

Konsekvenser af indikatorer

Professor Peter Dahler-Larsen
Institut for Statskundskab
Syddansk Universitet
Campusvej 55
5260 Odense S



KREVI.

FORORD

Brugen af indikatorer som evalueringsform vinder indpas inden for den offentlige sektor i Danmark. Fænomenet er kendt i udlandet, og i Danmark er evalueringsformen især brugt på sygehusområdet, hvor sygehusenes præstationer i en årrække er blevet målt ved hjælp af en række sammenlignelige indikatorer. Resultaterne bliver offentliggjort på www.sundhedskvalitet.dk. Med de 23 ældreindikatorer er det kommunale område nu også berørt, og der er mere på vej. I bl.a. kvalitetsreformen, økonomiaftalerne mellem regeringen og KL og senest i Finansloven 2008 lægges der op til øget brug af nationale indikatorer. De bløde velfærdsområder, såsom udsatte børn og unge, ældre- og dagtilbudsområdet, skal også evalueres ved hjælp af indikatorer, der leverer "sammenlignelige oplysninger om den faglige kvalitet" som det hedder i Finansloven 2008.

I den internationale forsknings- og evalueringslitteratur findes en del undersøgelser og diskussioner af indikatorer og deres konsekvenser. Men der er en stor mangel på danske undersøgelser og i tilknytning her til også mangel på viden om indikatorers konsekvenser og anvendelighed i en dansk sammenhæng.

På den baggrund har KREVI etableret projekter, som på forskellig måde har til hensigt at kaste lys over brugen af indikatorer i evalueringssammenhæng. I et projekt undersøges positive og negative konsekvenser af nationale præstationsmålinger i en dansk sammenhæng med fokus på sygehusområdet. I et andet undersøges, hvordan brugen af indikatorer evt. kan understøttes af andre tiltag og dermed få en højere kvalitet.

Som led i disse projekter indgik KREVI et samarbejde med professor Peter Dahler-Larsen, der både nationalt og internationalt regnes for førende inden for evalueringsområdet. Peter Dahler-Larsens rolle i samarbejdet har både været at give sparring i forhold til de nævnte projekter, men også at bidrage med en mere teoretisk, litteraturbaseret viden om indikatorer og deres konsekvenser. Det er resultatet af sidstnævnte, som fremlægges i dette notat.

I notatet diskuterer Peter Dahler-Larsen indikatorbegrebet og indsætter det i en større evalueringsteoretisk sammenhæng. Det indebærer bl.a. en grundig gennemgang af en række anvendelsesformer og en diskussion af det Peter Dahler-Larsen kalder for "konstitutive virkninger". Notatet afsluttes med nogle forslag til, hvordan indikatorers negative virkninger kan "repareres" samt nogle overvejelser om, hvordan indikatorers virkninger kan studeres.

Peter Dahler-Larsens notat rejser en række spørgsmål til den "indikatorbølge", som i disse år skyller ind over offentlige organisationer i Danmark. Et af dem er, om indikatorerne overhovedet opfylder de tiltænkte mål. Et andet er, om det kan betale sig at udvikle tunge indikatorsystemer, eller om ressourcerne med fordel kunne bruges anderledes.

Dette er væsentlige spørgsmål, som kræver en grundig debat. Det er KREVIs forhåbning, at dette notat, og det øvrige arbejde med indikatorer, som foregår i KREVI, kan stimulere en sådan debat.

Tim Jeppesen, Direktør
KREVI, maj 2008

INDHOLDSFORTEGNELSE

1 FORMÅL OG AFGRÆNSNING	4
1.1 FORMÅLET MED DETTE NOTAT	4
1.2 HVAD ER EN INDIKATOR?	4
1.3 KRÆVER EN INDIKATOR EN STANDARD, ET INDIKATORSYSTEM OG NOGLE SANKTIONER?	5
1.4 HVAD MENER VI MED "KONSEKVENSER"?	5
1.5 ER DET SANDSYNLIGT AT MAN KAN IDENTIFICERE ENTYDIGE "EFFEKTER, DER HOLDER" AF INDIKATORER?	6
2 BEGREBER OM KONSEKVENSER AF INDIKATORER	7
2.1 ER VI PÅ BAR BUND BEGREBSLIGT, NÅR VI SKAL STUDERE KONSEKVENSER AF INDIKATORER?	7
2.2 HVILKE ANVENDELSESFORMER FOR EVALUERING FINDES DER?	7
2.3 HVILKE KONTEKSTUELLE FORHOLD KAN TÆNKES AT SPILLE IND PÅ KONSTITUTIVE VIRKNINGER?	23
2.4 HVORDAN BEGREBSLIGGØRES OG VURDERES "KONSTITUTIVE VIRKNINGER"?	24
2.5 HVAD BETYDER DET, AT INDIKATORER INDGÅR I MANagementsYSTEMER TIL EVALUERING/DOKUMENTATION OSV.?	25
3 RESSOURCETRÆK TIL INDIKATORER	27
3.1 HVOR BLIVER RESSOURCER TIL EVALUERING/DOKUMENTATION/KVALITETSINDIKATORER AF?	27
4 FORSLAG OM REPARATIONER AF INDIKATORERS NEGATIVE VIRKNINGER	29
4.1 HVILKE FORSLAG HAR VÆRET NÆVNT TIL MODVIRKNING AF INDIKATORERS UØNSKEDE VIRKNINGER?	29
5 AT STUDERE KONSEKVENSER AF INDIKATORER	30
5.1 HVORDAN KAN MAN STUDERE KONSEKVENSER AF INDIKATORER?	30
6 LITTERATUR	32



KREVI.

Olof Palmes Allé 13
8200 Århus N
www.krevi.dk
post@krevi.dk

T: 7226 9970
F: 7226 9980

ISBN 978-87-92258-08-3 (elektronisk version)

Grafisk design: Designit

1 FORMÅL OG AFGRÆNSNING

1.1 Formålet med dette notat

Som indledende fase i et samarbejde mellem KREVI og Peter Dahler-Larsen:

- at opstille nogle relevante spørgsmål om emnet "konsekvenser af indikatorer" og
- at give nogle svar på, hvordan det begrebslige og metodiske landskab på dette område kan skitseres og opdeles.

1.2 Hvad er en indikator?

Inden for dette notats faglige område gælder det samme som for mange samfundsvidenskabelige områder, nemlig at klare teoretiske definitioner ikke altid passer på folkets hverdagsbetegnelser. Begreber om indikatorer, evaluering og kvalitet bruges både i teoretisk og i praktisk sammenhæng, men der er ikke et gudgivet katalog over et sæt entydige begreber.

Vi vil her som arbejdsdefinition på en indikator hævde: *En indikator er en kvantitativ størrelse, der beskriver indsats, præstationer og/eller udfald af offentlig virksomhed, og implicit eller eksplicit og med en vis stabilitet benyttes i vurdering heraf med tanke på, at denne vurdering skal benyttes i praktiske handlingssituationer, herunder f.eks. politiske, ledelsesmæssige eller fagprofessionelle beslutninger.*

Et tal er ikke pr definition en indikator. Det er tallets brug/tolkning/indfletning i en bestemt sammenhæng, som beskrivelse af noget, til vurdering af noget, med henblik på brug til noget, der gør det til en indikator. En indikator er med andre ord et kvantitativt evalueringskriterium.

Det bemærkes, at "indikator" således indgår som et led i en temmelig generel betragtning om evaluering. Det er med andre ord en evalueringsteoretisk ramme, emnet her indsættes i. Man kunne have gået efter mere specifikke definitioner af f.eks. indikatorer, benchmarking, standarder, kvalitetsmåling, osv. Men det vil næppe være frugtbart at antage, at der må en særskilt teori til for hver af disse. Det vil være urimeligt at antage, at konsekvenser af indikatorer kræver en helt særlig teori til forskel fra f.eks. teorien om konsekvenser af kvalitetsmåling.

En konsekvens af ovenstående definition er, at mennesker i praksis kan benytte indikatorer, uden at de kalder det indikatorer. Mange menneskers omtale af "nøgletal" synes at svare til det, der her kaldes indikatorer. Det er dog ikke altid eksplicit heller, hvad der nøjagtigt skal til, for at et tal er et nøgletal. For en række betegnelser så som performance indicators, performance measurement, præstationsmålinger og monitorering gælder på samme måde, at ingen af de nævnte betegnelser vurderes at have tilstrækkelig præcision og teoretisk styrke til at kunne bruges definatorisk i denne sammenhæng. Desuden er det fra KREVI's side ønsket, at notatet skal fokusere på indikatorer. Men de betragtninger, der i dette notat gøres gældende om indikatorer, vil i vidt omfang også gælde for nøgletal, performance indicators mm., så længe de falder inden for ovenstående definition.

1.3 Kræver en indikator en standard, et indikatorsystem og nogle sanktioner?

Som indikatorbegrebet er defineret ovenstående, er der en vis stabilitet eller jævnlighed ved indikatoren. Med andre ord er det, der interesserer os her, at indikatoren optræder mere end en enkelt tilfældig gang.

Det antages ofte om indikatorer, at den normative tolkning af indikatoren er åbenbar, således f.eks. at "0" er dårligt og "100" er godt. Men det indgår ikke i definitionen.

Til en indikator kan være knyttet en eksplicit standard, altså en tærskelværdi for, hvornår scoren er god eller god nok. Tærskelværdien kan være normativt velbegrundet eller mere eller mindre pragmatisk fastsat (f.eks. ud fra landsgennemsnittet). Som standard kan man bruge sidste års tal, et landsgennemsnit, en statistisk beregnet størrelse efter kontrol for en række omgivelsesbetingelser, eller et vilkårligt tal.

Den amerikanske evalueringsforsker Robert Stake (2004) siger, "there is no good way of fixing standards." Det er vel mere eller mindre rigtigt på forskellige områder. Hvad angår trafikdræbte er de fleste enige om, at nul er et rigtig godt tal. I hvert fald for "godt". Men "godt nok" er måske et andet tal. Vanskeligere er det på områder, hvor en bestemt praksis er et resultat af nøje overvejelser og ofte modstridende hensyn, f.eks. tvangsfjernelser, kejsersnit og bæltefikseringer. I praksis bliver et gennemsnit, f.eks. et landsgennemsnit, den mere eller mindre implicitte standard. Det er problematisk på flere måder. En af dem er, at et stigende antal kejsersnit ikke nødvendigvis er et gode, blot fordi det sker i hele landet samtidig. (Munk 2007).

Vi vil i det følgende operere med, at der til en indikator kan knyttes, men ikke nødvendigvis knyttes, en eller flere implicitte eller eksplicitte standarder. Standardens tydeliggørelse kan veksle over tid, også efter en given indikator er etableret. For eksempel kan man som konstruktør af et indikatorsystem undlade at sætte standarder, og standarder kan så påføres af journalister eller politikere senere. Man kan også vedtage officielle standarder, som så glemmes eller ignoreres, når tallene når frem.

Indikatorer indgår typisk i forskellige organisatoriske og styringsmæssige systemer og koncepter, så som "benchmarking", "kvalitetsmåling", "kvalitetssikring", "monitorering" osv. Hvilke systemer det er, og hvordan de virker, kan være afgørende for konsekvenserne af indikatorer, men det indgår ikke i sig selv i definitionen på en indikator.

Ligeledes kan der til en given indikator knyttes forskellige sanktioner. For eksempel kan en resultatkontrakt, hvori indgår indikatorer, være hård eller blød, det vil sige indikatorerne kan være mere eller mindre eksakt og formuleret, og konsekvenserne af forskellige scorere kan være hårdt eller blødt defineret. Igen: det kan ske, men ingen eksakt bestemmelse heraf indgår i selve definitionen på en indikator.

Men afklaringen er vigtig, for mange mennesker underforstår bestemte standarder, systemer og sanktioner, når de taler om indikatorer. Og vi vil senere komme tilbage til, at en indikators konsekvenser selvfølgelig afhænger stærkt af de organisatoriske og politiske omgivelsesbetingelser, herunder f.eks. institutionaliserede standarder, systemer og sanktioner.

1.4 Hvad mener vi med "konsekvenser"?

Det er et bredt arbejdsbegreb, der har til sigte at fange, at "noget sker." Mange forskellige ting kan ske. (Det er måske en konsekvens af Villy Søvndals fremlæggelse af GPS-data fra hans kampagnebil, at mange overbevises om hans flid og vilje til at tale med mange men-

nesker i hele landet. Men det er også en konsekvens af samme tal, at studerende kan fremlægge beregninger, der viser, at han visse steder må have kørt 150 kilometer i timen)].

Det er først, når man beslutter sig for *bestemte* begrebslige perspektiver og empiriske metoder, at man kan indfange *bestemte* konsekvenser.

Med "konsekvenser af" mener vi noget kausalt. At noget sker som følge af noget andet. Kausale spørgsmål kan i almenhed angribes på to måder, variansbaseret og procesbaseret. I en variansbaseret strategi søger man at isolere årsagen så meget som muligt og derefter finde dens effekt ved at sammenligne variationer i årsager og resultater på tværs af kontekster, hvor alle andre variable så kontrolleres mere eller mindre stramt. Mange mener, at det mest stringente design hertil er det randomiserede eksperiment, og at dette design er i stand til at bestemme effekter entydigt. Andre benytter en procesbaseret analyse, hvor en given indsats ikke udfolder sig isoleret, men altid i samspil med forskellige kontekster. Effekter er resultat af indsats + kontekst (Pawson og Tilley 1997).

Når vi i det daglige taler om "konsekvenser af" X, underforstår vi ofte et helt sæt af kontekstuelle betingelser, som vi ikke problematiserer, mens vi udpeger X som årsag til disse konsekvenser. Vi burde måske mere præcist tale om, at X har de nævnte konsekvenser, når bestemte kontekstuelle betingelser samtidig er til stede eller medvirker.

