 pct. om året i den anvendte undersøgelsesperiode fra 1995-2010. For delbranchen, dagligvarehandel, er tallene ligefrem bemærkelsesværdige, idet de angiver et produktivitsfald på hele 1,4 pct. om året i den undersøgelsesperiode 1995-2007, der anvendes af Produktivitetskommissionen for underbrancherne i detailhandlen: dagligvarer, varehuse, apoteker og materialister, tøj og sko samt øvrige butikker.

I forretningsservice, der udøves i virksomheder med advokater, revisorer, konsulenter, ingeniører, arkitekter, landmålere o. lign., har der i Nationalregnskabet i undersøgelsesperioden fra 1995-2010 ligefrem været et endnu mere bemærkelsesværdigt fald i timearbejdskraftproduktiviteten på i gennemsnit 2 pct. om året.

Produktivitet går normalt kun én vej, nemlig opad. I begyndelsen af en nedgangskonjunktur er det dog ikke usædvanligt, at der kan indtræffe et produktivitsfald, idet arbejdsstyrken ofte ikke umiddelbart afstemmes hermed (hoarding), fordi der kan herske usikkerhed om, hvordan konjunkturudviklingen udvikler sig.

Men i det konkrete tilfælde vedr. dagligvarehandlen drejer det sig om en 12 årig periode, der i øvrigt afsluttes med en voldsom overophedning af dansk økonomi i 2006 og 2007, hvorfor hoarding-argumentet ikke bidrager til at forklare den anførte negative produktivitsudvikling. Og for forretningsservice

er det anførte produktivetsfald endnu mere misvisende.

Ses der på udviklingen bag de anførte gennemsnitstal for dagligvarehandlen, jf. figur 1, sættes der yderligere spørgsmålstejn ved talgrundlaget bag time-arbejdskraftproduktiviteten i dagligvarehandlen.

Hvordan skal det f.eks. økonomisk kunne forklares, at time-arbejdsproduktiviteten i dagligvarebranchen fra 1995 til 2003 er faldet med over 25 pct.? Der var jo heller ikke borgerkrig i denne periode i Danmark!

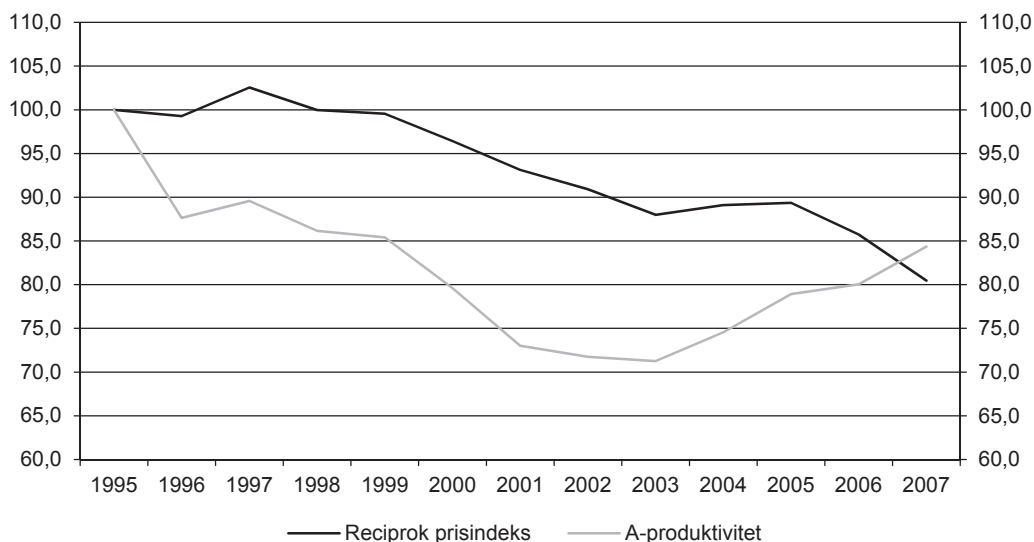
Der kan umiddelbart peges på tre årsager til den lidet troværdige udvikling i den af Produktivitetskommissionen anvendte opgørelse af produktivetsudviklingen for dagligvarehandlen: 1) anvendelse af time-arbejdskraftproduktivitet frem for totalfaktorproduktivitet, 2) nedgang i arbejdskraftens kvalifi-

kationsniveau og 3) den specielle opgørelse af produktionsværdien i handelserhvervene, herunder dagligvarehandlen, ved bruttoavance og de hertil hørende problemer i relation til omregningen til faste priser af produktionsværdien.

Det er efter min vurdering (Sørensen, 2014) givetvis den tredje årsag, der er helt afgørende for de ret så mærkelige produktivitetstal. Kommissionen afviste (Produktivitetskommissionen, 2014d) imidlertid meget bastant min kritik af deres – efter min vurdering – alt for ukritiske anvendelse af nationalregnskabsstallene. Men de har efter det for mig foreliggende ikke forholdt sig til, om de selv kunne tro på de beregnede tal for produktivetsudviklingen i dagligvarebranchen.

Da jeg imidlertid ikke kan tro på så mærkelige tal, som er anvendt i relation til dagligvarebranchen, har jeg i en konstruktiv dialog

Figur 1: Timearbejdskraftproduktiviteten og den reciproke implicite prisindeks for produktionsværdien i dagligvarehandlen 1995-2007. Indeks med 1995=100



Kilde: Danmarks Statistik. Statistikbanken.

med Danmarks Statistik nærmere undersøgt de beregningsmetoder, der anvendes i denne sammenhæng.¹

Denne undersøgelse kan sammenfattes i to overordnede hovedresultater: 1) den deflator, der anvendes til at omregne produktionsværdien, dvs. bruttoavancen, i dagligvarehandlen, er overvurderet af to årsager, hvorfor den mængdemæssige udvikling i produktionsværdien er undervurderet og 2) mangelfulde data om bruttoavancens opsplitning i mængder og priser kan resultere i betydelige fejlvisninger. Det er især det andet hovedresultat, der må antages at være afgørende i den betragtede periode, hvor åbningstiderne blev markant udvidet i tilknytning til en række liberaliseringer af lukkeloven.²

Deflatoren i tilknytning til beregningen af bruttoavancen i faste priser

Det er som udgangspunkt vigtigt at være opmærksom på, at der til grund for udarbejdelsen af nationalregnskabsstatistik kun er et ufuldstændigt datagrundlag til rådighed, hvorfor der må anlægges en række antagelser, udøves en række skøn og undlades en række korrektioner på grund af manglende grundlag herfor, ligesom nationalregnskabsstatistik er styret af en række internationale konventioner hidrørende fra såvel FN som EU. En sådan konvention er bl.a. den specielle opgørelse af produktionsværdien i handels erhverv, herunder dagligvarebranchen, ved bruttoavancen. Denne konvention skal bl.a. ses i tilknytning til, at analyse af nationalregnskabstal med anvendelse af input-output-tabeller derved bliver langt mere oplysende.

