

Interne evalueringer

Redskab til kontrol eller udvikling af offentlige organisationer?

Finn Hansson
WP 6/2005

MPP Working Paper No. 6/2005 ©

Juni 2005

ISBN: 87-91181-92-5

ISSN: 1396-2817

**Department of Management, Politics and Philosophy
Copenhagen Business School**

Blaagaardsgade 23 B

DK-2200 Copenhagen N

Phone: +45 38 15 36 30

Fax: +45 38 15 36 35

www.mpp.cbs.dk

Indholdsfortegnelse:

Evalueringssamfundet – refleksivitet og modernisering	4
Forskellige former og typer evaluering	5
Evalueringer og NPM-reformpolitik	11
De usynlige interne evalueringer i evalueringslitteraturen	16
Interne evalueringer og organisationsudvikling	19
Vidensproduktion og interne evalueringer.	25
Empowerment evaluation - deltagelse i vidensproduktion:	29
Kvalitet i interne evalueringer – hvem kontroller viden og hvordan	33
I stedet for en slutning: Lake Wobegon effekten i evalueringer	41

Evalueringssamfundet – refleksivitet og modernisering

At vi lever i et samfund hvor evalueringer er blevet et fast indslag i hverdagen er ubestrideligt. Dahler-Larsen taler i sin sidste bog (2004) om en evalueringsbølge der skyller henover os, Nepper Larsen taler om en evalueringsfeber (2004), men man kunne også bruge en formulering, som viderefører det kritiske perspektiv på kontrolfunktionen i evalueringer, som ligger i Powers meget udbredte begreb, revisions-samfundet (The audit society) og tale om at vi lever i et evalueringssamfund. Der er under alle omstændigheder tale om en udvikling, som de moderne, vestlige samfund har gennemlevet de seneste 10-15 år, hvor brug af evalueringer i den offentlige sektor er gået fra at være et relativt sjældent syn, ofte forbundet med en vurdering af større programmer og projekters virkninger, når de var afsluttet, (ex post facto evalueringer) til nu at være en mere eller mindre integreret del af den moderne organisations liv. Baggrunden for denne udvikling skal ikke behandles nærmere her (se Hansson 1998, Dahler Larsen 1998, 2004, Dahler-Larsen og Krogstrup 2001, Foss Hansen 2003), men har udspring i nødvendige strukturelle reformer og omlægninger af de senmoderne vestlige samfund. Disse forandringer som samtidig også indebar nye krav og funktioner i forhold til de traditionelle former for politikudvikling der var baseret på omfattende systematiske undersøgelser som grundlag for politiske beslutninger¹, og som vi kender det fra Danmark, hvor eksempelvis SFI i en årrække i 1960'erne og 70'erne var storleverandør af omfattende redegørelser og undersøgelser, som indgik som beslutningsgrundlag for politiske reformer. De omfattende samfundsforandringer har også medført ændringer i grundlaget for denne politikudviklingsmodel, for med Ulrich Beck's ord er den sociale omverden under så hastige forandringer, at en empirisk kortlægning af mere stabile sammenhænge er højst usandsynlig². Store undersøgelser af sociale forhold baseret på studier af udviklingen over længere tidsforløb som vi kender så mange af fra 1950'erne og frem til 1970'erne kan i mindre og mindre omfang præsentere gyldige resultater, i og med at grundforudsætninger for det undersøgte er under konstant forandring.

¹ Dette var ikke helt situationen i USA, hvor koblingen mellem et stort og decentralt samfund og en svag forbundsstat tidligt gjorde det nødvendigt med en mere fleksibel form for refleksiv politikudvikling og kontrol, se Albæk (1988).

² "The highly specialized empirical social research presumes a constancy of categories and hence a high and actually rather rare social stability."(Beck 1997)