1.5 Er det sandsynligt at man kan identificere entydige "effekter, der holder" af indikatorer?

Nej, det er ikke sandsynligt. "Indikatorer" er – ligesom mange andre aktuelle koncepter for styring og ledelse ikke ret meget i sig selv ud over nogle løse ideer. Disse ideer i sig selv er ikke konstante og distinkte og unikke. De fører først til praktiske konsekvenser, når noget konkret skal måles, og når der bliver investeret i dem i form af politik, ledelse, organisation, mening, magt, modstand osv. Disse "investeringer" er ganske forskellige fra kontekst til kontekst, ligesom investeringen i "den samme" indikator kan veksle meget over tid. Det er faktisk endda typisk, at indikator-systemer har en ganske kraftig udviklingsdynamik. Positionen i dette papir er derfor, at det vil være lidet frugtbart at satse på at udlede universelle årsags-virkningssammenhænge angående indikatorers konsekvenser. Men gennem begrebsdannelse, ideer og undersøgelser kan vi blive gradvis klogere på nogle sandsynlige konsekvenser, alt efter konteksten.

2 BEGREBER OM KONSEKVENSER AF INDIKATORER

2.1 Er vi på bar bund begrebsligt, når vi skal studere konsekvenser af indikatorer?

Vores forestillinger om konsekvenser af indikatorer hænger sammen med vore forestillinger om, hvilke problemer indikatorer skal være med til at løse. Antager vi, at der findes bestemte kvalitetsproblemer, f.eks. for dårlige præstationer i enkelte institutioner, kan det eventuelt hjælpe på *dette problem* at definere indikatorer med tilknyttede standarder. Andre kvalitetsproblemer – f.eks. manglende innovation blandt samtlige institutioner – vil ikke blive afhjulpet, men måske endda blive værre. Med andre ord er vore antagelser om problemer i virkeligheden mere eller mindre tavst med i billedet, når vi skal begrunde indførelsen af indikatorer og når vi udnævner bestemte ændringer til at være indikatorernes ”konsekvenser.” Det betyder imidlertid ikke, at disse ændringer er ren fiktion.

Der findes ingen specifik teori om konsekvenser af indikatorer. Det betyder ikke, at vi er på helt bar bund. Vi kan tage udgangspunkt i anvendelsesdebatten inden for evaluering, hvor der gennem en længere årrække er etableret en typologi med syv typer af evalueringsanvendelse. Jeg vil derfor i det følgende lade indikatorer og evaluering være let udskiftelige begreber og bruge disse syv typer som ramme for en nogenlunde systematisk gennemgang af, hvilke konsekvenser af indikatorer, man kan forestille sig.

2.2 Hvilke anvendelsesformer for evaluering findes der?

Kontrol
Læring
Oplysning
Strategisk
Taktisk
Symbolsk
Konstitutiv

Per definition udarbejdes indikatorer ligesom evaluering generelt for at spille en rolle i praktiske handlingssituationer. Evalueringsfeltet har løbende været optaget af dette hensyn, som måske har fungeret som den vigtigste drivkraft bag udvikling af nye tilgange og modeller (Hellstern, 1986). Samtidig har tilhængere og udøvere af evaluering haft en tilbøjelighed til at fokusere på ganske bestemte udvalgte former for anvendelse af evaluering, især de positive og tilsigtede – altså hvad man kan kalde de uantastelige måder at anvende evaluering på. Til gengæld har man måske haft en udpræget blind plet (Mark & Henry, 2004) i forhold til anvendelsesformer, som er mere problematiske og diskutabile. Det betyder ikke, at man inden for evalueringsskredse ikke har diskuteret forekomsten af politisk misbrug af evalueringer. Det har man gjort (Christie & Alkin, 1999). Men det har vedvarende været ud fra en ide om, at misbrug var let at kende, når man mødte den, - det har været ud fra en klar sondring mellem det antastelige og det uantastelige, mellem legitimt og illegitimt. Man har ikke altid taget højde for, at den ene mands brug kan være den andens misbrug. Det samme gør sig gældende hvad angår debatten om indikatorer (f.eks. i form af performance indicators). (Perrin 1998).

Nu drøftes syv begreber om evalueringers anvendelsesformer, som er udviklet inden for forskningen i evaluering. Hver af dem giver et bud på, hvori konsekvenserne af evaluering

består. Især de første par stykker af de syv anvendelsesformer svarer i øvrigt til, hvad mange i almindelighed mener, er evalueringsfunktion helt generelt.

De syv anvendelsesformer er analytiske begreber. De er betragtningsmåder lige så meget, som de er virkelighedsudsnit. Hver anvendelsesform indebærer en måde at betragte virkeligheden på, som har sine forudsætninger, og som er selektiv. Samtidig med, at anvendelsesformerne hver især kaster lys over aspekter af virkeligheden, efterlader de andre dele uden for opmærksomhedsfeltet. Der er ingen garanti for, at de tilsammen udtømmer ethvert tænkeligt udfald af indikatorer.

Det er kun som begreber, at de kan adskilles. I virkelighedens verden er det meget sandsynligt, at flere af disse anvendelsesformer forekommer samtidig og/eller er filtret sammen. Måske er det umuligt at garantere, at evaluering får et eneste ganske bestemt udfald. Måske konvergerer konsekvenserne af en evaluering ikke i et enkelt endeligt udfald. Måske er konsekvenserne af evaluering mangfoldige og ufærdige. Måske er det umuligt at love, at hvis blot evaluering, in casu indikatorer tilrettelægges på en bestemt måde, så vil deres udfald være definitive, forudsigelige og gode. Vi vil nærme os denne problematik ved at se på anvendelsesformerne en for en. De mest komplicerede kommer til sidst.

Evaluering som kontrol

Den første anvendelsesform er kontrol. Det er legitimt i den demokratiske, parlamentariske styringskæde, at der udøves en eller anden form for kontrol [Vedung 1998]. Evert Vedung (1998) peger således på, at evaluering i forhold til det repræsentative demokratis styringskæde kan tjene hele fire legitime kontrolfunktioner: Borgerne kan holde øje med politikerne; politikerne kan holde øje med embedsmændene; embedsmændene kan holde øje med fagfolkene, der udfører en bestemt indsats. Og endelig kan brugere og borgere holde direkte øje med de fagfolk, som de modtager ydelser fra. I den engelsksprogede litteratur opereres ofte med et begreb om "accountability", der har at gøre med evnen til og krav om at afrapportere, hvis man udfører en opgave betroet af en opdragsgiver. Tankegangen er kontraktlig.

Når der udøves kontrol, er det typisk det centrale spørgsmål, om de underordnede dele af systemet har leveret ydelser, der lever op til de formelt fastsatte målsætninger i systemet som helhed. Den typiske beslutningssituation er relativt endimensionel: Skal indsatsen fortsætte eller nedlægges, skal den have flere penge eller færre penge, og skal der i øvrigt uddeles positive eller negative sanktioner. Kontrolanvendelsen kan kendes på, at der sendes en positiv eller negativ sanktion tilbage fra en overordnet instans i styrekæden til en underordnet instans. Den positive sanktion kan bestå i, at der ikke kommer nogen negativ. Det er ikke en relevant kritik mod kontrol-anvendelsen, at den ikke giver læring og udvikling. Meningen med evaluering til kontrol er at gøre det muligt for kontrollanten at udøve kontrollen.

I et kontrolsigte ønsker man typisk at beskrive relativt komplicerede aktiviteter og udfald heraf ved hjælp af relativt simple kriterier. Dette er forudsætningen for, at der er god styringsøkonomi i et givet kontroltiltag, og at kontrollen netop understøtter en beslutning om en sanktion, hverken mere eller mindre. At foreslå forbedringer er ikke nødvendigvis på dagsordenen.

At beskrive noget komplekst ved hjælp af noget komplekst er sjældent ressourcemæssigt fornuftigt. I et kontrolperspektiv kan der således være organisatoriske grunde til at vælge relativt simple evalueringskriterier, også selv om andre kan mere, at kriterierne ikke er fuldstændigt dækkende og valide udtryk for kompleksiteten i de processer og ydelser, der evalueres. Her kan indikatorer have fordele frem for mere vanskeligt håndterbare evalueringskriterier. Operationelle indikatorer står her for god styringsøkonomi. Hvordan kompleksitetsreduktionen sker, kan være af underordnet betydning set fra de kontrollerendes perspektiv. Kontrollen er også på plads, selv om kontrolkriterierne er indholdsmæssigt diskutabile.

Teoretisk kan dette forklares, for eksempel i lyset af moderne udgaver af systemteori a la Luhmann. De "regulative instanser" vil ofte stille sig tilfreds med, at reguleringerne gør en eller anden forskel, uden at de bekymrer sig om, hvilke forskelle andre steder i systemet, som denne forskel gør (citeret hos Andersen 2004).

Derfor kan man af og til opleve, at kontrol bygger på temmelig besynderlige kriterier. De handler f.eks. om processer, selv om man ideologisk taler for resultatstyring (Dahler-Larsen 2001). De kan også have relativt lav validitet, og alligevel bruges til kontrol.

Ikke overraskende har kontrolevalueringer et ret dårligt image hos de kontrollerede. Det er ikke rart at få en kompleks indsats belyst ved hjælp af simple kriterier, især når det kan føre til sanktioner.

Mange frygter også, at kontrolkriterier skævvrider billedet af det, der kontrolleres. For eksempel vil et stærk kontrolorienteret fokus på indikatorer for skolernes målopfyldelse indebære en opprioritering af eksogene mål for undervisningen på bekostning af endogene mål i undervisningen.

Selv om der kommer indvendinger af denne type, fungerer evalueringer til kontrol alligevel i det omfang indikatorerne indgår i udøvelsen af opsynshavers kontrolfunktion. Legitimiteten kan ligge i den funktion, som evaluering har i forhold til opsynet og/eller i hele forvaltnings-systemet, ikke i forhold til indikatorernes betydning for producenter af en ydelse.

Nogle af de aktuelle udfordringer angående kontrol handler om "intelligent accountability" (O'Neill 2006), altså en form for kontrol, der er intelligent, ikke-indgribende, letfodet, omkostningslet, relevant og som ikke blander sig for meget i det daglige arbejde, således at et skel mellem daglig ledelse og udøvelse af kontrol kan opretholdes.

Man kan således godt gå ind for en mere intelligent udøvelse af kontrol, mens man er imod eksisterende kontrolsystemers udformning. Mange hævder således, at de nuværende autoritativt stillede krav til kontrol af offentlige ydelser gør det svært at anerkende autoritetsudøvelsen. Atter andre er helt lodret imod kontrol som sådan, men det er en vanskelig position at indtage i et demokratisk perspektiv.

Evaluering som læring

Begrebet læring bruges i denne sammenhæng udskifteligt med udvikling. I begge tilfælde henvises til en forbedring af indsatsen, eventuelt af indsatsens organisering eller rammebetingelser. Sigtet med evaluering til læring er ikke at holde opsyn og uddele sanktioner, men at forbedre indsatsen på et område. Der er derfor typisk tale om andre metoder og data end kontrolevalueringernes. Der er ikke så stor interesse for målinger i forhold til standardiserede kriterier, men mere for procesbeskrivelser og kontekstnære metoder. Indikatorer kan dog indgå som anledninger til at kaste opmærksomhed på, hvorfor man scorer, som man gør.

I et læringsperspektiv er der imidlertid typisk tale om en helt anden mentalitet hos de involverede parter. Hvor kontrolevalueringerne typisk leder til, at man viser sine bedste sider og skjuler sine dårlige, så godt man kan, forudsætter læringsevalueringer en åben og tillidsfuld fremlægning især af ens egne svageste sider, for det er her, at læringspotentialet er størst. Ikke mindst derfor opererer læringsorienteret evalueringsarbejde ofte med deltagerorienterede og dialogprægede metoder. Indikatorer må udvikles, tolkes, bearbejdes og forstås. Det gøres også for at øge sandsynligheden for, at evalueringsresultaterne opleves som brugbare for de personer, der er ansvarlige for at udvikle indsatsen. Ejerskab, involvering, refleksion og relevans er ofte nøgleord i disse processer.

Med andre ord flyder læringen ikke automatisk ud af indikatorerne selv. Der må knyttes nogle organisatoriske processer til, før indikatorer kan anvendes læringsorienteret.

Læring er næsten per definition altid til det bedre. Læring repræsenterer et trin opad i et indsigts- eller kompetencehierarki. Man bliver gennem læring altid bedre til at løse bestemte opgaver eller klare "fremtidens krav og udfordringer" (som den, der bruger læringsbegrebet åbenbart er bekendt med). (En undtagelse af Levitt og March [1988], der opererer med "overtroisk læring", altså at personer og organisationer kan drage forkerte konsekvenser af erfaringer).

Evaluering til læring opererer således ofte med en tavs antagelse om, hvad det er, der skal læres i forhold til og dermed hvordan indsigts- og kompetencehierarkiet er ordnet (før man kan begynde at stige opad). Denne tavse antagelse kan være rodfæstet hos en evaluator eller i antagelserne bag et anonymt indikatorsystem eller være mere eller mindre konsensusbaseret blandt de vigtigste interessenter. Afgørende for hele læringsbegrebet er imidlertid, at læring altid er et gode.