Den indbyrdes sammenhæng i nationalregnskabsstatistikken betyder, at der skal være nogenlunde overensstemmelse med den prisudvikling, som forbrugerprisindekset viser for dagligvarer, og den prisudvikling, som kan beregnes ud fra dagligvarebutikkernes indkøbspriser forhøjet med de bruttoavancer fordelt på varegrupper, som Danmark Stati-

stik anvender. Disse fordelinger af bruttoavancen i løbende priser, der foretages med udgangspunkt i virksomhedernes regnskabstal for den samlede bruttoavance, er nærmest helt skønsmæssige. Hertil kommer, at bruttoavancen på brancheniveau – også i løbende priser – ligeledes er skønsmæssigt fastsat for dagligvarebranchen, idet en række virksomheder forhandler andet end dagligvarer, bl.a. Dansk Supermarked. Men det er ifølge sagens natur ikke muligt at afgøre, om og i hvor høj grad disse fejlkilder evt. medvirker til de mærkelige produktivitetstal for dagligvarebranchen, idet disse korrektioner kan gå begge veje.

Forbrugerprisindekset overvurderer imidlertid prisudviklingen. Dette skyldes grundlæggende, at forbrugerprisindekset beregnes som et fastvægtsindeks, hvor det endvidere er lagt til grund, jf. konventioner om EU-forbrugerprisindekset HICP (Harmonised Index of Consumer Prices), at der ses bort fra forskydninger i butiksstrukturen. Det skal understreges, at dette er et helt bevidst valg. Imidlertid fører dette valg til skæve resultater for udviklingen i dagligvarebranchens produktion i faste priser – dvs. bruttoavancen i faste priser – og dermed også for produktivtetsudviklingen.

Som tidligere anført (Sørensen, 2014) bør der i perioden 1995-2007 ikke ses bort fra discountbutikkernes stigende markedsandel, når produktivitetstal skal fortolkes. I 1995 stod discountbutikkerne således for ca. 15 pct. af omsætningen af dagligvarer i Danmark, i 2007 var denne andel steget til 28,8 pct. Det er endvidere ikke urimeligt at antage, at priserne for identiske varer er ca. 10 pct. lavere i discountbutikker end i alm. dagligvarebutikker.³

Men denne forskel i prisniveauet kommer ikke til udtryk i forbrugerprisindekset, idet det implicit antages, at forskellen alene modsvaret en mængdeforskel. Rent teknisk sker

det ved at anvende Jevons-indeks i beregningen af prisudviklingen for de ca. 500 primærgrupper i forbrugerprisindekset. Herved ses der (Sørensen, 2002: 132-145) bort fra prisniveauforskelle.

Denne neutralisering af forskelle i prisniveau sker også, når vægtgrundlaget i beregningen af forbrugerprisindekset med mellemrum skiftes, og der foretages en sammenkædning af prisudviklingen. Vægtgrundlaget i forbrugerprisindekset blev i perioden 1995 til 2007 opdateret i 1996, 1999, 2002 og 2005. Når der kædes, inddrages de forskellige butiksførers andele af omsætningen. Men for at undgå hop i prisudviklingen i tilknytning til kædningen, ses der også i kædningssituationen bort fra forskelle i prisniveau. Disse antagelser betyder, at det kun er de relative prisstigningstakter mellem discountbutikker og almindelige dagligvarebutikker, der kommer til udtryk i forbrugerprisindekset. I virkeligheden kan dette føre til paradoksale resultater. Antag f.eks., at discountbutikkerne i indtrængningsfasen midlertidigt har fastholdt et ekstra lavt prisniveau for at vinde markedsandele, og at de efterfølgende – efter et vægtskift i forbrugerprisindekset – hæver priserne mere end den almindelige prisudvikling og derfor indsnævrer prisniveauforskellene til almindelige dagligvarebutikker. I et sådant forløb vil discountbutikkernes højere markedsandel føre til en ekstra – fejlagtig – stigning i forbrugerprisindekset.

Jeg kan på det teoretiske plan godt forstå, at det er muligt at opfatte prisforskelle mellem discountbutikker og alm. dagligvarebutikker som udtryk for en ren mængdeforskel begrundet i antagne kvalitetsforskelle. Men det er en ret ekstrem antagelse. Med denne antagelse burde der naturligvis også korrigeres for de kvalitetsforskelle, dvs. mængdeforskelle, der er mellem varer indkøbt i de sædvanlige åbningstider og i de øvrige åbningstider, hvor lønniveauet er højere, jf. nedenfor. I det

mindste bør der tages hensyn hertil i fortolkningen af forbrugerprisindeks mv.

Begge de anførte forhold betyder, at forbrugerprisindekset umiddelbart overvurderer prisstigningerne på dagligvarer. Denne overvurdering betyder, at der også er en afsmitende virkning til den deflator, der anvendes, når bruttoavancen i løbende priser, dvs. produktionsværdien i løbende priser, omregnes til faste priser. Hermed undervurderes også produktivitetsudviklingen. I ikke mindst Danmark, hvor discountbutikkerne har fået en meget stor markedsandel, er dette af betydning. Spørgsmålet er, hvor meget dette kan ændre på den målte – og forkert opgjorte – produktivitetsudvikling i dagligvarebranchen.

Først opstilles i tabel 1 skøn over, hvor stor betydning den manglende hensyntagen til prisforskelle mellem discountbutikker og alm. butikker i forbrugerprisindekset kan betyde som følge af, at der ses bort fra den øgede markedsandel for discountbutikker. Som det fremgår af tabel 1, baseres dette skøn på: 1) et 10 pct. lavere prisniveau i discountbutikker end i almindelige dagligvarebutikker og 2) en stigning i discountbutikkernes markedsandel fra 15 pct. til 28,8 pct., som det var tilfældet i perioden fra 1995 til 2007.

Det fremgår af tabel 1, at fejlvisningen hidrørende fra beregningsreglerne for forbrugerprisindekset kun er på 1,4 pct. for hele perioden 1995-2007. Denne skævhed slår imidlertid flere gange igennem, nemlig med en faktor på ca. 5, da bruttoavancen udgør omkring 20 pct. af omsætningen i dagligvarebranchen. Denne fejlkilde fra forbrugerprisindekset bidrager til at forklare en del af de mærkelige produktivitetstal for dagligvarebranchen, men er langt fra tilstrækkelig, da en fejlvisning på ca. 7 pct. for hele perioden 1995-2007 kun kan forhøje den årlige produktivitetsudvikling fra -1,4 pct. til -0,8 pct. Denne konklusion forstærkes yderligere af, at

Tabel 1: Overvurdering af prisniveauet i perioden 1995 til 2007 i forbrugerprisindekset som følge af stigende markedsandel til discountbutikker

	Prisniveau ^a	Omsætningsandel i 1995, pct.	Vægtet prisniveau i 1995	Omsætningsandel i 2007, pct.	Vægtet prisniveau i 2007	Overvurdering af prisniveau 1995→2007 ^b
Discount	0,90	15	13,50	28,8	25,92	
Andre forretn.	1	85	85,00	71,2	71,20	
I alt			98,50		97,12	1,4 pct.

a. Med prisniveauet i almindelige dagligvarebutikker sat til 1.

b. Prisfald på grund af ændret butiksfordeling, som ikke indgår i forbrugerprisindekset.

det ikke nødvendigvis gælder, at udviklingen i forbrugerpriserne fuldt ud slår igennem i den deflator, der bruges i relation til opgørelsen af bruttoavancen i faste priser i dagligvarebranchen.