Forskellige former og typer evaluering

At evaluere eller vurdere er noget der forekommer og indgår i alle sociale sammenhænge og er tæt indvævet eller sammenflettet med social handlen i både mikro og makro sammenhænge. Vi vurderer og bruger vurderingen reflektivt hele tiden når vi handler - både vores private liv og ikke mindst i organisations- og arbejdslivet. Det nye er, at der i det senmoderne samfund er sket en række afgørende forandringer i samfundets stabilitet, som har fjernet eller i det mindste reduceret betydningen af de traditionelle handlingsrationaler, som er baseret på værdier og sociale normer for både enkeltpersoner og organisationer af enhver slags. Beck fremhæver den tiltagende individualisering i den anden eller sen-moderniten (Beck 1999, 2000), hvor normer og andre sociale reguleringsmekanismer ikke længere har samme stabilitet som tidligere og samtidig skal anvendes i samfundsmæssige konstellationer, der hastigt forandres. Svækkelsen af stabile normer og/eller af de sociale sammenhænge de skal fungere i, både i privatliv og i samfundsliv, kombineret med presset for hastigt at tilpasse sig nye krav om ændringer og indstille sig på nye vilkår er – sammen med det forhold, at meget få områder i de senmoderne samfund kan holde sig ude fra disse krav – med til at sætte nye vilkår for hvorledes vi opfatter tid og hvilke krav vi stiller til grundlaget for beslutninger i valgsituationer³. Det er i denne samfundsmæssige situation i efterkrigstiden, at vi skal finde evalueringernes genese. Det handler i meget høj grad om at finde nye sociale teknologier til at udvælge og underbygge beslutninger og politikker, og tilpasse dem til nye krav og orienteringspunkter. Dette forstærkes i situationer, hvor hverken eksisterende normer og traditionelle værdisæt rækker til og hvor velfærdssamfundets højt udviklede traditioner for rationel og langsigtet plantækning også er kommet til kort. Dette sker ikke mindst når en af plantækningens forudsætninger, det løbende input fra de omfattende empiriske studier af udviklingstendenser, som kunne kortlægges og ekstrapoleres, også har nået sine grænser.

I denne situation tilbyder evalueringer en form for praktisk anvendt samfundsforskning som, modsat de store, empiriske surveyundersøgelser med lange produktionstider, kan levere relative hurtige svar på de stillede spørgsmål. Det er ikke noget helt nyt. Allerede i 1968 kunne Charles Wright konstatere, at

”two trends stand out in modern attitude toward evaluation. First, evaluation has come to be expected as a regular accompaniment to

³ Dette ses tydeligt i den nye organisationsteori, hvor forestillinger om lineære sammenhænge og kontinuerlig til må vige for en orientering mod processer og relationer. Se også Chia 2002.

rational social-action programs. Second, there has been a movement toward demanding more systematic, rigorous and objective evidence of success. The application of social science techniques to the appraisal of socialaction programs has come to be called evaluation research.”

Wright fortsætter og påpeger, at

“A scientific approach to the assessment of a program’s achievements is the hallmark of modern evaluation research. ... Since the burden is on the evaluator to provide firm evidence on the effects of the program under study, he favors a study design that will tend toward maximizing such evidence and his confidence in the conclusions drawn from it. ... the first concern is to obtain basic evidence on effectiveness, and therefore most research resources are allocated to this goal.” (Wright 1968, 197-198)

Men siden Wright beskrev dette dengang relativt nye felt, som især havde udviklet sig i USA i forbindelse med 60ernes store politiske og sociale reformer (Albæk 1988) er det gået hastigt frem med dette område.

Der er i de senere år vokset en omfattende evalueringslitteratur frem som en konsekvens af, at evalueringsområdet har udviklet sig kraftigt hen mod en professionalisering af evalueringer som et særligt ekspertområde. Udførelser af evalueringer har for længst forladt de traditionelle vidensproducerende institutioner, universiteter og forskningsinstitutter, og findes i dag som en fremtrædende aktivitet på linie med andre ekspertydelser (konsulentfirmaer), som en særlig del af den statslige styring og kontrol (statslige evalueringsinstitutter, rigsrevision)⁴ samt integreret som en særlig aktivitet i større offentlige organisationer. I denne uoverskuelige verden findes et udtal af definitioner og andre beskrivelser af evalueringer, der på systematisk forsøger vis at beskrive dette felt. Trods denne mangfoldighed af definitioner og beskrivelser er det dog muligt at gruppere de fleste et sted imellem de to nedenstående definitioner, som fra hvert sit hold forsøger at nærme sig en mere systematisk beskrivelse af hvad evalueringer er. Den ene definition ser evalueringer som et forsøg på at belyse sammenhængen mellem værdisæt og intervention, og den anden fokuserer på den systematiske brug af samfundsvidenskabelige metoder i vurdering af programmer.

⁴ Foss Hansen, H. (2003). *Evaluering i staten. Kontrol, læring eller forandring?* København: Samfundslitteratur.