Af samme grund skildres modstand mod læring som regel som uhensigtsmæssig og uheldig. Modstand mod læring skyldes en psykologisk brist eller en mangel på kognitiv kapacitet. Ligeledes tales i dag ganske ofte om organisatorisk læring; i så fald består barrieren mod læring i uhensigtsmæssige kommunikations- og samarbejds mønstre på det organisatoriske plan. Problemet er i begge tilfælde, at man beklageligvis ikke er i stand til at lære. Der er ikke tale om decideret uvilje. Læringsproblemer henføres således ikke til en strid om værdier mellem evaluator og den lærende. På den måde "udglatter" læringsanvendelsen eventuelle værdikonflikter og taler i stedet om øget kompetence, udvikling, læring, forandringsparathed osv. (Om evaluerings udglattende funktion se endvidere Schwandt, 2003).

Forholdet mellem kontrol og læring:

Kontrolanvendelse og læringsanvendelse opstilles ofte som diametrale modsætninger. Der kan da også som skitseret være ganske god grund til at understrege de typiske forskelle mellem dem både i sigte, design og faktisk anvendelse. Hvis kontrol svarer til summativ evaluering og læring til formativ, argumenteres det fra formativ side, at læring kun sker, hvis der ikke samtidig er eksterne trusler og inspektion. Derfor holder teatre prøver uden publikum før premieren. Læring forudsætter et klima af åbenhed, tillid, og tolerance for fejl. Det er den modsatte mentalitet, der fremmes af ekstern kontrol, lyder argumentet.

Tilhængere af formativ evaluering vil derfor ofte sige: enten-eller. De har imidlertid det problem, at et eventuelt samfundsmæssigt behov for opsyn ikke forsvinder blot fordi man i det enkelte evalueringsarbejde vælger at sigte mod formativ anvendelse. I den henseende er formativ og summativ ikke funktionelle alternativer.

Den amerikanske evalueringsforsker Michael Scriven argumenterer for, at sondringen mellem summativ og formativ evaluering udelukkende er en sontring angående anvendelsen. Den samme evaluering kan altså bruges til kontrol eller læring, alt efter hvem der modtager evalueringsresultaterne og hvad der gøres ved dem. Synspunktet er sundt, i den forstand at sigte og design ikke behøver at sige særlig meget om den faktiske anvendelse. Endvidere kan forskellige parter på forskelligt organisatorisk niveau samtidig anvende evalueringsdata summativt og formativt. Man kan f.eks. sige nej tak til en enkelt udbyder af en ydelse (summativt), samtidig med at man udvikler hele politikken på dette ydelsesområde (formativt).

Begrebsligt er der visse fællestræk mellem kontrol- og læringsanvendelsen. Begge er rimeligt klart afgrænset i tid og rum. Det handler om som umiddelbar følge af evalueringen at fortsætte, nedlægge eller udvikle den pågældende indsats. Som regel kan både tilhængere af kontrol og tilhængere af læring enes om, at det er vigtigt at sikre en afklaring af formålet

med evaluering, inden man går i gang. Denne forskrift er kun meningsfuld under den antagelse, at der er en snæver sammenhæng mellem formålet og den faktiske anvendelse.

Man kan bruge formålserklæringen til at forudsige virkningerne med. I kontrolperspektivet kan man ikke se andre virkninger end den ønskede kontrol. I læringsperspektivet kan man ikke se andre virkninger end den ønskede læring. Til gengæld får man i hvert perspektiv nøjagtig det, som man forventer. Funktion er lig med sigte. Denne antagelse deles ofte af begge anvendelsesforståelser (men altså ikke af Scriven 1991).

I nogle kredse har man ofte satset på intern læringsorienteret evaluering, hvis man følte sig truet af ekstern evaluering. Som begrebet om evalueringsanvendelse til læring her er stillet op, kan man imidlertid også se svaghederne og de underbelyste sider ved denne strategi. Hvis læringsudfaldet på forhånd er garanteret (man kan jo altid sige, at der er sket en eller anden form for læring), og hvis læring altid er et gode, stilles der ikke ret store krav til selve evalueringsarbejdets egen kvalitet. Det er faktisk netop et problem ved intern "læringsorienteret" evaluering, at det af og til degenerer, så man slet ikke kan se selve evalueringselementet.

Yderligere et problem ved "læring" ud fra evaluering er, at det kan være svært objektivt at registrere, om læring faktisk er indtrådt. Er det nok, at nogen har opnået en eller anden erkendelse som følge af evaluering, eller skal der også ske en ændring i indsatsen? Hvis ja, hvilke krav stilles der så til denne ændring, for at vi kalder den "læring"? Skal det være en forbedring? Hvordan kender vi forbedringer?

Oplysende evalueringsanvendelse

Den amerikanske evalueringsforsker Carol Weiss har undersøgt instrumentel anvendelse (altså anvendelse i direkte lineær forlængelse af evalueringens formål) af evaluering og er nået frem til, at den finder sted ret sjældent (Weiss & Bucuvalas, 1980). Hun har i stedet lagt et langt bredere perspektiv i tid og rum på evalueringsanvendelse. I stedet peger hun på den oplysende brug som et positivt aspekt ved evalueringsarbejde. Her anvendes evalueringsresultater ikke direkte i forhold til evaluanden. I stedet spredes viden fra evalueringer i form af "knowledge creep", gradvis videnspredning, inden for forskellige faglige felter, policy-felter og organisatoriske felter. Viden, begreber og ideer spredes langsomt og gradvist. Evalueringer kan føje sig til hinanden ligesom forskningsresultater, der oplyser større vidensfelter og/eller resten af samfundet (uden at noget enkeltresultat nødvendigvis kan påberåbe sig særlig sandhedsværdi.) En af fiduserne ved evaluering til oplysning er, at resultaterne således kan bidrage til en bredere vidensudvikling i samfundet.

På længere sigt antager man imidlertid, at de mest pålidelige evalueringsresultater vinder over de andre. De gode ideer spredes, og de dårlige taber i den oplyste debat. Selve det at skabe debatten om dårligt virkende indsatser kan også være en oplysende funktion af evaluering. Oplysende evaluering kan således være gavnlig også selv om den, der oplyses, ikke deler værdier med evaluerings præmisser eller ikke har bedt om de pågældende evalueringsdata til et specifikt formål (så som læring eller kontrol).

Den oplysende evalueringsanvendelse er således positivt ladet. Man kunne hævde, at de samme processer, som leder til oplysende brug af evalueringer, også kan lede til formørkende brug. Fortalere for oplysende evalueringsanvendelse vil dog fastholde, at evalueringer i det lange løb er principielt formålstjenlige. Det formål, der tilgodeses, er imidlertid det større offentlige oplysende formål, ikke den enkelte evaluerings direkte formål. Forholdet mellem en evaluerings formål og funktion er således langt mere subtilt end tilfældet er ved kontrol og læring. Den enkelte evaluering har nemlig ofte til formål at betjene en konkret opdragsgiver med en specifikt efterspurgt evalueringsviden, mens evalueringens oplysende funktion sjældent ophøjes til at være det primære evalueringsformål, medmindre der er tale om egentlig forskning med evaluering som sidefunktion. Mange evaluatore og i særde-

leshed evalueringsforskere vil trods alt gerne satse så meget som muligt på evalueringernes oplysende funktion, idet man er opmærksom på, at en overdreven fokus på den enkelte opdragsgivers anvendelse af den enkelte evaluering på længere sigt kan svække både evaluators frihed til at problematisere overordnede policy-beslutninger og svække den overordnede, langsigtede samfundsmæssige gavn af evalueringer (Henry, 2000).

Tilhængere af den oplysende evalueringsanvendelse hænger således en stor del af deres forventninger på en knage af håb: Den vigtigste anvendelse er ikke en funktion af evalueringens formelle sigte, snarere tværtimod. I stedet håber man på en indirekte og langsigtet oplysende funktion af evaluering, hjulpet på vej af en offentlighed, der medvirker til at optimere udbyttet af mange evalueringer over tid. Offentligheden antages således at rumme en "medvirkende motor", der trækker evalueringerne i retning af formålstjenlighed.¹

Hvorvidt indikatorer spiller eller kan spille en tilsvarende rolle, afhænger i vidt omfang af den tolkningsmæssige og politiske sammenhæng, som de indgår i. Tal svarende til, hvad vi i dag kender som indikatorer, har i vidt omfang været benyttet som socialvidenskabeligt oplysningsmiddel for at vække den offentlige debat og gøre opmærksom på f.eks. antallet af epidemiramte, trafikdræbte, ledige, syge, osv., alt efter om man offentligt gjorde eller ikke gjorde noget. Alt efter tolkningsrammen kan indikatorer således godt spille en alment oplysende og politisk debatskabende rolle.

Weiss synes imidlertid at nå frem til denne funktion som en slags omvej eller bieffekt, efter at der er udført et evalueringsarbejde med et mere afgrænset sigte. Inden for den oplysende evaluering er man i øvrigt meget opmærksom på, at denne "almengørelse" af evalueringsresultater afhænger af, at der er formuleret og testet en nogenlunde troværdig teori om, hvorfor bestemte indsatser har ført til bestemte resultater. Kun så kan man uddrage en konklusion fra en evaluering, der lader sig bruge til bredere oplysning. Nøjagtig samme pointe gælder brugen af indikatorer. Kun hvis der bydes på en troværdig forklaring på hvilke indsatser, der fører til de ønskværdige scorer, kan iagttagelsen af indikatorer føre til en almen udbredelse af gode eksempler.

Carol Weiss' begreb om oplysende evalueringsanvendelse er udviklet i en tid, hvor medier ikke spillede samme rolle, som de gør i dag. I dag bærer nyhedsmedier, internationale organisationer og internettet evalueringsresultater hurtigt videre i tid og rum. Indikatorer formidles via medier, bla. digitale. I den forbindelse tiltager de sig nogle funktioner, som næppe let lader sig indordne under den oplysende funktion, som Weiss fandt, f.eks.

- Indikatorer bliver til nyheder. Nyhedsmedier bringer historier, hvor evalueringsresultater sættes ind i nyhedsmæssige sammenhænge. Nyheder er ofte sort-hvide og konfliktprægede. Nyheder kan ikke rumme en sags lange og komplekse forhistorie. Nyheder glemmes hurtigt. Sådanne kriterier spiller formodentlig med, når evalueringsresultater præsenteres i medierne. Mediernes overskrifter om skolen de senere år har formodentlig været et vigtigt indspark i skolepolitikken.
- Indikatorer former som images. Et image er en forestilling, som udmærker sig ved sin billedskabende kraft. For eksempel er forestillingen om et internationalt læsekapløb et image, som har knyttet sig til de offentlige billeder af de senere års internationale læseundersøgelser (Stronach, 1999).
- Indikatorer kan bruges som PR og reklame. Offentlige organisationer, der befinder sig i en konkurrencesituation, vil formodentlig i stigende grad benytte eventuelle

¹ Ifølge en dansk forsker (Rieper, 1996) er vi i Danmark imidlertid ikke særlig gode til at organisere os sådan, at evalueringernes oplysende funktion fremmes (det afhænger af netværk, tidsskrifter, konferencer mm. Og det handler om ikke at være blufærdig over andres vurdering af ens arbejde). Det afhænger tillige af et grundlæggende fokus på virkemidler snarere end på udførende enheder.

positive indikatorer i markedsføringsmæssig sammenhæng. Tolkningen af dem kan ske med markedsføringssproget som kode.

- Indikatorer som forbrugerinformation. Her benyttes de som en slags forbrugertest, der kan understøtte en beslutning om at vælge den pågældende institution til eller fra. Jo mere institutionen opfattes som et forbrugergode, jo mere vil denne brug af evalueringsinformation kunne vinde indpas.
- Offentliggørelse af indikatorer kan endvidere virke som et forstærket incitament for den enkelte institution til at forbedre sine resultater. Incitamentet kan formodentlig både sætte sig igennem på indholdet i ydelsen såvel som i forsøg på at påvirke målingen af ydelsens kvalitet. Det kan både virke i retning af at styrke kvalitetsarbejdet på f.eks. den enkelte skole samt i retning af at demotivere medarbejdere og ledere på skoler, hvor det af ydre grunde er vanskeligt at skabe gode resultater.
- Indikatorer kan endvidere virke som socialt mærkat. Det sker, når bestemte institutioner og/eller bestemte sociale kategorier stemples svarende til deres evalueringsscore (f.eks. skoler nederst på en rangliste). Det er en kendt sag fra socialforskning, at sådanne mærkater kan skabe deres egen realitet, ikke mindst når de mærkede efterfølgende forsøger at frigøre sig fra mærkningen eller forsøger at udnytte den. Ikke mindst hvis indikatorer knyttes sammen med ressourcetildeling til institutioner, kan der igangsættes en god eller dårlig udviklingsspiral for hver enkelt skole.

Disse mulige anvendelser af evaluering på den offentlige arena antyder, at evalueringer kan have funktioner, der ikke let lader sig indordne under Carol Weiss' optimistiske begreb om "oplysning" (Boyle, Breul og Dahler-Larsen 2008).

Courtney mfl. (2004: 1152) mener, at indikatorer kan bruges til at blive klog på sammenhænge mellem policy, indsats og resultat, altså en slags offentlig, policy-orienteret oplysende brug af indikatorer.