Stigende konkurrence og længere åbningstider

Det følgende eksempel tilsigter at pege på, hvad der kan være hovedårsagen til, at nationalregnskabstallene viser en så negativ produktivitetsudvikling for dagligvarebranchen, som det var tilfældet i perioden fra 1995 til 2007, hvor timearbejdskraft-produktiviteten i henhold hertil skulle være faldet med hele 1,4 pct. i gennemsnit pr. år.

Antag, at stigningen i discountbutikkernes markedsandel i dagligvarebranchen fra 15 pct. i 1995 til 28,8 pct. i 2007 har resulteret i et fald i bruttoavanceprocenten fra 21 til 18 i løbet af denne 12-årige periode. Antag endvidere, at den mængdemæssige omsætning i branchen er uændret, hvorfor det eneste, der er sket, er en ændret fordeling af omsætningen fra almindelige dagligvarebutikker til discountbutikker med en lavere bruttoavanceprocent til følge. Faldet i bruttoavanceprocenten fra 21 til 18 betyder under de givne forudsætninger, at bruttoavancen i løbende priser er faldet med 3/21-dele eller med 14,3 pct. Det antages endvidere, at dette alene er udtryk for et prisfald, hvorfor bruttoavancen i faste priser er uændret.

Antag alternativt, at bruttoavanceprocenten ikke ændres, idet en udvidelse af åbningstiderne antages at betyde, at priseffekten fra discountbutikkernes øgede markedsandel netop opvejes af øgede omkostninger som følge af den udvidede åbningstid. Preiseffekten fra discountbutikkernes større markedsandel er naturligvis stadig isoleret set 14,3 pct. Dette antages under de givne forudsætninger opvejet af prisstigninger begrundet i øget åbningstid, idet der i øvrigt – fysisk set – omsættes samme kvanta som i udgangseksemplet. Heraf følger, at der, grundet kvalitetsforøgning som følge af øgede køb uden for den sædvanlige åbningstid, har været en mængdemæssig stigning i bruttoavancen i faste priser på 14,3 pct., idet bruttoomsætningen i løbende priser jo var uændret.

Esensen af dette eksempel er, at når produktionsværdien opgøres ved et differencebeløb, som bruttoavancen jo er, kan manglende oplysninger om årsager til fastsættelse af bruttoavanceprocenten have meget stor betydning for opsplitningen i hhv. pris- og mængdeeffekt – i det konkrete eksempel næsten 15 pct. Her ligger givetvis den væsentligste årsag til de mærkelige negative produktivitetstal i dagligvarebranchen.

Det var netop dette forhold, som jeg også tidligere (Sørensen, 2014) især har lagt vægt på i min kritik af kommissionens talbehandling for dagligvarebranchen. Produktivitetskom-

missionen afviste imidlertid blankt denne kritik, jf. boks 1.

Boks 1. Produktivitetskommissionens afvisning af min kritik

»Dette (at kvalitetsforbedringer skal henføres til mængdeændringer, tilføjelse) er Danmarks Statistik helt på det rene med, hvorfor der også tages højde for problemet. Af dokumentationen af metoder og kilder til fastprisberegninger fremgår følgende (Danmark Statistik, 2002: 23-24):

»For konsumgrupperne er udviklet et system, som automatisk korrigerer for ændringer i avanceprocenterne i årets priser, under forudsætning af, at dette bringer resultatet nærmere udviklingen i forbrugerprisindekset. Idéen bag tilretningen er, at ændringer i avanceprocenten i løbende priser for en dels vedkommende har baggrund i en ændret kvalitet af servicen i handelsleddet, hvilket skal afspejle sig i fastprisværdien af avancen.«, Produktivitetskommissionen (2014d, side 2).

Afvisningen i boks 1 skyder imidlertid helt ved siden af, idet det jo forudsættes, at der er korrigeret tilstrækkeligt for kvalitetsforbedringer i beregningen af forbrugerprisindekset. Som det fremgår af det ovenfor anførte, er dette ikke tilfældet, idet de anførte to forhold også slår igennem i en overvurdering af prisudviklingen i forbrugerprisindekset. Produktivitetskommissionens afvisning kan derfor overhovedet ikke tillægges vægt, jf. også at det citerede fra Danmarks Statistik vedrører afstemningen i nationalregnskabet.

At det givetvis er omregningen fra bruttoavancen (produktionsværdien) fra løbende til faste priser, der giver anledning til de mærkelige produktivitetstal for dagligvarebranchen, underbygges yderligere af illustrationen i figur 1, der foruden udviklingen i timearbejdskraftproduktiviteten også viser udviklingen i prisindekset for produktionsværdien. Det er den inverse/reciprokke prisudvikling, der er vist på indeksform med 1995=100. Dette fremhæver, at der er en positiv korrelation mellem de to kurver i figur 1 – en korrelation, der i øvrigt ville være højere, hvis de årlige

fejlvisninger i prisudviklingen blev akkumuleret.⁴

Den positive korrelation underbygger yderligere, at prisudviklingen overvurderes, og dermed at produktivitetsudviklingen undervurderes.

Fra negativ til positiv produktivitetsudvikling i dagligvarebranchen

Det er bruttoværditilvæksten og ikke produktionsværdien (bruttoavancen), som Produktivitetskommissionen har anvendt i deres produktivetsberegninger. I dagligvarebranchen udgjorde værditilvæksten i gennemsnit under 2/3 af produktionsværdien i perioden 1995-2007. Det betyder, at den samlede effekt af de to anførte mulige fejlvisninger på 7+14 pct. skal forhøjes med 3/2, altså til 33 pct., når effekten på bruttoværditilvæksten skal beregnes.⁵ Forøges bruttoværditilvæksten i faste priser i dagligvarebranchen med 33 pct. over perioden 1995 til 2007, forøges den årlige produktivetsstigning fra -1,4 pct. til 1 pct.

Ovenstående forhold er mere end tilstrækkelige til at påvise, at de af Produktivitetskommissionen fremlagte tal for produktivitetsudviklingen i dagligvarebranchen, -1,4 pct. om året i perioden 1995-2007, er helt skæve. Timearbejdskraftproduktiviteten har givetvis været positiv i denne periode på hele 12 år. Men det er på det foreliggende grundlag ikke muligt at afgøre, om den årlige produktivetsstigning har været større eller mindre end 1 pct.

Et er, at Produktivitetskommissionen publicerede meget mærkelige tal for produktivitetsudviklingen i bl.a. dagligvarebranchen. Noget andet og langt mere kritisabelt er, at kommissionen drog meget vidtgående konklusioner (Sørensen, 2014) på basis af disse tal, herunder at planloven skulle ændres for at muliggøre etableringen af langt flere hypermarkeder.