Det omfattende evalueringsbegreb finder vi eksempelvis hos Evert Vedung, for hvem

"en ambitiös utvärdering försöker besvara två frågor. Motsvarer resultatet det som man kan förvänta eller önska sig utifrån något värdekriterium? Är dette resultat på något sätt en produkt av interventionen?" (Vedung, 1991/1997).

I en af de mest udbredte lærebøger finder vi en definition på evalueringer med fokus på systematik og brug af samfundsvidenskabelig metode. For Rossi og Freeman (1993) er evalueringer

"the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs".

Forskellen på de to definitioner er, at Vedung sammenkobler værdier og mål med evalueringen, dvs. inddrager den politiske proces, mens Rossi og Freeman først og fremmest ser på tilgangen, dvs. metoden i evalueringsprocessen. Eleanor Chelimsky, mangeårig leder af evalueringskontoret i Government Accountability Office (GAO), USA, fremhæver i modsætning til disse to definitioner, specielt formålsaspektet ved evalueringer, hvor især 3 centrale og adskilte formål understreges:

- evaluation for accountability (e.g., measurement of results or efficiency),
- evaluation of development (e.g., the provision of evaluative help to strengthen institutions),
- evaluation for knowledge (e.g., the acquisition of a more profound understanding in some specific area or field) (Chelimsky 1997, 10).

Ikke mindst Chelimsky's fremhævelse af disse tre centrale formål introducerer anvendelsesaspektet som helt centralt for forståelsen af evalueringers rolle og funktion i det moderne samfund. Anvendelse af evalueringer har i litteraturen været diskuteret ved brug af meget forskellige opdelinger og modeller (se f. eks. Albæk 1996, Krogstrup, Albæk og, Foss Hansen i Dahler-Larsen og Krogstrup 2001, Hansson 1997, 1999 og Dahler-Larsen 2001). Man har talt om politisk anvendelse ud fra en slags social engineering forståelse, om en organisationskulturel anvendelse til udvikling og senere udvidet mod organisatorisk læring, om anvendelse af evalueringer til rationalisering af beslutningsprocesser og som politisk redskab og til at legitimere beslutninger og sent ankommet også som mere integreret element i organisationers liv (interne evalueringer). Disse aspekter kan så igen kobles sammen med spørgsmål om evalueringers opdragivere og

brugergrupper, om evaluatorers roller, om metodetilgange, som kan være kvantitative eller kvalitative og om præsentationer og formidling i forskellige miljøer og sammenhænge.

En ting har disse forskellige definitioner fælles, nemlig at de ser evalueringer som baseret på systematisk brug af viden produceret ved hjælp af samfundsvidenskabelige metoder og begreber og som opererer med en ret traditionel model for forskerens rolle i evalueringer, ret lig den som traditionelt har været den anvendte samfundsforskers. Denne kan bedst som idealtipe beskrives med Poppers begreb om socialt ingeniørvæsen (social engineering), hvor den rationelle anvendelse af videnskabelig viden er i højsædet, idet

"the engineer or the technologist approaches institutions rationally as means to serve certain ends, and as a technologist he judges them wholly according to their appropriateness, efficiency, simplicity etc." (Popper 1995/1945, p. 24).

Men som Hirdman (1992) med velvagt ironi påpeger, så er der tale om en ganske udemokratisk opfattelse af den videnskabelige videns politiske og sociale overlegenhed⁵.

"What unites the "utopian" social engineer and the "step by step" social engineer, to use Popper's definition, is undeniably that their revolutionizing actions are legitimated in terms of a *scientific* title. If the difference between the two is no longer crystal clear, it is because we ourselves are somewhat skeptical about the scientific purity of the social and human sciences."