Strategisk og taktisk evalueringsanvendelse

Begreber om instrumentel såvel som oplysende evalueringsanvendelse har været kritiseret for at overse, i hvor høj grad evalueringer (og indikatorer) er politiske fænomener. Ud fra en opfattelse af organisationer som præget af interesser, ideologier og magtkampe, er der derfor udviklet begreber om hhv. strategisk og taktisk evalueringsanvendelse. I begge disse tjener evalueringer til at fremme skjulte, uofficielle formål. Således bruges evalueringsresultater strategisk til at enten torpedere eller hvidvaske bestemte aktiviteter, institutioner eller faggrupper. Der kommer ingen reel ny viden frem. Evalueringen bruges kun som ammunition til at fremme positioner, der allerede var indtaget, før evalueringen blev gennemført. Den strategiske evalueringsanvendelse kan udføres med større eller mindre finesse i det kunstneriske udtryk. Evalueringsarbejdet kan være styret af en mere eller mindre ideologisk dagsorden. Der kan skrives et mere eller mindre selektivt evalueringskommissorium. Evaluator kan vælges med større eller mindre omhu. Evaluators arbejdsbetingelser kan undervejs påvirkes. Informationer kan skjules eller formidles til de rette. Udvalgte evalueringsresultater kan blæses ud af proportioner. Sluttelig torpederes uønskede indsatser og institutioner. Alt dette er imidlertid ikke overraskende, men ganske forudsigeligt, idet det kan bruges til at fremme den underliggende, men skjulte politiske dagsorden.

Indikatorer kan i den sammenhæng udvælges, så de skaber det bedst mulige eller værst mulige billede af det, der overvåges, f.eks. i kraft af en bestemt standardsætning. De kan også designes således, at de skaber en konkurrence og et tilsigtet modsætningsforhold mellem de instanser, der monitoreres.

Den taktiske evalueringsanvendelse tjener også skjulte formål, men her er det evalueringsprocessen snarere end evalueringsresultaterne, der bruges. Når evaluering bruges taktisk, afviser man en kritisk diskussion med henvisning til, at en bestemt aktivitet evalueres. Man

bruger evaluering som en politisk "syltekrukke", hvor ubekvemme sager gemmes væk eller henlægges til bedre tider. Hvis et indikatorsystem er sat op, kan man således fortælle, at "direktoratet holder udviklingen i tallene under nøje opsyn", mens man forbereder eventuelle tiltag, som kan sættes ind "i det rette øjeblik" eller "når tiden er moden."

Det er imidlertid svært at identificere den strategiske eller taktiske evalueringsanvendelse på noget ydre kendetegn. Det er f.eks. ikke i sig selv forkasteligt at bruge en dårlig score på en indikator til at nedlægge en indsats med, som reelt er dårligt fungerende. Brugen er kun strategisk, hvis man alternativt alligevel havde ønsket at nedlægge indsatsen, selv om evalueringsresultaterne havde været positive. En klar identifikation af en strategisk evalueringsanvendelse forudsætter således data om den kontrafaktiske situation (altså en situation, der faktisk ikke indtræffer i praksis) (Vedung, 1998). Alternativt kunne man identificere en strategisk anvendelse på, at en bestemt politisk aktør havde forkastelige motiver, altså at vedkommende spekulerede i at bruge evaluering til et politisk formål. Men det er empirisk vanskeligt at få indsigt i en politisk aktørs indre motivation. Hvordan kan man med sikkerhed vide, hvad en politisk aktør egentlig er ude på? Et tilsvarende ræsonnement gør sig gældende ikke blot for den strategiske, men også den taktiske evalueringsanvendelse. At evalueringer tager tid er givet, men det er illegitimt at spekulere i at nøle. Det er spekulatøren i tidsforbruget snarere end tidsforbruget i sig selv, der gør en evalueringsanvendelse til taktisk. Det betyder i praksis, at den strategiske og den taktiske evalueringsanvendelse er ganske vanskelig at identificere. Med andre ord er det let at have mistanke om, at andre bruger indikatorer strategisk og taktisk, men at eftervise en sådan påstand forskningsmæssigt er temmelig krævende.

En af de vigtigste funktioner af begrebet om strategisk og taktisk evalueringsanvendelse er måske derfor netop den retoriske i forhold til en given evalueringsproces. Man kan søge at delegitimere indikatorer ved at henvise til, at de er strategisk eller taktisk motiveret. Men det er svært at bevise. Tilsvarende er det svært at bevise, at en evaluering ikke er strategisk eller taktisk motiveret. Når Danmarks Evalueringsinstitut for eksempel på sin hjemmeside anfører, at instituttet ikke har en skjult dagsorden, så er dette udsagn ikke i sig selv specielt overbevisende. Hvis man havde haft en skjult dagsorden, havde man jo pr. definition ikke fremlagt den på hjemmesiden.

lagttager man evaluering fra et strategisk og taktisk perspektiv, vil man se enhver udvikling i evalueringsbilledet som et tegn på en overordnet politisk dagsorden. Og man taler gerne om de næste, endnu usynlige skridt i denne udvikling, som lige nu er i gære, hvis man ellers har afkodet den skjulte dagsorden rigtig. Denne tanke kan af og til have et skær af konspirationsteori over sig. Men denne forestilling kan også kritiseres, hvis man læser nyere teori om viden-sanvendelse, hvor der lægges vægt på, at viden har sociale konsekvenser, som ofte er ikke-lineære og uforudsigelige.

Den symbolske evalueringsanvendelse

Når evalueringer anvendes symbolsk, handler det ikke om at bruge evalueringsresultaterne, hverken i forhold til evaluanden eller i forhold til oplysning på den større, samfundsmæssige arena. I stedet tjener evaluering til at symbolisere noget. Den måske vigtigste funktion af evaluering er måske at signalere, at der er overblik og kontrol. Det illuderes ofte, at hvis nogle indikatorer på et problem er under intens måling, så er der også "styr på" problemet. I et demokrati er det vigtigt, at der gives beretninger om ansvarlighed (March & Olsen, 1989).

Strengt taget kan man ikke slutte, at problemet er ved at blive løst, fordi det er ved at blive målt. Men målingen signalerer måske en villighed fra et kontrollerende systems side til at være opmærksom på problemet. Symbolpolitik kalder man det, når en af de vigtigste funktioner af en politisk beslutning er at vise handlekraft, beslutsomhed og ønske om ansvarsplacering.

I et symbolpolitisk perspektiv er den vigtigste funktion af lovbefæstede og standardiserede indikatorer måske at vise, at der er styr på et indsatsområde.

Det ligger imidlertid i begrebet om symbolsk evalueringsanvendelse, sådan som det kan bruges og understøttes af en hel del organisationsteori, at man samtidig undgår, at indikatorer har nogen andre væsentlige konsekvenser. Symbolikken er det funktionelle alternativ til en egentlig indholdsmæssig anvendelse. Der er tale om organisatorisk hykleri (Brunsson 1986), idet der opretholdes en stor forskel mellem, hvad der siges og gøres, eller når der blot løbende udstedes løfter om, at man holder øje med udviklingen, og vil gøre det (endnu) bedre i fremtiden. Organisationer er i stand til at optage nye organisationsopskrifter så som resultatmålinger under stor tvetydighed, således at der opretholdes stor afstand mellem den teoretiske model for resultatmålinger og organisationens faktiske opførsel (Modell 2004).

De institutionelle forventninger har at gøre med overordnede værdier og signaler om styring snarere end til de faktiske resultater, lyder i hvert fald dele af organisationsteorien, specielt her den nyinstitutionelle organisationsteori (Meyer og Rowan 1977: 341). (På tilsvarende vis er det vigtige ved sikkerhedskontrollen i lufthavne, at man er overbevist om, der er en sikkerhedskontrol, ikke at sikkerhedskontrollen faktisk fanger terrorister).

Hvis man i dag forlanger af skolen, at den skal resultatmåles, fint, så resultatmåles den, men det behøver ikke at have nogen væsentlig indflydelse på det arbejde, der udføres i forhold til elevernes resultater. Denne pointe underbygges og illustreres fint af en artikel Andersen (2004). Her drøftes en evaluering af en større indsats for implementering af målstyring i skolen. Den viser først og fremmest, at lærere ikke særlig godt kan huske kommunale målsætninger og skolens målsætninger. Målsætningsarbejdet blev opfattet først og fremmest som en bureaukratisering. Konceptet for målstyring og evaluering cirkulerer mellem skoler og mellem kommuner, men har meget begrænset indflydelse på undervisningen. En af nøglerne til forklaring af denne situation er, at man fra en styringssynsvinkel er optaget af, at styringskoncepterne følges, men uinteressert i meldinger om, hvordan styringskoncepterne virker på indholdet af det styrede. Ingen interesserer sig for at følge op på, om målstyring, resultatmåling og evaluering egentlig påvirker elevernes resultater (Andersen 2004: 463).

Kontrollanter er optaget af, at deres kontrolsystemer sættes op. Information om, at et kontrolsystem er sat op, kan gøre arbejdet meget lettere for kontrollanter, end hvis de selv skulle foretage mere indgående kontrol. Derfor fokuserer en hel del moderne evalueringssystemer på kontrolprocesser og organisatoriske procedurer (Munro 2004: 1079). Ikke mindst inden for auditering kender man denne effekt, men den er alment udbredt.

Man kan heller ikke udelukke et tæt parløb her mellem en ledelsesmæssig og en politisk interesse i symbolsk evalueringsanvendelse. Også politikere vil gerne give indtryk af, at der er politisk greb om sagerne, men samtidig giver indikatorer mulighed for at skubbe ansvaret ud til kommuner og institutioner, hvis der opstår en møgsag eller der scores dårligt på indikatorerne.

Vi er her ved at have fat i, at meget "symbolske" effekter altså også er "reelle nok." En væsentlig indvending går derfor på, at det symbolbegreb, der her opereres med, er for fattigt, fordi det skiller symbol fra indhold ligesom indbildning fra virkelighed. Hverken teoretisk eller empirisk kan det symbolske imidlertid holdes fast som rent symbolsk i betydning ikke-virkelig (Tompkins, 1987; Røvik, 1998). Derfor må vi have fat i mere forunderlige og omsigribende virkninger: de konstitutive.

Konstitutiv evalueringsanvendelse og deres "mekanismer"

Den syvende og sidste, såkaldte konstitutive evalueringsanvendelse, opererer med, at evalueringer over formende og skabende indflydelse på virkeligheden. Handlinger og tolkninger påvirkes på måder, som ikke altid er officielle. Denne indflydelse kan række hinsides det planlagte, tilsigtede og gennemskuelige.

Som Michael Power (1996: 293) er inde på, er de vigtigste funktioner af indikatorer, evaluering og kvalitetsmåling måske ikke at registrere kvaliteten, men at konstruere og definere kvalitet. Pointen er, at indikatorer kan øve indflydelse på de tolkninger og handlinger, der bestemmer den socialt konstruerede virkelighed.

Der findes i evalueringslitteraturen en række benævnte "mekanismer", der søger at indfange denne effekt. De er næppe gensidigt udelukkende. De fleste af nedenstående nævnt af Smith (1995):

- Indikatorfiksering. Man hæfter sig stift ved det, der måles i indikatoren på bekostning af større målsætninger. Man lader f.eks. alt over minimumsstandarden synke ned. Det sidstnævnte er også kendt som "tærskleffekten". (Bevan og Hood 2006: 521).
- Tunnel vision. Man får af den grund snæversyn, og bliver blind for f.eks. afledte effekter.
- Suboptimering. Man scorer så højt som muligt på lokale mål, afdelingsmål osv. på bekostning af større mål (se også McBriarty 1988).
- Nærsynethed. Man har fokus på kortsynede mål snarere end langsigtede.
- Misrepræsentation. Data manipuleres, enten ved direkte snyd eller ved julelege i afrapporteringen.
- Creaming. Man udvælger de klienter/brugere, der giver de bedste chancer for en pæn score.
- Referral. Man henviser klienter til andre instanser, for at de ikke skal ødelægge ens score på en indikator (se også Vulliamy og Webb 2001).
- Gaming. Denne betegnelse dækker forskellige strategiske adfærdsmæssige spil. Gaming indbefatter blandt andet, at man prøver at sætte en standard så lavt, at den bliver lettere at indfri. Det betyder også, at man ikke ønsker at score særlig højt i år, hvis dette års niveau bliver til næste års standard. Fænomenet er kendt fra Sovjet (!) (Smith 1995) og kaldes "the ratchet effect" (på dansk pal-nøgleeffekten eller skraldenøgle-effekten).²
- Ossification (forbening). Man satser på nutidens indikatorer og bliver dårligere til innovation. I det lys kan indikatorer fremme førsteordenslæring, men stille sig i vejen for anden-ordens, mere reflekteret læring, hvad enten denne er organisatorisk eller politisk motiveret.

En række af disse sammenfattes i begrebet "performance paradox", som betegner den situation, at mere måling af kvalitet ikke følges af øget kvalitet.

Mens de ovennævnte mekanismer leder i retning af decideret utilsigtet anvendelse af indikatorer - (Smith (1995) ser da også hele spørgsmålet om anvendelse af indikatorer som et spørgsmål om en principals opsyn med en agent, hvor man altså identificerer sig med principalens intentioner) - kan man også forestille sig mere subtile mekanismer, der afhænger af fortolkninger, hvor skellet mellem rigtigt og forkert, mellem tilsigtet og utilsigtet bliver meget vanskeligere at opretholde. I stedet handler det om, hvad opsynet skaber. "Audit make things auditable", som Power (1997).

² For at undgå et sammenrend af interesser mellem udførende led og de ledere, der afrapporterer opadtil (som fører til afrapportering af Potemkin-kulisser), har man ønsket mere uafhængige instanser til overvågning og evaluering (se f.eks. Bevan og Hood 2006: 530).

Her vil indikatorer blive udsat for tolkning og handling på baggrund af tolkning, uden at det nødvendigvis er indlysende, hvorfor en bestemt intention skal betragtes som eneste mulige målestok.

Konstitutive virkninger i forskellige domæner

Sådanne konstitutive virkninger af evalueringer kan sætte sig i forskellige aspekter af virkeligheden.