Produktivitet og virksomhedsdynamik (Produktivitetsparadokset)

Hvor forsigtig man skal være med uden videre at anvende produktivitetstal, fremgår også af Produktivitetskommissionens betragtninger (Analyserapport 2: kapitel 3) om virksomhedsdynamik, som kommissionen tillægger en afgørende betydning for produktivitsudviklingen, jf.:

»Virksomhedsdynamik – undertiden kaldet kreativ destruktion – er væsentlig for produktivitsudviklingen. Virksomhedsdynamik betyder, at højproduktive virksomheder udvider produktionen og vinder markedsandele, samtidig med at de lavproduktive indskrænker eller helt lukker og erstattes af virksomheder med højere produktivitet.«

Virksomhedsdynamikkens betydning for produktivitsudviklingen uddybes (Analyserapport 2: 29) efterfølgende med:

»Virksomhedsdynamikkens bidrag til produktivitsvæksten kan opdeles i tre dele. Omfordeling af ressourcer blandt etablerede virksomheder, etablering af nye virksomheder og ophør af virksomheder.

Omfordeling af ressourcer blandt etablerede virksomheder giver et positivt bidrag til produktivitsvæksten, hvis virksomheder, der har en produktivitet over gennemsnittet, øger antallet af beskæfti-

gede, og hvis virksomheder, der har en produktivitet under gennemsnittet, mindsker antallet af beskæftigede. Tilsvarende giver etablering og ophør af virksomheder et positivt bidrag, hvis de har en produktivitet henholdsvis over og under gennemsnittet.«

Og videre (Analyserapport 2: 31):

»Ved vurderingen af resultaterne i tabel 2 er det vigtigt at holde sig for øje, at der er problemer med at måle produktiviteten i serviceerhvervene ... Det er vanskeligt at opdele nominel produktivitsvækst i bidrag fra kvalitet, mængde og pris. Den registrerede stigning i deflatorerne for flere af brancherne i tabel 2 har været overraskende høj. Det kan derfor ikke udelukkes, at den reale produktivitsvækst i disse brancher er undervurderet som følge af vanskelighederne ved at omregne fra nominel til real vækst.

Som dokumenteret ... har den registrerede reale produktivitsvækst siden 1995 i brancher som fx vidensservice og landtransport imidlertid været væsentligt lavere herhjemme end i en række relevante sammenligningslande som Sverige, Holland, Tyskland og USA. Disse forskelle i registreret produktivitsvækst er så store, at de næppe fuldt ud kan forklares ved målefejl. Det er derfor Produktivitetskommissionens vurdering, at de hjemmemarkedsorienterede servicebrancher i Danmark har haft en svag produktivitsudvikling i forhold til relevante sammenligningslande.«

De tal, som Produktivitetskommissionen

Tabel 2: Timeproduktivitet mv. i de fem største hjemmemarkedsorienterede serviceerhverv 2001-2009

Erhverv	Nominel vækst, pct. p.a.	Prisstigning, pct. p.a.	Real vækst, pct. p.a.	Heraf bidrag fra dynamik, pct. p.a.
Detailhandel	2,9	1,0	1,9	0,8
Arkitekt- og ingeniørvirksomhed	4,2	6,3	-2,0	-2,6
Juridisk bistand, bogføring og revision	4,6	5,5	-0,9	-1,2
Handel med biler og Motorcykler	3,0	3,5	-0,5	-1,0
Landtransport	1,4	5,3	-3,8	-2,7

Anm.: Vækst i arbejdsproduktivitet (værditilvækst per fuldtidsbeskæftiget). Prisindeksene er opgjort med A-metoder inden for juridisk bistand, bogføring og revision; både A- og B-metoder inden for landtransport; og B-metoder inden for detailhandel, arkitekt- og ingeniørvirksomhed og handel med biler og motorcykler.

Kilde: Analyserapport 2, side 31.

har beregnet for de fem største hjemmemarkedsorienterede serviceerhverv, og som her er gengivet i tabel 2, viser tydeligere, at der ikke kan fæstes lid til de produktivitetstal, der for en række erhverv kan udledes af nationalregnskabstallene og/eller grundlaget herfor. For de to underbrancher i forretningsservice, som indgår i tabel 2, arkitekt- og ingeniørvirksomhed samt juridisk bistand, bogføring og revision, fremgår det, at det er dynamikbidraget, der umiddelbart er den afgørende årsag til den registrerede negative realvækst i timearbejdsproduktiviteten. Så mærkværdige tal burde have fået Produktivitetskommissionen til helt anderledes kritisk at forholde sig til talgrundlaget. Det skal dog anføres, at der for detailhandel er beregnet et positivt bidrag fra virksomhedsdynamik for perioden 2001-09 på 0,8 pct. om året. Men som det også fremgår af figur 1, der dog alene vedrører underbranchen dagligvarehandel under detailhandel, er det især i perioden frem til 2002, at der synes at være alvorlige problemer med datagrundlaget.

Uddannelsesområdet

Det anføres (Slutrapporten: 13): »... der (er) meget væsentlige svagheder ved det danske uddannelsessystem, som uden tvivl hæmmer produktiviteten. ... Skal produktivetsproblemet løses, er der behov for gennemgribende reformer i store dele af uddannelsessystemet« og videre (Slutrapporten: 81): »... Produktivitetskommissionen vurderer, at det er en væsentlig årsag til Danmarks produktivetsproblem, og at der er behov for vidtgående reformer«.⁶

Det fremgår af kommissoriet, at kommissionen skal komme med konkrete anbefalinger, der kan styrke produktiviteten. Med den vægt, som kommissionen tillægger uddannelsesområdet – især inden for de videregående uddannelser – er det bemærkelsesværdigt, at kommissionen ikke har analyseret konkrete modeller for finansiering og styring af disse, men begrænser sig til at fastslå, at »der kræ-

ves en radikal gentænkning af den måde, institutionerne er styret og finansieret på«.

Rent diagnosemæssigt er Analyserapport 4 givetvis den bedste af de 6 analyserapporter. På i hvert fald et helt centralt område kommer kommissionen dog heller ikke på dette område igennem. Det anføres flere gange i Analyserapport 4, at Danmark ligger helt i top (blandt OECD-landene), når uddannelsesudgifterne relateres til BNP. Dette gælder endda, selv om udgifterne til uddannelsesstøtte/SU-støtte, hvor Danmark jo giver »verdens suverænt højeste studiestøtte«, ikke er inkluderet. Samtidig med at udgifterne ligger helt i top, er antallet af undervisningstimer på en række studier drastisk nedskåret gennem de senere år. Følgende citater fra Analyserapport 4 viser dette:

»Der undervises meget lidt på mange videregående uddannelser. På en række universitetsuddannelser modtager de studerende under ti timers undervisning om ugen (inklusive vejledning)«, side 13.

»Hvis studieintensiteten på læreruddannelsen blev løftet, så den svarer til et fuldtidsstudium, hvor de studerende bruger 37,5 timer om ugen, vil det svare til to års ekstra uddannelse med samme studieintensitet som i 2010. Samtidig skal de 13 timers undervisning ses i lyset af, at der i 1986 var 25 ugentlige undervisningstimer på lærerseminarierne«, side 87.

»Mange studier tilbyder få undervisningstimer, ofte under 10 timer om ugen. Det betyder, at studieintensiteten på mange uddannelser svarer til et deltidsjob«, side 117.

Men Analyserapport 4 giver ikke svar på, hvad de mange ressourcer bruges på, når det nu ikke er udgifter til undervisning, der synes at kunne forklare den danske rekord i uddannelsesudgifter. Er det til bureaukrati og hertil stærkt stigende udgifter i form af bl.a. højere lønninger til ledelseslaget eller til hvad?

Som sagt er diagnosedelen udmærket i Analyserapport 4. Og det er en så dystre vurde-

ring, der fremsættes, at man kun kan undre sig over, at der ikke har været langt større reaktioner herpå. Følgende citater viser dette. Citaterne er her opdelt i fem hovedgrupper: generelt om uddannelsessystemet, grunduddannelser, ungdomsuddannelser, videregående uddannelser og styringen af uddannelsessystemet.

Generelt om uddannelsessystemet

»Resultaterne er nedslående.«, side 7

»... det er gået den forkerte vej med uddannelsessystemet ...«, side 10

»... får (vi) nok ud af de ressourcer, vi allerede bruger på uddannelsessystemet. ... nej«, side 27

»... det danske uddannelsessystem er blevet overhalet af andre lande«, side 29

»Derimod er der tegn på, at læringen hos elever og studerende halter, og at de kompetencer, mange opnår gennem uddannelse, ikke er direkte anvendelige på arbejdsmarkedet«, side 75

Grunduddannelser via folkeskoler/friskoler o. lign.

»Internationalt set klarer skolebørn i Danmark sig middelmådigt i tests af færdigheder, til trods for at vi er et af de lande, der bruger flest penge på folkeskolen«, side 10

»Derfor er det bekymrende, at internationale sammenligninger tegner et billede af en middelmådig dansk folkeskole til trods for, at Danmark er et af de lande, der bruger flest ressourcer på skoleområdet«, side 34

Ungdomsuddannelser (gymnasier o. lign./erhvervsskolerne)

»... et gennemgribende eftersyn af hele ungdomsuddannelsesområdet for at vurdere, hvordan det sikres, at uddannelserne i højere grad giver eleverne de rette kompetencer. Eftersynet anbefales at omfatte både de enkelte uddannelser og samspillet mellem dem«, side 105

»Der er behov for, at hele ungdomsuddannelsesområdet gentænkes«, side 106

Gymnasier⁷

»... kan det tyde på, at kravene for at bestå en studentereksamen er faldet.

Gymnasiereformen fra 2005 og den nye karakterskala har formentlig bidraget til den udvikling ...«, side 11

»Man vidste meget lidt om konsekvenserne af reformen (2005-gymnasiereformen), da den blev indført, og der foreligger endnu ikke nogen systematisk viden ...«, side 17

Kompetencer med høj erhvervsindkomst til følge: »Målt på det kriterium har gymnasiet været inde i en uheldig udvikling«, side 101

Taxameterfinansiering: »Det giver gymnasierne en økonomisk tilskyndelse til at sænke det faglige niveau ...«, side 101

Karaktergivningsbias: »... gives signifikant højere karakterer i B-niveau-fag som billedkunst, dramatik, idræt, mediefag, psykologi og religion«, side 105

Erhvervsuddannelser

»... de danske erhvervsuddannelser har Europarekord i frafald«, side 7

»... gennemsnitsalderen for en nyuddannet i Danmark (er) 28 år. Det er ti år mere end i Sverige«, side 37

»... gennemsnitsalderen på en elev, der fuldfører et uddannelsesforløb, er høj – ca. 28 år. At kalde erhvervsuddannelserne for ungdomsuddannelser er derfor i praksis blevet misvisende, da mange af eleverne er voksne«, side 95

Videregående uddannelser

»Flere og flere unge søger ind på de videregående uddannelser, der giver højest risiko for arbejdsløshed og lavest løn«, side 7

»... antallet af studiepladser og dimittender er vokset mest på uddannelser, der giver lav produktivitet målt på indkomst og beskæftigelsesgrad«, side 12

»Vi uddanner kort sagt forkert«, side 12

»Der undervises meget lidt på mange videregående uddannelser«, side 13

Kommunikation og kulturfag under humaniora: »I

1991 udgjorde de 10 pct. af det samlede humanistiske dimittender. I dag er tallet 60 pct.», side 51

Lærerruddannelsen er ikke et fuldtidsstudium: »Hvis studieintensiteten på lærerruddannelsen blev løftet, så den svarer til et fuldtidsstudium, hvor de studerende bruger 37,5 timer om ugen, vil det svare til to års ekstra uddannelse med samme studieintensitet som i 2010. Samtidig skal de 13 timers undervisning ses i lyset af, at der i 1986 var 25 ugentlige undervisningstimer på lærerseminarierne«, side 87

»Mange studier tilbyder få undervisningstimer, ofte under 10 timer om ugen. Det betyder, at studieintensiteten på mange uddannelser svarer til et deltidsjob«, side 117

Styringen

»Kombinationen af selveje og taxametersystemet er uheldig«, side 15

»Det vil kræve en radikal gentænkning af styrings- og finansieringsmodellerne for de videregående uddannelser«, side 15 og 118

For få og utilstrækkelige kompetencer: »Den korte forklaring på dét er, ..., at der i uddannelsessystemet er alt for svage tilskyndelser til at fokusere på kvalitet og på uddannelsernes anvendelighed på arbejdsmarkedet«, side 75

»Reformer af uddannelsessystemet har hidtil kun i begrænset omfang været baseret på solid evidensbaseret viden om, hvordan den bedste indlæring og den bedste opfyldelse af de øvrige mål med uddannelser sikres«, side 75

»Hovedparten af landets uddannelsesinstitutioner er selvejende og finansieret på en måde, der tilskynder dem til at optage flere elever og studerende og til at sikre hurtig gennemførelse af uddannelsen«, side 77

»Samlet tegner der sig et billede af, at det nuværende SU-system er en dyr og ineffektiv metode til at opfylde de målsætninger, der ligger til grund for systemet. Produktivitetskommissionen opfordrer derfor til, at man ser fordomsfrit på systemet i forbindelse med en reform, der sigter mod at styrke de studerendes tilskyndelser«, side 115

»Når uddannelsesinstitutionernes finansiering alene afhænger af antallet af eksamensbeviser, der uddelles, træder uddannelsernes kvalitet og relevans på ar-

bejdsmarkedet i baggrunden. Taxametersystemet giver dem ingen tilskyndelse til at begrænse optaget på uddannelser med høj dimittendledighed og lav gennemsnitsløn, tværtimod. Et lavere optag betyder færre ressourcer, og det kan føre til afskedigelser«, side 120

»Vil vi have uddannelser, der giver dimittenderne høj produktivitet på arbejdsmarkedet, er det nødvendigt at finansiere og styre uddannelsesinstitutionerne på en måde, så det bliver et succeskriterium. Som beskrevet i de forrige afsnit er det ikke tilfældet i dag«, side 123.