Jeg vil i det følgende operere med, at konstitutive virkninger kan sætte sig i følgende:

- en problemforståelse ud fra et verdenssyn
- ydelsens indhold
- en tidsramme for en ydelse
- sociale relationer og identiteter knyttet til leveringen af en ydelse.

Lad os først undersøge, hvordan indikatorer hænger sammen med et bestemt syn på verden eller et fortolkningsmæssigt univers.

Indikatorer viser hen til måder at forstå verden på. At sociale fænomener fra en kultur vurderes ganske anderledes i andre kulturer, viser, at en bestemt måde at evaluere noget på er relativ i forhold til hele den verdensorden, som ethvert socialt fællesskab konstruerer.

Indikatorer kan forstås som policy-instrumenter, og sådanne producerer specifikke effekter uafhængigt af deres erklærede formål og de strukturerer et politikområde ud fra deres egen logik (Lascoumes og Le Gales 2007: 10).

På nogle indsatsområder, f.eks. det sociale, har indsatsen kun været beskrevet af en meget "flydende" og rudimentær terminologi, så som f.eks. "at yde omsorg" eller "at være der for borgeren." Når der indsættes indikatorer og standarder, har selve introduktionen heraf dermed en betragtelig diskursiv og formende effekt på indsatsområdet (Munro 2004: 1084).

Indikatorer kan medvirke til at fremme en bestemt forståelse af et politikområde. Testresultater er med til at definere, hvad meningen med skolen er. Når man inden for undervisningen af dansk for voksne fremmedsprogede lægger vægt på beskæftigelse som succeskriterium, afspejler det et andet syn på den aktivitet, end hvis man havde satset på kultur- og sprogforståelse.

Indikatorer til sammenligning af kommuner, institutioner eller andre producerende enheder kan tolkes i lyset af den præmis, at variationer mellem enhederne er politisk uacceptable. Det betyder ikke, at det, der måles, er et validt og anerkendt kriterium for "kvalitet", at der er kontrolleret for enhedernes socioøkonomiske forskelle, at udligning af forskellene er økonomisk efficient og/eller at der er tale om en legitimt vedtagen politisk standard. Men indikatorerne kan som konstitutiv effekt have at fremme den politiske billeddannelse, at variation er illegitim (Smith 1995: 296).

Suen (2003) viser i en fin artikel, hvordan eksamenssystemet i et land (Kina) gennem flere århundreder ikke blot afspejler bestemte magt- og kulturforhold i landet, men også medvirker til at vedligeholde disse. Samtidig har dette system forhindret landets uddannelsessystem i at rette sig mod udvikling af andre kompetenceområder. Det skulle vise sig senere at give væsentlige konkurrencemæssige ulemper for landet.

I en case skildrer Arlie Hochschild (2004), hvordan moderne travle mandlige ledere kan lade sig evaluere af deres familiemedlemmer ud fra et nyt koncept om "personlig familieevaluering." Ved et første øjekast er det jo progressivt, at familiens perspektiv bringes så tydeligt frem på lederens indre radarskærm. På den anden side risikerer evalueringssyste-

met også at medføre, at lederen må anlægge et objektiverende syn på sine familierelationer, der ender med at underminere eller i hvert fald redefinere disse relationer.

Her er vi ved en kernepointe angående konstitutive virkninger af evaluering. På den ene side afspejler evalueringer et bestemt syn på verden og en bestemt problemforståelse. På den anden side sender evalueringer også en mere eller mindre tydeligt formuleret symbolsk fortolkningsramme tilbage til virkeligheden.

Et klassisk retorisk og politisk kneb er at forvandle diskussioner om, hvordan sociale indretninger bør være til et faktuel spørgsmål om, hvordan de er (Mortensen 2004). Når først et evalueringssystem er etableret, er det jo et faktum, at der evalueres i forhold til dets kriterier. Det er altså for fattigt blot at forestille sig, at evalueringer skal registrere virkeligheden. Når de virker konstitutivt, er de også med til at fastslå, hvordan virkeligheden skal undersøges, hvilke data om virkeligheden, der kan gælde, og hvad der er problemet i virkeligheden lige nu.

Evalueringer virker ikke mindst gennem sproget. Evalueringssproget udnævner noget til at være "en evaluand". Mere eller mindre underforstået er der et vist problem ved denne evaluand, for eksempel "der er for store forskelle i effektiviteten på institution X og Y". Indikatorer kan udtrykke, hvad der er vigtigt ved evaluanden. Der markeres en problemforståelse, som virkeligheden har at forholde sig til.

Konstitutive virkninger på en ydelses indhold

Det er en klassisk iagttagelse i organisationsteorien, at måden, en organisation evalueres på, kan komme til at påvirke dens opfattelse af, hvad der er indholdsmæssigt og sagligt væsentligt ved arbejdets resultater (Ridgway, 1956; Jasinsky, 1956) - inklusive, at en organisation, der evalueres dysfunktionelt, vil komme til at fungere dysfunktionelt. F.eks. blev den polske møbelindustri før jerntæppets fald vurderet på, hvor mange tons møbler, den kunne producere pr. år. Det medførte, at det polske møblement blev verdens tungeste.

I en undervisningspraksis kan effekten være den såkaldte "teaching-to-the-test". Det betyder, at undervisningen kommer til at handle om at score så højt som muligt på givne test eller evalueringskriterier. Disse kriterier bliver målsætninger i sig selv.

Jo mere en bestemt indikator bliver af afgørende politisk og organisatorisk betydning, jo større er samtidig den eventuelle belastning af indikatorens integritet. Vulliamy og Webb (2001) viser blandt andet, hvordan en indikator (i dette tilfælde udelukkelse af elever fra skoler) bliver til en social konstruktion, som er et resultat af mange forskellige aktørers reaktionsmønstre. Disse reaktioner kan selvfølgelig påvirke indikatorens validitet og reliabilitet.

Man taler i den forbindelse om den såkaldte indikatorfiksering, altså at evalueringskriteriet eller indikatoren så at sige ender med at blive mere vigtig end det gode arbejde som sådan. Indikatorfikseringen findes, når simple målbare evalueringskriterier kommer til at stå for hele det gode arbejde, uanset om disse kriterier er dårlige udtryk for kvaliteten i det, der skal måles. I metodisk sprog vil man her tale om, at indikatoren har *lav validitet*, den måler ikke det, den skal måle. Indikatorfikseringen findes imidlertid også i mere avanceret form. Den findes, når indikatoren begrebsligt set *bliver* definitionen på kvalitet. Det sker, hvis diskussionen af, hvad god læsning er, erstattes af en operationel (målbar) definition på læsning eller med andre ord netop en indikator. En avanceret indikatorfiksering sikrer ikke blot, at adfærden styres i retning af indikatoren, den sørger også for at sætte sig definatorisk på det begreb, som hele indsatsen handler om.

Hvis man i en kommune for eksempel beslutter sig for, at skolekvalitet er lig med høj score på bestemte tests, er det svært bagefter at argumentere for, at testen har lav validitet i forhold til de politiske målsætninger.

Derfor er det både politisk og ledelsesmæssigt en stor udfordring at beslutte sig for indikatorer. Man risikerer, at indikatorens integritet nedbrydes. Man risikerer også, at man får, hvad man beder om, og ikke andet. Tager man begrebet om konstitutive virkninger alvorligt, så er indikatorer politisk kommunikation. De er udtryk for, hvad man vil. Mennesker forholder sig til dem. De har virkninger.

Konstitutive virkninger på en ydelses tidsorientering

Der kan opstå konstitutive virkninger ikke blot i forhold til det saglige indhold i en offentlig indsats, men også i forhold til dens tidsorientering. I mange tilfælde omformes mere kvalitative forhold, som er svære at måle, i vores kultur til enten penge eller tid. Ganske ofte bliver mere eller mindre valide evalueringskriterier udtrykt i tid eller "oversat" til noget tidsligt (Tilbury 2004; Munro 2004: 1080). Når kvalitet er svær at måle, vælger man i stedet indikatorer, der siger noget om, hvor hurtigt det går, hvilket fører til et disproportionelt fokus på tidsindikatorer (Munro 2004: 1080).

I England havde man problemer med lange ventetider på skadestuerne. Derfor lavede man et kvalitetssikringssystem, der skulle overvåge, hvor lang tid, der gik, fra den enkelte patient kom ind, til man faktisk mødte noget sundhedspersonale. Derfor ansatte skadestuerne nogle såkaldte "hello-sygeplejersker", der blot gik rundt og sagde "hello" til nyankomne patienter.

Også i England lavede politikerne en aftale med nogle hospitaler om at reducere ventelisterne til gengæld for en ekstrabevilling. Ventelisternes omfang blev defineret som antal måneders gange antal personer fra første konsultation på hospitalet til den faktiske indlæggelse. Det viste sig hurtigt, at projektet lykkedes. Imidlertid klarede hospitalerne opgaven ved at undlade at tage patienter ind til første konsultation.

Man har bedt ambulancer vente uden for skadestuen med patienter for at være sikker på, at man inden for dørene kunne være klar til at behandle patienten inden for den afsatte ventetid (Bevan og Hood 2006: 531). Bevan og Hood (2006) refererer til empiriske resultater, der tyder på betragtelige konstitutive virkninger af sådanne ventetidsindikatorer.

En hospitalsafdeling i Danmark kvalitetsvurderes f.eks. løbende på, om den inden for senest 3 dage efter undersøgelse af en patient tilsender en epikrise (beskrivelse af undersøgelsens resultater) til patientens praktiserende læge. Den pågældende afdeling sender imidlertid ofte en blodprøve til klinisk-kemisk afdeling, som er mere end 3 dage om analysen. For at leve op til kvalitetsdefinitionens deadline fremsender afdelingen alligevel en epikrise til praktiserende læge, selv om brevet er indholdsmæssigt tomt. Når undersøgelsesresultaterne senere faktisk forelægges, fremsendes den egentlige epikrise. For at leve op til kvalitetskriteriernes tidsdimension må afdelingen således sende to epikriser, en for at leve op til kvalitetssystemet og en for at leve op til formålet med en epikrise.

På en dansk skadestue uddeler man et papkort med teksten: "Dit behandlingsbehov er vurderet til prioritet 4. Det betyder, at behandlingen tilstræbes påbegyndt inden for 3 timer." På bagsiden står: "behandlingen er påbegyndt, når du kaldes ind fra venteværelset. Der kan opstå yderligere ventetid imellem de forskellige faser i behandlingen. Denne ventetid er ikke medregnet i Skadestuens kvalitetsmål for ventetiden." De konstitutive virkninger heraf er ikke kendte, men man kan lave velbegrundede hypoteser herom.

Resultaterne af undervisning, ikke mindst af samfundets nye medlemmer, har i princippet et vældigt langt tidsperspektiv. Indikatorer må i princippet forkorte effektbilledets tidshorisont, i og med at der skal findes udtryk her og nu for undervisningens værdi og/eller effekter i evalueringsøjeblikket. Dette problem kan ikke løses metodisk; begrænsningen er rent logisk. På grund af gabet i tid mellem undervisningens principielt meget langsigtede virkninger

ger og indikatorers bundethed til nutiden er undervisning moden for en række forestillingsmæssige og ideologiske konstruktioner af, hvad der kan tænkes at ske efter evalueringen og fremover som resultat af undervisningen. Men indikatorer må alligevel operere med en eller anden vedtagen deadline. Det, som kan fanges inden deadline, tæller med.

På langsigtede og forebyggende indsatsområder må denne effekt formodes at være størst.

Evalueringsens konstitutive effekt langs den tidsmæssige dimension består i, at man kan komme til at indrette en indsats efter, at (erkendelsen af) dens evaluerbare effekter skal vise sig på det tidspunkt, hvor en given evaluering finder sted. Hvis klare mål gælder for 3., 6., og 9. niende klasse, er der hermed sat tidslige grænser for, hvornår gode målbare effekter bør have vist sig.

En evalueringsmæssig deadline vil, som flere af ovenstående eksempler viser, sjældent være neutral over for bestemmelsen af indsatsens indhold. Den vil også kunne påvirke de involverede mennesker og deres relationer.

Konstitutive virkninger på mennesker og deres relationer

Samtidig med at en indikator kommunikerer om evaluandens indhold og dens tidsperspektiv, metakommunikerer den også om de involverede parter identiteter og relationer.

Af særlig identitetsmæssig interesse er i vor tid kategorien "brugeren". Ganske ofte former evalueringer i vor tid forholdet mellem de involverede parter som et forhold mellem en bruger og komplementært dertil en serviceleverandør. Brugeren er historisk en relativt ny kategori (Vedung, 1998). Konstituerende for brugeren er typisk en individualiseret valgsituation og et sæt præferencer, der konfronteres med et givet servicetilbud udmønter sig i en tilfredshedsgrad. Det bemærkes, at brugeren ikke bestemmes ud fra et ansvar for politikdannelsen eller dens konsekvenser. Det bemærkes også, at brugeren (ligesom sin nære slægtning forbrugeren) samtidig med sin tilsyneladende sociale statusfremhævelse er genstand for en lang række kontrol-, socialiserings- og normaliseringsmekanismer (Bauman, 1983).

Gennem tests og evalueringer placeres skolens elever i forhold til, om deres faglige og personlige udvikling er foran, på niveau med, eller bagud i forhold til det forventede på deres klassetrin, ligesom et eventuelt særligt behov for specialundervisning kan komme på tale. Gennem deltagelse i eksamener, tests og evalueringer lærer nogle elever gradvist at blive egnede subjekter for evaluering.

Også på de offentligt ansattes side kan evalueringer naturligvis være medkonstituerende. For eksempel kan lærere skildres som fagprofessionelle, som konsulenter og sparringspartnere på elevernes læreprocesser, som serviceleverandører og som ansatte. Mange har traditionelt antaget, at læreridentiteten som andre fagprofessionsidentiteter er mere eller mindre givet med uddannelsen og jobbet; nyere forskning anlægger imidlertid det perspektiv, at fagprofessionel identitet står til løbende konstruktion.