Hvis ovenstående barske kritikpunkter skal forsøges sammenfattet, kan det måske gøres således:

Taxametersystemet har korrumpet uddannelsesinstitutionerne. Bestyrelserne og/eller ledelserne af vore uddannelsesinstitutioner har ladet dette ske uden aktiv modstand. Jeg har f.eks. ikke hørt rektorer fra universiteter eller bestyrelser fra universiteter tage afstand fra taxametersystemet med de hertil hørende moral hazard-problemer eller korrupperende konsekvenser. Det har godt nok på det seneste været fremført fra universiteters ledelser, at de nu vil tage affære. Men kan der sættes lid til disse uforpligtende tilkendegivelser? Er realiteten ikke snarere som det udtrykkes i revyvisen ...« Det lover vi hinanden, det gør vi aldrig, aldrig mere? Ministre og ministerier med ansvar for uddannelsesområdet har vendt det blinde øje til, idet de gang på gang er blevet advaret om bl.a. de meget negative læringsmæssige konsekvenser af taxametersystemet. Blot nogle få eksempler herpå. I 2004 var der en omfattende debat om de negative konsekvenser af taxametersystemet med udgangspunkt i Politikens kronik »Disneyland er et dårligt forbillede« fra 9. juni 2004, der var skrevet af fem universitetsprofessorer. Kronikkens overskrift henviser til, at taxametersystemet giver samme incitamenter, som gælder for Disneyland. I Disneyland er et succeskriterium antallet af besøgende gennem tælleapparaterne, og med taxametersystemet blev antallet af studerende, der klarede eksaminer uanset eksamenskravene, en

økonomisk gulerod for universiteter. Pølsefabrik, faglig katastrofe fra skoler til universiteter var nogle af de overskrifter, som denne kronik, der inddrog uddannelsessystemet fra grundskolen til universitetsniveau, udløste.

I denne sammenhæng vil jeg også tillade mig at anføre, at jeg allerede i en Politiken-kronik, »Universiteternes forfald« fra 14. februar 1994 forudsagde, at taxametersystemet (dengang benævnt STÅ-systemet) ville føre til kvalitetsfald grundet taxametersystemets incitament. Universiteterne ville selvfølgelig forsøge at skjule eller benægte dette, men for økonomer er det standardviden, at dette vil ske. Skjult handling afledt af incitament har i den økonomiske litteratur fået en endog meget sigende betegnelse: moralsk hazard. Der skulle derfor på ingen måde profetiske evner til for at forudsige, at kvalitetsforringelser uundgåeligt ville blive en konsekvens af taxametersystemet. Men disse åbenbare konsekvenser blev alligevel benægtet af de ansvarlige.

To ufuldendte opgaver: produktivitet og uddannelse

Produktivitetskommissionen fik en meget omfattende opgave, der endvidere skulle løses på meget kort tid. Næppe nogen kommission ville have kunnet løse denne opgave tilfredsstillende inden for den udstukne tidsramme.

Men dette fritager ikke kommissionen for ansvar for det udførte arbejde, selv når denne præmis lægges til grund. I denne sammenhæng skal det endvidere tages i betragtning, at kommissionen havde mulighed for at indhente analyser fra andre. Det er nærliggende i relation hertil at inddrage, at der i kommissionens arbejde ikke er anført overvejelser over, hvordan kommissionen skulle løse den stillede opgave.

Noter

1. Jeg skal hermed udtrykke min store påskyn-delse og respekt for Danmarks Statistik og de medarbejdere, som jeg i relation hertil har udvekslet synspunkter med. Deres aktive medvirken har været en nødvendig forudsætning for, at jeg har kunnet uddybe mine vurderinger. Og denne medvirken er sket, selv om det har stået klart, at jeg ville sætte spørgsmålstejn ved den anvendelse af nationalregnskabsallene, som Produktivitetskommissionen gjorde brug af i relation til produktivitetsudviklingen i dagligvarebranchen. Min kritik har dog aldrig været rettet mod, at Danmarks Statistik bruger de konventioner i relation til nationalregnskabsopgørelse m.v., som Danmarks Statistik er forpligtet af, eller mod de begrænsninger, der er begrundet i et ufuldstændigt datagrundlag. Kritikken har udelukkende været vendt mod, at fagøkonomer ukritisk bruger tal, som de burde vide er ret så usikre, især når disse tal viser en meget mærkværdig udvikling.
Det anførte om talgrundlaget er alene udtryk for min vurdering og derfor ikke nødvendigvis for Danmarks Statistiks vurdering.
2. En skematisk oversigt herover fremgår af min hjemmeside www.en-vis-mand.dk under fanebladet Netop Nu.
3. På basis af en af andre defineret varekurv har en miniundersøgelse, som jeg selv har foretaget, indikeret, at prisniveauet er mindst 10 pct. lavere, jf. notatet *Prisundersøgelse* på min hjemmeside www.en-vis-mand.dk under fanebladet Netop Nu.
4. Anvendelsen af Jevons-indeks for primærgrupperne i forbrugerprisindekset gør, at dette ikke sker. Det skal understreges, at det er det implicitte prisindeks for produktionsværdien (brutto-avancen), der er anvendt i figur 1.
5. Der er et langt bedre grundlag for at beregne forbruget i produktionen, der udgør forskellen mellem produktionsværdien og værditilvæksten, hvorfor der her er set bort fra evt. fejlvisninger i opgørelsen af forbruget i produktionen i faste priser.
6. I denne forbindelse skal anføres, at STÅ-systemet, som taxametersystemet dengang blev benævnt, blev udbredt til universiteter m.v. fra 1993.
7. Side 77-78 er der en generel sønderlemmende kritik af gymnasiereformen fra 2005. Eksplo-

sion af i antallet af kursister på gymnasiale suppleringskurser efter 2005-gymnasireformen – fra omkring 2000 om året til over 10.000 om året – er også en af de stærkt negative konsekvenser af denne reform, se side 11-12.

Litteratur

Andersen, Ellen (1964a), »Måling af kapital«, *Nationaløkonomisk Tidsskrift*, 1964(102): 135-52.

Andersen, Ellen (1964b), »Mere om måling af kapital«, *Nationaløkonomisk Tidsskrift*, 1964(102): 235-48.

Danmarks Statistik (2002), *Nationalregnskab – Fastprisberegninger – Kilder og metode*, København.

Produktivitetskommissionen (2013a), »Danmarks produktivitet – hvor er problemerne?« *Analyserapport 1*, København, april.

Produktivitetskommissionen (2013b), »Konkurrence, internationalisering og regulering«, *Analyserapport 2*, København, maj.

Produktivitetskommissionen (2013c), »Styring, ledelse og motivation i den offentlige sektor«, *Analyserapport 3*, København, september.

Produktivitetskommissionen (2013d), »Uddannelse og innovation«, *Analyserapport 4*, København, december.

Produktivitetskommissionen (2014a), »Infrastruktur«, *Analyserapport 5*, København, januar.

Produktivitetskommissionen (2014b), »Offentlig-privat samarbejde«, *Analyserapport 6*, København, februar.