I et sådant perspektiv tilbyder professionelle og organisatoriske processer som evaluering nogle byggesten, som læreren kan arbejde videre med i sit løbende identitetsarbejde (Halford & Leonard, 1999: 107). Indikatorer giver tilbagemeldinger at tolke på, og enhver bruger tilbagemeldinger fra omverdenen i sit identitetsarbejde. Forskellige slags evalueringer giver imidlertid ganske forskellige muligheder. Indikatorer baseret på standardiserede mål har en tendens til at trække det professionelle element ud af den professionelle; en professionel er nemlig i princippet en, der har et mandat til at bruge kompleks viden til at træffe varierende beslutninger med i forhold til komplekse situationer. Er der høj grad af standardisering i arbejdet (som nogle indikatorer stipulerer), så er behovet for professionel ageren så meget mindre. Andre evalueringer, der er baseret på fagprofessionel viden, lægger i stedet op til, at den fagprofessionelle netop træder i karakter.

De stærkeste tendenser for tiden inden for skolen trækker i retning af at standardisere forventningerne til lærernes indsats og resultater. Det samme gør sig gældende på andre områder. Mange ønsker også at kunne bruge løndifferentiering som redskab i en sådan udvikling. Statskundskabsstuderende og økonomer (de kommende administrative ledere!) lærer i vidt omfang, at fagprofessionelle er egennyttmaksimerende rationalister, som udnytter asymmetrisk information til at skaffe sig fordele med – og desuden ønsker at lave så lidt som muligt. Denne opfattelse fører til, at man anbefaler klare indikatorer og klare incitamentsystemer. Mange mener, at der vil være yderligere gevinster at hente ad den vej. Hvad det dominerende paradigme imidlertid overser, er at man samtidig kan tabe de gratisgevinster, der kunne have været ved et samvittighedsfuldt og hårdt arbejdende lærerkorps præget af en stærk offentlig etos og en højprofileret professionsetik. Disse kvaliteter kommer selvfølgelig ikke ud af ingenting. Men indikatorer over deres egen indflydelse i forhold til professioners selvforståelse og holdningsmåde.

Her er det også af betydning, hvilke sammenligninger den fagprofessionelle eventuelt udsættes for. På nogle arbejdspladser evalueres den fagprofessionelle i forhold til gennemsnittet af sine kolleger. Alt andet lige vil et sådant system incitere til, at man ikke engagerer sig i kollegialt samarbejde og erfaringsudveksling. Indikatorer kan fremme konkurrence og en oplevelse af splittelse mellem dem, der sammenlignes.

Endelig kan der ske en konstitutiv identitetsmæssig påvirkning af den institution eller organisation, der er ansvarlig for en bestemt offentlig ydelse. Dels kan institutionen selv begynde at tale om sig selv og præsentere sig selv i lyset af evaluering. I et projekt om udvikling af evalueringsmodeller fandt man imidlertid, at en skole kunne "indkode" sig selv både som virksomhed, pædagogisk miljø, politisk institution og som fællesskab. Disse fire indkodninger gav forskellig anledning til at iscenesætte evaluering som et spejl for skolens liv (Dahler-Larsen 2003). Også omverdenens tolkning af en bestemt organisation/institution sættes naturligvis i spil. I nogle tilfælde kan indikatorer skabe et positivt eller negativt billede, alt efter hvordan en standard sættes. Vælges et givet år som standard i forhold til det følgende, sker en unødigt afstraffelse af organisationer, der allerede i standardåret gjorde en god indsats. Courtney mfl. (2004: 1150) viser et eksempel på, hvordan en organisation, der gør det godt på et indsatsområde, kan blive ringere stillet på et andet, fordi de mest vanskeligt stillede klienter er efterladt på dette andet område.

Konstitutive virkninger forplanter sig

I figur 1 opsummeres de tre aspekter af en ydelse, som konstitutive virkninger kan ramme, nemlig den indholdsmæssige, den tidslige og den identitetsmæssige. Det foreslås endvidere, at de konstitutive virkninger af evaluering kan manifestere sig i tre forskellige slags ændringer, kaldet førsteordens, andenordens og tredjeordens ændringer.

I førsteordensændringer påvirkes alene det sprog, som man bruger til at beskrive problemer, kriterier og løsninger med. I andenordensændringer træffes en beslutning – en sanktioneret begivenhed, der markerer en officiel opmærksomhed, for eksempel organisatorisk og/eller politisk opmærksomhed. Det kan for eksempel være en beslutning om at tildele ekstra timer til et hidtidigt underprioriteret indsatsområde. I tredjeordensændringer påvirkes den institutionelle orden på et bestemt område. Et eksempel herpå er oprettelsen af et evalueringsinstitut. I amerikansk evalueringsforskning taler man f.eks. om "washback effects", når evalueringssystemer virker tilbage på læseplanerne og undervisningens tilrettelæggelse.

Figur 1 Typer af konstitutive virkninger

	Ændringer af:		
Domæne:	Første orden: tale	Anden orden: beslutning	Tredje orden: institutionel struktur
Problemforståelse/verdenssyn			
Indhold			
Tid			
Sociale identiteter og relationer			

Figur 1 kan bruges til at vise, at ikke alle konstitutive virkninger af evaluering når lige langt. Nogle evalueringer når kun at demonstrere en smule af deres eget evalueringssprog og klinger derefter af. Andre kan imidlertid have langt mere vidtrækkende virkninger. Konstitutive virkninger kan forplante sig både på langs og på tværs i figuren. En evaluering af, hvorvidt elever har nået fastsatte mål i tredje, sjette og niende klassetrin, er formuleret som tidslige krav, men kan have konsekvenser både for undervisningens indhold og for sociale identiteter og relationer – specielt for de børn, som viser sig ikke at nå de fastsatte mål.

En analyse af "Fælles Sprog", et monitoreringssystem på ældreområdet kendt blandt andet for sine "stregkoder", har således vist, at systemet gennem sin løbende overvågning medvirker til at flytte forankringen af omsorgen fra den indholdsmæssige og den sociale dimension til den tidslige, idet hele Fælles Sprog er bygget op på tidsenheder afsat til varetagelsen af bestemte på forhånd kodede omsorgsproblemer. Til gengæld virker den stramme tidsstyring naturligvis tilbage på den leverede omsorgs faktiske indhold (de konkrete ydelser som varierer fra gang til gang) og på relationen mellem de ældre og hjemmehjælperne (dobbel umyndiggørelse) (Højlund & la Cour, 2001: 98-99).

Virkninger af indikatorer kan forplante sig i flere led, når aktører reagerer på evalueringer, og der reageres på reaktionerne. Hvis en skole for eksempel offentligt præsenteres som "nederst på en rangliste" kan dens lærerkreds og ledelse beslutte sig for at sætte bestemte handlinger i gang. Skolen kan forsøge at skaffe ekstra midler begrundet i dens særligt vanskelige forhold. Der kan også ske en øget uddifferentiering af elever til specialklasser for, at disse ikke skal forstyrre det evalueringsbillede af skolens "normale elever", som den først og fremmest bliver målt på. Specialklasseeleverne og deres forældre kan reagere på denne uddifferentiering. Samtidig kan nogle af de fagligt stærkere elever og deres forældre samt nogle af lærerne forlade skolen for ikke at blive delagtiggjort i skolens negative image i offentligheden. Skolen kan også sætte supplerende evalueringsarbejde i gang for at fremme andre billeder af skolen end ranglistens evalueringsbillede.

Et andet eksempel på konstitutive virkninger, der forplanter sig, er det forhold, at en hel del evalueringsarbejde i Danmark de senere år formodentlig er inspireret af eller udsprunget af internationale undersøgelser. Når læseundersøgelserne først i 1990'erne skildrede Danmark som helhed, opstod der en logisk pres i retning af også at begynde at skildre de enkelte skolers resultater. I første omgang dannede man evalueringsinstituttet, som ikke fik til opgave at lave sammenligninger. Senere kom lovgivningen om offentlig indsigt i gennemsnitlige karakterer for hver skole. Nogen begynder også at presse på med øget ledelsesmæssig og/eller offentlig indsigt i de enkelte læreres kvalitet, selv om dette perspektiv har noget længere udsigt. Selv om sammenhængene næppe er direkte og lineære, så skaber evalueringer følgevirkninger af ganske forskellig art meget forskellige steder i skolesystemet.

Med andre ord er der både en vis omsættelighed og en vis afsmitning mellem de forskellige lag i skemaet.

2.3 Hvilke kontekstuelle forhold kan tænkes at spille ind på konstitutive virkninger?

I almindelighed kan man således ikke regne med, at det er forhåndsbestemt, hvorvidt og hvor meget indikatorer vil påvirke. En oplagt indvending mod begrebet om konstitutive virkninger er naturligvis således, at ikke alle indikatorer når at have nogen væsentlige konstitutive virkninger. Mange glemmes formodentlig, inden de når at gøre nogen væsentlig forskel.

Inden for organisationsteori sonderer man mellem forskellige reaktionsmønstre (Dahler-Larsen 2004):

- Afvisning, hvor indikatorer så vidt muligt holdes ud fra en bestemt praksis.
- Hykleri, hvor indikatorer gøres op, men det samtidig sikres, at virkningerne kun omfatter udtalelser og erklæringer på det retoriske plan.
- Tæmning, hvor indikatorer gøres op, men der sorteres i dem og i deres anvendelser så man søger at sikre, at der kun findes virkninger af en type, som man kan leve med i en given organisatorisk eller politisk sammenhæng.
- Kolonialisering, hvor indikatorers konstitutive virkninger over en bred kam og i dybden kommer til at præge efterfølgende praksis inden for et givet område.

Vi kan udvikle rimelige hypoteser om, at virkningerne af evaluering vil være stærkere (de nederste på listen) under nogle omstændigheder end andre. Man kan formode, at den konstitutive effekt afhænger af:

- Graden af institutionalisering af indikatorer/evaluering.
- Hvor stærkt f.eks. faglige og kollegiale værdier fastholdes i arbejdet.
- Sagområdet, herunder om indikatorer opleves som relevante og valide for sagområdet.
- Hvilket bolværk den lokale organisationskultur og fagprofessionalismen evt. udgør.
- Hvilke belønnings- og sanktionssystemer, der i øvrigt spiller ind på en offentlig ydelses tilrettelæggelse og udførelse, samt hvorvidt forskellige kontrol- og evalueringssystemer eventuelt spiller sammen med eller modvirker nogle indikatorer og/eller om flere indikatorsystemer modvirker hinanden.
- Offentliggørelse af indikatorer vil formodentlig også medføre en skærpelse af de konstitutive virkninger (Endahl & Utzon 2002). Men forskning peger på, at "framingen" af de offentliggjorte data også er vigtig (Andersen og Dahler-Larsen 2008).
- Analyseniveauet, som indikatorerne oparbejdes på (institution, afdeling, individ osv.).
- Indikatorernes tilblivelseshistorie. I kommunerne kan nationale indikatorer virke anderledes end kommunernes egne, ligesom man generelt kan forestille sig, at frivilligt etablerede indikatorer virker anderledes end påtvungne.
- Indikatorernes alder. Økonomer kan fortælle, at lav rente og faldende rente ikke har samme effekter. På tilsvarende vis kan man forestille sig, at indførelsen af nye indikatorer ikke har samme effekter som indikatorer som sådan. Der kan være opmærksomheds- og tilvænnningseffekter. Der kan således også være forskel på et indikatorsystem og et foranderligt indikatorsystem. Lawn (2007) er inde på, at det er helt almindeligt, at indikatorsystemer forandrer sig. Udviklingen af systemet er normen. Det er ukendt, hvilken betydning nye indikatorer har i forhold til indikatorer som sådan. Der findes teori om indikatorers løbende nedslidning (jf. "performance paradox", se van Thiel og Leeuw (2002)).

En undersøgelse fra FTF/Catinet (Dahler-Larsen 2008) tyder også på, at det gør en forskel fra sektor til sektor, hvorvidt dokumentationen opfattes af de fagprofessionelle som fagligt

meningsfuld og medvirkende til kvalitetsudvikling. Af mindre betydning ser det ud til at være, om dokumentationsarbejdet tager tid fra borgerne og om det har udviklet sig hastigt de seneste par år. Disse tolkninger er af meget overordnet karakter og må måske modificeres, når det gælder specifikke indsatsområder og kontekster.

2.4 Hvordan begrebsliggøres og vurderes ”konstitutive virkninger”?

Tidligere bidrag inden for evalueringsforskningen har identificeret paradoksale virkninger af evaluering (van Thiel og Leeuw 2002), men ofte med det samme benævnt disse patologiske (Bouckaert mfl. 1991), perverse (Power 2005: 335) eller utilsigtede (Smith 1995; Courtney et al 2004)). Patologisk er en teoretisk utilfredsstillende medicinsk/biologisk metafor, fordi der reelt ikke findes nogen konsensus om en ”sund” måde at evaluere på, og det er ikke indlysende, hvad det i givet fald skulle betyde.

Utilsigtet er ved nærmere eftertanke også et problematisk begreb. Ganske vist kan der selvfølgelig forekomme virkninger, som er klart utilsigtede. Men faktisk at demonstrere, at en given virkning er utilsigtet, kræver en indsigt i aktørers intentioner, som vi ikke har adgang til. Det ville også kræve en specifik definition af, hvis intentioner, der ville tælle med. Hos mange identificerer man sig implicit med en tiltænkt rationel policy-arkitekt, eller stipulerer blot, at initiativet er ”well-intended” (Werner og Asch 2005: 1243).