Produktivitetskommissionen (2014c), »Det handler om velstand og velfærd«, *Slutrapport*, København, marts.

Produktivitetskommissionen (2014d), *Svar på Christen Sørensen's kritik af produktivitetskommissionen*, Notat af 24. marts.

Sørensen, Christen (2002), »Pris- og mængdeindeks«, Økonomisk Samfundsbeskrivelse, Systime Academic.

Sørensen, Christen (2014), *Produktivitetskommissionens analyse af produktivitetsudviklingen i detailhandel*, marts (Notat udarbejdet til De Samvirkende Købmænd. Se hjemmesiden www.envis-mand.dk under fanebladet Netop Nu).

Anmeldelse

**Kasper Møller Hansen og
Karine Kosiara-Pedersen, red. (2014)**

Folketingsvalgkampen 2011 i perspektiv
København: Jurist- og Økonomforbundets
forlag, 436 sider

Danmark har været velsignet med systematiske studier af danske vælgere siden folketingsvalget i 1971. Valget blev starten på »Det Danske Valgprojekt«, som gennemfører repræsentative vælgerundersøgelser efter hvert valg. Hovedformålet ligger lige for: At undersøge årsagerne til, at vælgerne stemmer som de gør. Projektet har rødder og inspiration i *The People's Choice* (1944), som analyserede det amerikanske præsidentvalg i 1940. Tilbage i 1940'ernes USA havde interview-teknikker og metoder til stikprøveudvælgelse nået et højt og anvendeligt niveau, som er en forudsætning for systematiske studier af vælgeradfærd. Værket står selv i dag står som en milepæl i valgeradfærdsforskningen.

Redaktørerne Kasper Møller Hansen og Karina Kosiara-Pedersen skriver sig med bogen *Folketingsvalgkampen 2011 i perspektiv* ind i en stolt tradition i dansk politisk sociologi: Det Danske Valgprojekt. Og bogen er et ambitiøst og veludført supplement eller alternativ til hovedpublikationen *Krisevalg. Økonomien og Folketingavalget 2011* (2013). Med sit omfattende og mangedaceterede datamateriale sætter bogen lup på folketingsvalgkam-

pen 2011 ud fra både vælgernes, mediernes, partiernes og kandidaternes synsvinkel.

Bogens ambition er tydelig, interessant og hidtil uset. Bogen beskæftiger sig indgående med præcist de 20 dage, folketingsvalgkampen i august/september 2011 varede. Dermed er bogen den første af sin art herhjemme, som giver en sammehængende fremstilling af, hvordan en valgkamp til Folketinget konkret udspiller sig i Danmark. En så systematisk og omfattende kortlægning af en dansk valgkamp er ikke tidligere set samlet i et værk. Alene af den grund er bogen værd at læse for både politikere, forskere og alle med interesse for valg, vælgeradfærd og valgkampe.

Partierne gør brug af en bred vifte af kampagneinitiativer for at vinde vælgenes gunst. Bogen sætter fokus på disse initiativer og vurderer hvordan de virker. Et simpelt kig på partiernes kampagneudgifter vidner også om, at en systematisk kortlægning af valgkampen er interessant. I 2004 havde partierne kampagneudgifter svarende til omkring 34 mio. kr. I 2011 var dette tal mere end firedoblet til 137 mio. kr. Bogen formår desuden at perspektivere 2011-valgkampen til tidligere valgkampe med henblik på at identificere tendenser og udviklingstræk.

Bogens indhold kan opdeles i fire hovedgrupper. Første gruppe omhandler valgkampens

centrale aktører – herunder partierne, kandidaterne, partimedlemmer og interesseorganisationer. Bogen sætter fokus på partiernes pengeforbrug under valgkampe og på hvordan kandidaterne konkret fører valgkamp. Desuden undersøger bogen hvem (interesseorganisationer, erhvervsliv og andre aktører), der støtter de politiske partier og med hvor mange penge.

Anden gruppe sætter fokus på aktiviteter i valgkampen. Dette gælder bl.a. klassiske aktiviteter som valgplakater, dør til dør-kampagner og politikerdueller i TV. Denne del af bogen sætter også fokus på *negativ campaigning*.

Tredje gruppe af kapitler samles under overskiften *politik*. Her rettes fokus bl.a. på Ole Borres velkendte skel mellem fordelings- og værdipolitik og den betydning det har, når vælgerne afgiver deres stemmer. Denne del af bogen analyserer også både mediernes og politikernes dagsorden under valgkampen.

Den sidste gruppe af kapitler – og der hvor alle med interesse for valg og valgkampe kan uddrage flere dokumenterede essenser – handler om *effekter*. Altså: Hvad virker og hvad virker ikke når der skal fiskes stemmer? Her fokuserer forfatterne bl.a. på effekten af kandidaternes kampagner, om valgkampe gør vælgerne klogere på politik samt sætter luppen på den evigt interessante gruppe af *tvivlere*.

En af bogens underfundige pointer handler om den værktøjskasse kandidaterne gør brug af under en valgkamp. På mange måder er den lige så klassisk og velkendt som den altid har været med valgplakater, annoncer i lokalaviser, gadearrangementer og omdeling af materialer. Her er alt ved det gamle. Men med til den historie hører også, at det er blevet langt mere krævende at være folketingskandidat. For værktøjsskassen har vokseværk. Den generelle tendens til professionalisering

af valgkampe, de sociale mediers frembrusen og et umætteligt hav af stadigt flere medier hæver omkostningerne for alle kandidater.

Bogens centrale akademiske bidrag centrerer om seks hypoteser, som bogen indledningsvist rejser og finder opbakning til i de løbende analyser. Og skulle nogen være i tvivl: Valgkampe gør en forskel og valgkampe flytter stemmer. Opskrifterne på hvordan findes overalt i bogen. Bogen viser blandet andet, at valgkampe er *empowerment* af vælgerne. De danner, oplyser og engagerer vælgerne med medierne som det centrale bindeled. Dermed spiller valgkampe en central demokratisk rolle. I analyserne vises det desuden, at danske vælgere er partipolitisk forudindtagede og meget selektive i hvilken eksponering de lader sig udsætte for. Forskerne finders således støtte i *minimumseffekt-tesen*, som handler om, at valgkampes største effekt reelt er at fastholde vælgere i det partivalg de allerede i modsætning til at flytte stemmer rundt i det politiske landskab. Og det på trods af, at ikke mindre end 44 pct. af vælgerne ved valgkampen i 2011 først bestemte sig under valgkampen.

Sammenfattende er bogen lækkerbiskken af tal og tabeller om valg og vælgere, for alle, der holder af valg og valgkampe. Bogens styrker er mange. Men især det klare fokus på selve valgkampen i kombination med det mest omfattende og bredspektrede datamateriale for nogen dansk valgbog gør analyserne interessante. Bogens eneste åbenlyse mangel er fokus på sociale medier. Hvilke politikere bruger dem? I hvilket omfang? Og med hvilken succes og effekt? Der er et stort behov for valgforskning i sociale mediers betydning for såvel dagsordensættelse som vælgernes stemmeadfærd. Lad det være en opfordring herfra til næste store projekt.