Lad os tage det eksempel, at brugerorienteret skoleinformation (eksempelvis med gennemsnitlige karakterer som indikator) uden kontrol for elevernes socioøkonomiske baggrund vil medføre flere bevidste valg af skole blandt familierne i Danmark. Muligvis vil det føre til øget social differentiering, fordi de ”gode” skoler er dem, hvor velstillede folks børn går. De velstillede vil oven i købet vælge skoler med velstillede børn. Muligvis vil statistiske eksperter hævde, at evalueringsinformationen vil få utilsigtede virkninger eller ”ikke kan bruges”, fordi ingen indikatorer giver et validt udtryk for skolens kvalitet, når der ikke tages højde for børnenes sociale baggrund. Men denne indvending forhindrer ikke indikatorerne i at blive brugt.

Fra de velstillede side vil informationen blot være så meget mere brugbar, fordi der ikke kontrolleres for sociale forskelle. Man kan direkte ud af tallene læse, hvor de dygtige børn går helt uafhængigt af, om dygtigheden skyldes hjem eller skole. Det styrker evalueringsinformationen set fra de velstillede synspunkt. Hvordan kan man vide, om denne effekt er ”tilsigtet” eller ”utilsigtet”? Det er et ubehageligt teoretisk spørgsmål, fordi det er så umanerligt vanskeligt at afgøre. Tæller vi med, at nogle evalueringer virker tilbage på de begreber, de siger at måle, for eksempel via den såkaldte avancerede indikatorfiksering, bliver det umuligt at argumentere, at evalueringen giver en ”skæv” måling. Hvad vi i stedet er henvist til, er at notere, at indikatorer for skolekvalitet findes, og at de kan have virkninger. Mange virkninger. Interessante virkninger. Andre virkninger end de officielle.

Denne betragtning har en række vigtige konsekvenser. En vigtig konsekvens er, at hvad vi betragter som ”validiteten” af en given indikator afhænger af de mange lokale handlinger, som brugen af indikatoren sætter i gang. Man kan forestille sig, at jo mere en indikator bliver brugt som officiel målsætning og som led i officiel styring og politikdannelsen, jo mere belastes den med sociale konstruktioner som gør dens ”validitet” vanskeligere at stole på (Vulliamy og Webb 2001: 367).

Konsekvensen af denne betragtning er også, at konstitutive virkninger ikke nødvendigvis er negative. Konstitutive virkninger handler om at påvirke handlinger og tankegange inden for et felt i kraft af den måde, der iværksættes måling på. Om disse handlinger og tankegange vurderes negativt, afhænger helt og holdent af vurderingsinstansen. For nogen er konstitutive virkninger en fiks måde at tilvejebringe omstilling i den offentlige sektor på. For andre er

det ideologisk magtudøvelse. Man må med andre ord være eksplicit i sit perspektiv både teoretisk og normativt, hvis man vil bruge begrebet om indikatorers konstitutive virkninger på en klar måde.

Moderne vidensteoretikere sætter iagttagelser som disse i perspektiv. Stehr (1994: 95) definerer viden ud fra sin socialt produktive karakter, dvs. viden former virkeligheden. Viden er med i virkeligheden. Virkeligheden er allerede vidensformet eller sociologisk, som Giddens (1994) ville sige. En klassisk antagelse om at viden "bruges" på virkeligheden er for instrumentel og/eller bureaukratisk. Denne antagelse overser, at magten er fragmenteret, at virkeligheden reagerer tilbage på vidensproduktionen, og at vidensproduktionen-i-sin-kontekst er refleksiv, altså selv-kommenterende og selv-konstituerende (Giddens 1994: 44-45).

Det betyder blandt andet, at en af de vigtigste virkninger af indikatorer kan være at incitere til mere evalueringsarbejde eventuelt fra ganske andre vinkler.

2.5 Hvad betyder det, at indikatorer indgår i managementsystemer til evaluering/dokumentation osv.?

De senere år er fokus inden for evalueringsfeltet flyttet en del. Hvor man tidligere havde fokus på den enkelte evaluering og dens anvendelse i situationen (Patton 1997) (jævnfør f.eks. begrebet "evaluability assessment" og dets vigende relevans), er man i dag optaget af evalueringskultur, evalueringssystemer og opbygning af evalueringskapacitet. Samtidig drejer det sig ikke om enkelte evalueringsstudier, men om strømme af evalueringsinformation (Rist og Stame 2006). Indikatorer indgår heri. Indikatorer er ofte netop ofte med i systemer til løbende monitorering.

Medvirkende til denne udvikling har formodentlig været:

- Dårlige erfaringer med uforudsigelighed i enkeltstående evalueringer.
- Dårlige erfaringer med manglende anvendelse og opfølgning på enkelte evalueringer (den klassiske anvendelsesproblematik i evaluering).
- Ønske om en mere systematisk dækning af alle indsatsområder med evaluering.
- Ønske om at reducere kompleksiteten i overvågning og kontrol. Af denne grund bevæger f.eks. revision sig fra at kigge i bilagene til at overvåge den reviderede organisations egne revisionssystemer jf. Power (2005).
- Inspiration fra managerialismen i New Public Management (ud fra nøgleord som performance, management, indikatorer, effektivitet, kontrakter mm).
- Ønske om risikohåndtering såvel organisatorisk som politisk. Usikkerhed om udfald og mulig kritik kan "omformes" til håndtering af kalkulerbar risiko ved hjælp af organisatoriske procedurer til måling og overvågning.
- En politisk logik (støttet af mediernes fokus på "møgsager") giver politikere incitamenter til at sætte overvågningssystemer op uden dog selv at tage ansvar for selve udfaldet. Politikere holder øje og træder i karakter ved at placere ansvar når møgsagen er der, men tager ikke selv ansvaret. Til denne logik er et indikatorsystem egnet (Lawn 2007).

Men nogle af konsekvenserne ved sådanne overvågningssystemer er formodentlig:

- Stigende abstraktion af evalueringskriterier (i og med at de primært udspringer af et ledelses- og opsynsperspektiv og bruges til sammenligning mellem indsatsområder).

- Kvalitet defineres i stigende grad af organisationer, der er specialiseret i at overvåge kvalitetssystemer (Lindeberg 2007). Derved anonymiseres i indikatorsystemer de valg af værdier, der leder frem til evalueringskriterier (Greene 1999).
- Der holdes gang i spil om at undgå risiko og sikre ansvarsplacering andre steder (Hood 2002; Rothstein et al 2006; Power 2004).
- Der skiftes generelt fokus i evalueringsarbejdet fra "kritisk vurdering" til "sikring".
- Kvalitetsbegrebet bliver mere defensivt.
- Anvendelsesproblematikken skifter formodentlig ansigt. Manglende anvendelse bliver mindre relevant, symbolsk anvendelse og konstitutiv anvendelse bliver mere relevant. Organisationer må "konstruere sig", så de bliver til at holde opsyn med (Power 1997: 51).
- Formodentlig stiger ressourcebeslaglæggelsen til kvalitetsdokumentation også. Godt nok "rationaliserer" og systematiserer de her nævnte systemer kvalitetsdokumentationen, men samtidig foreskrives også obligatorisk måling og dokumentation og evaluering på områder, hvor man måske tidligere havde foretaget en mere kølig vurdering af fornuften ved evaluering i den foreliggende situation (jf. "evaluability assessment"). I hvert fald er Rigsrevisionen opmærksom på omkostningssiden ved rutineret og institutionaliseret evaluering.

Med andre ord må en analyse af indikatorers virkninger sættes i perspektiv af, hvordan overvågningssystemer baseret på indikatorer i øvrigt vinder indpas i vores samfund.

Courtney mfl. (2004: 1152) nævner muligheden for, at indikatorer bruges ikke blot til ansvarsplacering og til opsyn, men også i øget grad til at lære om sammenhængen mellem policy, indsats og resultat.

3 RESSOURCETRÆK TIL INDIKATORER

3.1 Hvor bliver ressourcer til evaluering/dokumentation/kvalitetsindikatorer af?

Omkostninger i tid, ressourcer, energi (personale) til kvalitetsdokumentation, herunder indikatorer, er ved at blive et debatpunkt i Danmark. En undersøgelse fra FTF/Catinet af offentlige ledere tyder på omfattende og stigende dokumentationskrav, mens mange finder, at tiden til dokumentation kun i begrænset omfang er givet godt ud. Problemerne herved kan være:

- nedslidning af personale
- ressourcerne kunne alternativt være brugt på bedre service

Det kan være vanskeligt at kortlægge præcist, hvor mange ressourcer kvalitetsdokumentation beslaglægger. Indikatorsystemer og tilsvarende systemer er ofte ikke indført efter nøje analyse af, hvor mange ressourcer de vil kræve, og de observerer sjældent deres egne omkostninger (Power 2005). Det gælder både det synlige og det usynlige ressourcetræk, både penge og arbejdstid. Der er formodentlig en del usynlige omkostninger, fordi kvalitetsdokumentation osv. er vanskelig at afgrænse fra andre arbejdsopgaver, især de steder hvor det er lykkedes at flette dokumentationsopgaverne ind i de daglige arbejdsrutiner (f.eks. journal-skrivning i sundhedsvæsenet). Pointen med en del dokumentations- og evalueringsarbejde de senere år har jo netop været at flette det meget sammen med det daglige arbejde og med en løbende udvikling. En række indikatorprojekter har da også søgt at sikre, at der ikke skulle samles nye, ekstraordinære data ind, og at indikatorerne i alle tilfælde kunne baseres på digital information. Antagelsen er her, at digital betyder tidsbesparende.

Opkomsten af en række nye enheder med særskilte opgaver og særskilt personale til dokumentation af kvalitet kan imidlertid trække i en anden retning, fordi deres eksistensberettigelse består i at sætte data sammen på måder, som er unikke og særegne for netop hver af disse institutioner. Derfor oplever offentlige myndigheder ofte at skulle afrapportere på forskellig måde om deres aktiviteter til flere forskellige eksterne kvalitetsinstanser.

Uanset måleproblemer må spørgsmålet om resourcebeslaglæggelsen til oparbejdning af indikatorer være et meget væsentligt spørgsmål. Også Rigsrevisionen (2005) er som nævnt opmærksom herpå.

Sammenfatning om perspektiver på indikatorer:

Fortalere for indikatorer kan hæfte sig ved mulighederne for at bruge indikatorer til

- kontrol
- læring
- oplysning.

Kritiske perspektiver på indikatorer omfatter:

- Indikatorer beslaglægger tid og penge (men har måske først og fremmest symbolsk anvendelse, idet deres ideal og deres faktiske virkning i praksis er stærkt forskellige [Modell 2004]).
- Indikatorer har en række utilsigtede virkninger (varianter af konstitutive virkninger, f.eks. triviell indikatorfiksering mm).
- Indikatorer kan (misbruges) bevidst politisk (f.eks. taktisk og strategisk brug).
- Indikatorer kan være med til at definere, hvad det gælder om på et politisk indsatsområde (varianter af konstitutive virkninger, f.eks. avanceret indikatorfiksering). Sidstnævnte kritik kommer i to varianter, dels en ideologisk, hvor "udfaldet" er ideologisk forudbestemt, f.eks. neoliberalt, eller i en sociologisk-konstruktivistisk, hvor "udfaldet" er noget, der folder sig ud som et resultat af mange kræfter og mange handlinger, ofte på en uforudsigelig måde.

4 FORSLAG OM REPARATIONER AF INDIKATORERS NEGATIVE VIRKNINGER

4.1 Hvilke forslag har været nævnt til modvirkning af indikatorers uønskede virkninger?

Der har været fremsat en række forskellige forslag til imødegåelse af uønskede virkninger af indikatorer. De ser dog i et vist omfang ud til at være gensidigt modstridende, hvilket dels kan skyldes emnets kompleksitet og dels hensynet til, at et måske nærmest perfekt indikatorsystem vil kræve urimeligt store ressourcer både i tid og penge.

Blandt forslagene har været en logisk undersøgelse af indikatorer (Munro 2004:1086), idet en del virkninger er relativt forudsigelige. Man kan f.eks. stille spørgsmålet: Hvis høj kvalitet afspejles i en god score på indikatoren X, er en høj score på indikatoren X også i det konkrete tilfælde et tegn på høj kvalitet?

Til imødegåelse af mere eller mindre meningsløse indikatorer har man foreslået øget inddragelse af lokale medarbejdere samt fagpersoner med indsigt i indsatsområdet. Man løber dog hermed i princippet en risiko for, at de indikatorer, der er lettest at leve op til, bliver valgt. Alene at lade den opsynshavende (principalen i en principal-agent model) vælge indikatoren har imidlertid også sine svagheder (van Thiel og Leeuw 2002: 276).

Så har der været foreslået helt uafhængige målinger. Disse har imidlertid en tendens til kun yderligere at bidrage til en stigende dokumentationsbyrde.

I samme retning trækker det, når man til modvirkning af få, problematiske indikatorer ønsker et øget antal for at opnå et mere dækkende billede af en indsats. Omfangsrige indikatorsystemer bliver nemlig også udsat for kritik. Både flere og færre indikatorer har været opstillet som generel medicin.

Især hvad angår ressourcetrækket til indikatorsystemer har man foreslået forskellige målinger til identifikation af disse på et tidligt tidspunkt, inden sådanne systemer etableres (i stil med de fra KREVI kendte VAKKS-undersøgelser). Et egentligt dokumentationsstop har også været foreslået (Dahler-Larsen 2008). Tankegangen er her, at man efter et vist tidspunkt ikke må indføre yderligere nye obligatoriske dokumentationskrav uden at luge ud i mængden af eksisterende krav.

Endelig er foreslået mere systematiske, forskningsbaserede studier af indikatorers virkninger.

5 AT STUDERE KONSEKVENSER AF INDIKATORER

5.1 Hvordan kan man studere konsekvenser af indikatorer?

Enhvert studie må forholde sig til følgende:

- Man må være eksplicit i valget af begrebsdannelse og interesseområde.
- Man må være eksplicit i sin afgrænsning af, hvilke indikatorer med hvilke evt. standarder tilknyttet hvilke enheder, der er tale om.
- Man må se indikatorerne i sammenhæng med de større politiske, organisatoriske og fagprofessionelle systemer og kontekster, de fungerer i.
- Man må være opmærksom på, at mange indikatorsystemer er under løbende udvikling og at det måske er en vigtig del af deres funktion [Lawn 2007].

Man kan i princippet vælge:

1. *Vurderinger af effekterne fra involverede aktører* (f.eks. interviews eller spørgeskemaer).

Svagheden herved er:

- Aktører overdriver ofte de positive effekter ved initiativer, de godt kan lide eller selv har taget samt de negative effekter ved fænomener, de ikke bryder sig om.
 - Aktører er ikke altid i stand til at se de væsentlige forandringer, der sker i de større systemer omkring dem.
 - Aktørers egne tilskrivninger af årsags-virkningsforhold er ideologiske eller upræcise.
2. *Procesorienterede analyser* (hvor man f.eks. via casestudier søger at afkode virkningsprocesser, hvem gør hvad som følge af hvad i hvilken kontekst). Der kan benyttes en flæthed af metoder, inkl. spørgeundersøgelser. Forskellen på (1) og (2) er imidlertid, at under (2) tager forskeren ansvaret for at slutte noget om kausale forhold.

Svagheden herved er:

- Undersøgelsen sker ud fra mere eller mindre eksplicite antagelser om hvilke effekter, der er værd at kigge efter. Væsentlige effekter uden for dette perspektiv overses. I case-orienterede, kvalitative analyser vil man forsøge at imødegå en sådan effekt gennem et fleksibelt design og et reflektivt, metodisk arbejde.
- Det eksakte årsags-virkningsforhold kan være svært at bestemme og temmelig kontekstafhængigt.

3. *Objektive mål* (ændringer i ventetider osv.)

Svagheden her er:

- Der fanges kun ændringer i de eksakte kriterier, som måles. Der findes eksempler på sådanne påvisninger af ændringer i objektive mål. Men andre væsentlige konstitutive ændringer overses. Og hvis den daglige praksis,

som det objektive kriterium skulle fange, ændres, så nedslides indikatorens validitet (Vulliamy og Webb 2001).

- De data, som er vanskelige at få fat i på grund af organisatoriske forhold, kan måske være de mest interessante.
- Kun hvis designet er stramt, kan man entydigt slutte noget om årsagsvirkningsforhold.
- Selv om den interne validitet kan være høj i et sådant studie, kan den eksterne validitet være begrænset.

6 LITTERATUR

- Andersen, Simon C. (2004): "Hvorfor bliver man ved med at evaluere folkeskolen", *Politica*, 4 452-68.
- Andersen, Vibeke N. & Peter Dahler-Larsen (2008): "The Framing of Public Evaluation Data: Transparency and Openness in Danish Schools", in *Open to the Public*, Richard Boyle, Jonathan D. Breul & Peter Dahler-Larsen (red.). New Brunswick: Transaction Publishers, 99-116.
- Bailey, K. M. (1996): "Working for Washback: A Review of the Washback Concept in Language Testing", *Language Testing*, 13 (3) 257-79.
- Bauman, Zygmunt (1983): "Industrialism, Consumerism and Power", *Theory, Culture and Society*, 1 (3) 32-43.
- Bevan, G & C. Hood (2006): "What's Measured is what Matters: Targets and Gaming in the English Public Health Care System", *Public Administration*, 84 (3), 517-534.
- Bouckaert, Geert & Walter Balk (1991): "Public Productivity Measurement: Diseases and Cures", *Public Productivity & Management Review*, XV (2) Winter 229-35.
- Boyle, Richard, Jonathan Breul & Peter Dahler-Larsen (red.) (2008): *Open to the Public: Evaluation in the Public Arena*. New Brunswick: Transaction.
- Brunsson (1986): "Organizing for Inconsistencies: On Organizational Conflict, Depression and Hypocrisy as Substitutes for Action", *Scandinavian Journal of Management Studies*, May, 165-185.
- Christie, Christina A. & Alkin, Marvin C. (1999): "Further Reflections on Evaluation Misutilization", *Studies in Educational Evaluation*, 25, 1-10
- Courtney, Mark E., Barbara Needell & Fred Wulczyn (2004): "Unintended Consequences of the Push for Accountability: The Case of National Child Welfare Performance Standards", *Children and Youth Services Review*, 26, 1141-54.
- Dahler-Larsen, Peter (2003): "Hvad er Selvevaluering?", i *Selvevalueringens Hvide Sejl*, Dahler-Larsen, Peter, Vibeke Normann Andersen, Kasper Møller Hansen & Carsten Strømbæk Pedersen. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, Peter (2004): *Evaluering og Magt*. Århus: Magtudredningen.
- Dahler-Larsen, Peter (2006): *Evalueringens kultur*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, Peter (2008): *Kvalitetens Beskaffenhed*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen, Peter (2001): "New Public Management og Evaluering", i *Modernisering av offentlige sektor: Udfordringer, Metode og dilemmaer*, Tor Busch, Erik Johnsen & Kurt K. Klausen (red.). Oslo: Universitetsforlaget, 359-76.
- Endahl, Lars A. & Jan Utzon (2002): "Fører offentliggørelse af Kvalitetsmålinger i Sundheds-væsenet til Bedre Kvalitet?", *Ugeskrift for Læger*, 164.

- Fountain, Jane E. (2001): "Paradoxes of Public Sector Customer Service", *Governance*, 14 (1).
- Giddens, Anthony. (1994): *Modernitetens konsekvenser*. København: Reitzel Forlag.
- Greene, Jennifer C. (1999): "The Inequality of Performance Measurements", *Evaluation*, 5 (2), 160-172.
- Halford, Susan & Pauline Leonard (1999): "New Identities? Professionalism, Managerialism and the Construction of the Self", i *Professionals and the Managerialism in the Public Sector*, Mark Exworthy & Susan Halford (red.). Buckingham: Open University Press.
- Hellstern, G. M. (1986): "Assessing Evaluation Research", i *Guidance, Control and Evaluation in the Public Sector*, F. X. Kaufmann, G. Majone & V. Ostrom (red.). Berlin: de Gruyter, 279-313.
- Henry, Gary T. (2000): "Why not Use?" i *New Directions for Evaluation*, Valerie J. Caracelli & Hallie Preskill (red.). San Francisco: Jossey-Bass Publishers.
- Hochschild, Arlie (2004): "Gennem sprækker i tidsfælden", i *Arbejdssamfundet*, Michael Hviid & Jens Tonbo (red.). København: Hans Reitzels Forlag, 109-31.
- Hood, Christopher (2002): "The Risk Game and the Blame Game.", *Government and Opposition*, 37 (1) 15-37.
- Højlund, Holger & Anders La Cour (2001): "Standardiseret omsorg og lovbestemt fleksibilitet - organisationsændringer på et kærneområde i velfærdsstaten", *Nordiske Organisasjonsstudier*, 2 (3) 91-117.
- Jasinsky, F. J. (1956): "Use and Misuse of Efficiency Controls", *Harvard Business Review*, xxxiv, 105-12.
- Lascombes, Pierre & Patrick Le Gales (2007): "Introduction: Understanding Public Policy through Its Instruments - From Nature of Instruments to the Sociology of Public Policy Instrumentation", *Governance*, 20 (1), 1-21.
- Lawn, Martin (2007): "The Platinum Metre: Standardising a European Education Policy Space.", Keynote speaker ved EES konference i Odense, oktober 2007.
- Levitt, Barbara & James G. March (1988): "Organizational Learning", *Annual Review of Sociology*, 14, 319-336.
- Lindeberg, Tobias (2007): *Evaluative Technologies: Quality and the Multiplicity of Performance*. København: Copenhagen Business School.
- March, James G. & Johan P. Olsen (1989): *Rediscovering institutions: The organizational basis of politics*. New York: The Free Press.
- Mark, Melvin M & Gary Henry (2004): "The Mechanisms and Outcomes of Evaluation Influence", *Evaluation* 10 (1), 35-57.
- McBriarty, Mark A. (1988): "Performance Appraisal: Some Unintended Consequences", *Public Personnel Management*, 17(4) 421-34.

- Meyer, J. & B. Rowan (1977): "Institutionalized Organizations: Formal Structure as Myth and Ceremony", *American Journal of Sociology*, 83, 340-63.
- Modell, Sven (2004): "Performance Measurement Myths in The Public Sector: A Research Note", *Financial Accountability & Management*, 20 (1), 39-55.
- Mortensen, Niels (2004): *Det Paradoksale Samfund*. København: Hans Reitzels Forlag.
- Munk, Lis (2007): "Forhandling og Udvikling af Standarder indenfor Fødselsområdet" Oplæg holdt på seminaret *Standarder og Kvalitet i Sundhedsvæsenet*. Center for STS. 23. November 2007. Aarhus Universitet.
- Munro, Eileen (2004): "The Impact of Audit on Social Work Practice", *British Journal of Social Work*, 34, 1075-95.
- O'Neill, Baroness O. (2006): "Making Accountability Intelligent", Keynote speaker ved EES Konference i London 2006
- Patton, Michael Quinn (1997): *Utilization-focused Evaluation. The New Century Text*. 3. edition. Thousand Oaks: Sage.
- Pawson, Ray and Nick Tilley (1997): *Realistic Evaluation*. London: Sage.
- Perrin, B. (1998): "Effective Use and Misuse of Performance Measurement", *American Journal of Evaluation*, 19, 367-79.
- Power, M. (1997): "From Risk Society to Audit Society", *Soziale Systeme*, 3, 3-21
- Power, Michael (2004): *Risk Management of Everything*. London: Demos.
- Power, M. (2005): "The Theory of the Audit Explosion" i *Oxford Handbook of Public Management*, Ferlie, Evan, Laurence E. Lynn & Christopher Pollitt. New York: Oxford University Press, 327-344.
- Ridgway, V. F. (1956): "Dysfunctional Consequences of Performance Measurements", *Administrative Science Quarterly*, 1 (2), 240-7.
- Rieper, O. (1996): *Bør Danmark have en evalueringspolitik?* København: Institut for Statskundskab, Københavns Universitet.
- Rigsrevisionen (2005): "14/04 Beretning om Statens Anvendelse af Evaluering", København: Statsrevisoratet.
- Rist, Ray C. & Nicoletta Stame (red.) (2006): *From Studies to Streams*. New Brunswick: Transaction Publishers.
- Rothstein, Henry, Michael Huber & George Gaskell (2006): "A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk.", *Economy and Society*, 35 (1), 91-112.
- Røvik, Kjell A. (1998): *Moderne Organisasjoner. Trender i organisasjonstenkingen ved tusenårsskiftet*. Bergen-Sandviken: Fagbokforlaget.
- Schwandt, Thomas A. (2003): "Back to the Rough Ground! Beyond Theory to Practice in Evaluation", *Evaluation*, 9 (3).

- Scriven, Michael (1991): "Beyond Formative and Summative Evaluation", i *Evaluation and Education: At quarter century*, Milbrey W. McLaughlin & D. C. Phillips (red.). Chicago: University of Chicago Press, 18-64.
- Shulha, Lyn M. & J. B. Cousins (1997): "Evaluation use: Theory, research, and Practice since 1986", *Evaluation Practice*, 18 (3), 195-208.
- Smith, Peter (1995): "On The Unintended Consequences of Publishing Performance Data in the Public Sector", *International Journal of Public Administration*, 18 (2/3), 277-310.
- Stake, R. E. (2004): *Standard-based and Responsive Evaluation*. Thousand Oaks: Sage.
- Stehr, Nico (1994): *Knowledge Societies*. London: Sage Publications.
- Stronach, I. (1999): "Shouting Theatre in a Crowded Fire: "Educational Effectiveness" as Cultural Performance", *Evaluation*, 5 (2), 173-93.
- Suen, Hoi K. (2003): "Some very Long-term and some Chronic Effects: The Invisible Dimension of the Consequential Basis of Validity", Paper præsenteret på "*International Symposium on Developing High-stakes Testing*" ved Korean Institute of Curriculum and Evaluation, Seoul, Korea
- Tilbury, Clare (2004): "The Influence of Performance Measurement on Child Welfare Policy and Practice", *British Journal of Social Work*, 34, 225-241.
- Tompkins, P. (1987): "Translating Organizational Theory: Symbolism over Substance", i *Handbook of Organizational Communication*, Frederic M. Jabil, Linda Putnam, Karlene Roberts & Lyman Porter (red.). Newbury Park: Sage.
- van Thiel, Sandra & Frans L. Leeuw (2002): "The Performance Paradox in the Public Sector", *Public Performance and Management Review*, 25 (3) 267-81.
- Vedung, Evert (1998): *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
- Vulliamy, Graham & Rosemary Webb (2001): "The Social Construction of School Exclusion Rates: implications for evaluation methodology", *Educational Studies*, 27(3) 357-70.
- Weiss, Carol H. & M. J. Bucuvalas (1980): *Social Science Research and Decision Making*. New York: Columbia University Press.
- Werner, Rachel M. & David A. Asch (2005): "The Unintended Consequences of Publicly Reporting Quality Information", *JAMA*, 293 (10), 1239-1244.