Kenneth Thue Nielsen
Kommunikationschef, cand.scient.pol.
LO • Landsorganisationen i Danmark

Abstracts

Observations on Using Scientific Evidence in Practice and Policy-making

Thomas A. Schwandt, professor,
University of Illinois, Urbana-Champaign,
USA

Given that the use of scientific evidence in practical and political decision making cannot basically be understood as an exercise in technical problem-solving, it seems reasonable to examine what the use of evidence actually entails in real-life situations. This paper explains six points: Evidence is not the same as data. Evidence is neither a »base« nor a »foundation« for decision-making. Scientific evidence is only one of several kinds of knowledge necessary for policy-making. Use is not the same as a simple transfer of knowledge from producers to consumers and its subsequent immediate and direct application. Use is a collective, interactive phenomenon. To understand use, we need to pay more attention to situated cognition, cognitive processes, and argumentation.

Evaluation as political insurance

Tobias Høygaard Lindeberg, ph.d.,
prodekan ved det Samfundsvidenskabelige
og Pædagogiske Fakultet,
Professionshøjskolen Metropol,
thli@phmetropol.dk

This article examines whether insurance can serve as a metaphor to comprehend evaluation and how it may influence the understanding of evaluation in society. The perspective adds a rationalistic counterpoint to leading studies that perceive evaluation as a ritualised practice. The article argues that evaluation can be understood as insurance in the sense that evaluation distributes the risk of decision-making parallel to the distribution of economic risk of insurance. Additionally, the article identifies central differences between insurance and evaluation, especially in regard to the two different logics at stake: market logic and hierarchy logic. The differences imply that the costs of evaluation – in contrast to costs of insurance – are not primarily born by those who benefit from the evaluation. Moreover, while risk sharing is typically desirable in an economic context, it is more ambiguous in a political context. Evaluation is often employed to enhance accountability, yet, when evaluation is analysed

as insurance, it becomes clear that evaluation also diffuse accountability.

Eight attention points when evaluating political reforms

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

Karen N. Breidahl, Institut for Statskundskab, Aalborg Universitet, knb@dps.aau.dk

Jan-Eric Furubo, Riksrevisionen i Sverige

Anne Halvorsen, institutleder, Institutt for sosiologi og sosialt arbeid, Agder Universitet, anne.halvorsen@uia.no

This article analyses the challenges related to evaluations of major policy reforms. It is based on a meta-evaluation of the evaluation of the NAV reform in Norway, who was both a significant reorganization of the central, regional and local government and a social policy reform. Meta Evaluations assess the usefulness of one or more evaluations and should not be confused with meta-analyses. The purpose of this meta-evaluation was to identify general principles for organizing the evaluations of large political-administrative reforms.

Based on the analysis eight crucial points of attention when evaluating large scale political-administrative reforms are elaborated. We discuss their reasons and argue that Denmark and other countries can learn from the experience of Norway.

Evaluation of large-scale national political-administrative reforms: A comparative analysis

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

Gunnar Gjelstrup, lektor, Institut for Statskundskab, Københavns Universitet, gg@ifs.ku.dk

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

Hanne Foss Hansen, professor, Institut for Statskundskab, Københavns Universitet, hfh@ifs.ku.dk

Large-scale national political-administrative reforms are a rare evaluand in the evaluation literature for at least two reasons. First such interventions are relatively less frequent than the countless changes happening in for instance the health care sector and the education systems. Second although their frequency seems to have increased in recent decades they are often not evaluated. This article provides a systematic comparative analysis of recent evaluations of two large-scale national political-administrative reforms. Two of the largest political-administrative reforms in the Nordic countries in the 2000s were the reform of labor and welfare administration in Norway (NAV reform) and the local government reform in Denmark. The two reforms have a number of characteristics in common – also with other major political and administrative reforms. They were adopted centrally by the Parliament (Folketing/Storting), they affected large parts of the public sector and the population, they had several and partly incompatible objectives and they developed and changed over time. Furthermore the two reforms were characterized by a focus on formal structural change as the main policy

instrument and included both the state, regional and local level of public administration. Finally and crucial in this context, both reforms have been evaluated. The systematic comparative analysis of the evaluations of the two reforms use four theoretical lenses: A rational instrumental perspective, a political interest-based perspective, an institutional cultural perspective and a chaos perspective. The analysis shows that despite the reforms similarities their evaluations were organized in very different ways and had very different impact. While the evaluation of the NAV reform in Norway was organized as a research-based external evaluation, the evaluation of Danish municipal reform was organized as an internal evaluation where key stakeholders got evaluation responsibilities. And while the evaluation of the NAV reform was used mainly for knowledge-building and creating debate, the evaluation of the local government reform was used instrumentally to adjust the division of labor between levels of government in selected areas of activity.

Performance management in Danish municipalities after the crisis

Eva Moll Sørensen, postdoc ved Institut for Statskundskab, Københavns Universitet, ems@ifs.ku.dk

Søren Kjær Foged, ph.d. ved Institut for Statskundskab, Københavns Universitet, skf@ifs.ku.dk

This article presents the results of six qualitative case studies of performance management in Danish municipalities after the economic crisis in 2009. It is argued that the municipalities have adapted their performance management practices to the new conditions of austerity, and that they aim for more strategic use of performance goals and measurements as a tool for prioritization of programs and

expenses. This development is, for example, expressed through the arrangement of political goals and targets into hierarchies and through the use of performance targets in the budgeting process. At the same time, the municipalities have deemphasized the development of common performance management systems for decentralized service providers.

A performance evaluation of sanctions on municipal accounts in Denmark, 2010-2013

Søren Kjær Foged, ph.d.-studerende ved Institut for Statskundskab, Københavns Universitet, skf@ifs.ku.dk

This article evaluates the low-based sanctions on the municipal accounts in Denmark introduced by the Danish government in 2011. The evaluation examines the effect of the sanctions relative to some selected evaluation criteria that are used to form hypotheses. The hypotheses concern the municipalities' ability to stay within the budgeted expenditures, the precision of the budget relative to the actual spending and finally whether the effect of the sanctions differ according to the municipalities' financial situation and the policy sector in question. The hypotheses are tested by the use of multiple regression on data from 2010 to 2013. The evaluation finds that the municipalities have indeed improved their ability to keep within their budget, but that the municipalities have not improved the spending precision relative to the budget. Furthermore, affluent municipalities are affected more by the sanctions than entities under higher fiscal stress, while schools, day-care and the specialized social area are the mostly affected policy sectors.

Do performance indicators lead to clarity for employees?

Peter Dahler-Larsen, professor,
Institut for Statskundskab,
Københavns Universitet, pdl@ifs.ku.dk

One of the main purposes of performance indicators is to communicate clearly to employees what needs to be done in the organization. However, significant empirical variation

is found in the extent to which two comparable groups of public employees (upper secondary level school teachers and language teachers) find that performance indicators provide clarity. Qualitative data that describe the characteristics of their indicator systems and the organizational contexts in which they operate help us understand the sources of clarity/uncertainty. Finally, the pros and cons of clarity are discussed in a public management perspective and in a democratic perspective.