Den tætte kobling til brugere, politikere og andre aftagere af evalueringer var som Hirdman påpeger, ikke baseret på en klar demokratisk tradition, men på en ganske elitær top-down implementering. Det varede dog ikke længe før der blev introducerede andre modeller for, hvorledes evalueringer

⁵ Yvonne Hirdman har vist hvorledes "social engineering" synspunkter havde en fremtrædende plads i svenske socialdemokraters forståelse af hvorledes man med befolknings- og socialpolitiske midler skulle forbedre samfundet i 1930erne, hvor de mest fremtrædende personer i denne politik var Gunnar og Alva Myrdal. Tidsmæssigt ligger denne praktiske politikudvikling af social engineering en del år før Popper udgiver sin klassiske bog. Hirdman bruger med en vis ironi udtrykket "democratic engineer" for at sammenfatte udviklingen fra en social engineer's praktiske anvendelse af videnskabelig viden til den nye situation, hvor videnskabelige viden anvendes og udvikles ud fra en systematisk planlægningsstilgang, dvs. en udtalt ekspertkultur som forsøges sammenkædet med en demokratisk politisk forståelse som det var fremtædende i det svenske socialdemokratis forestillinger om et planlagt samfund (Hirdman 1992).

og ikke mindst evalueringsforskeren skulle opfatte sig rolle og funktion i forhold til de programmer og projekter, der skulle evalueres. To modeller er her centrale, oplysnings- og interaktionsmodellerne⁶.

Oplysnings- og interaktionsmodellerne har rod i nogle af de mest kendte studier af den politiske anvendelse af evalueringer, som Carol Weiss har været pioner bag (Weiss 1979, Weiss og Bucuvalas 1980), og som siden er ført videre i en række anvendelsesstudier (Weiss 1999). Weiss og Bucuvalas peger på, at selvom grænserne for denne viden er klare, så kan disse grænser overskrides videnskabelig viden er omgivet af en særlig autoritet.

"Social science research provides only one component of knowledge. Yet, as we have seen in this inquiry, it can carry a special authoritativeness, in part because it is presumed to have the systematic and objective hallmark of science and in part because its concepts provide a powerfull vocabulary that shapes the definition of issues."(s. 276)

Hvad Weiss kunne vise baseret på empiriske studier fulgte helt i tråd med den kritik af rationelle modeller for anvendelse af samfundsvidenskabelig viden, som fremtrædende organisationsforskere som Charles Lindblom og Herbert Simon⁷ havde fremsat i forhold til managementlitteraturen. Fælles for kritikken var en problematisering af den enkle rationalitetsmodel for anvendelse af samfundsvidenskabelig teori, som lå bag social engineering modellen. Oplysnings- og interaktionsmodellerne for anvendelse af evalueringer fremhæver vigtigheden af at evalueringsresultaterne, den specifikke viden som er indsamlet i en evaluering, formidles bredt ud til offentligheden og ikke blot videregives til de umiddelbare opdragsgivere. Oplysningsmodellen har tydeligt hentet inspiration fra Habermas' teoretiske arbejde med at udvikle et begreb om en moderne borgerlig offentlighed (Habermas (1972, 1982) og modellen rummer en normativ 'forpligtelse' for evalueringsforskeren til at informere og formidle bredt. Interaktionsmodellen retter sig også mod oplysning og information, men modsat oplysningsmodellen fremhæver interaktionsmodellen specielt samspillet mellem politikere og evalueringsforskere med hensyn til implementering af resultater og fokuserer på den måde langt mere direkte på rådgivning, politikudvikling og beslutningstagning end oplysningsmodellen.

⁶ Se nærmere hos Hansson 1999 og Bryant 1995.

⁷ Begge har i adskillige publikationer fremført en radikal kritik af forestiller om linær rationalitet i beslutningsproceeser i organisationer og i det økonomiske liv, Lindblom i den klassiske artikel fra 1959, The Science of "Muddling Through" fulgt op med bogen Usable Knowledge fra 1979 og Herbert Simon i utallige bøger, f. eks. Administrative behaviour fra 1961 og Economics, Bounded Rationality and the Cognitive Revolution fra 1992.

Med kraftig inspiration fra begrænsninger i den ovennævnte habermasianske kritik af kapitalismens destruktion af den borgerlige offentlighed, er der vokset en model frem, som langt mere direkte end de ovenfor nævnte søger at inddrage de forskellige aktører og andre interessenter i og omkring en evaluering i et procesforløb. Denne sidste og nyeste model for samspillet mellem anvendt samfundsforskning er dialogmodellen (Bryant 1995), som ligger ganske tæt ved en mere kendt model for medinddragelse i forskningsprocessen, aktionsforskning⁸, og vel nærmest kan beskrives som en videreudvikling heraf. Denne model begrænser sig ikke til spørgsmål om formidling af viden, men inddrager også samspillet mellem evaluator og de evaluerede (lægpersoner) i vidensproduktionen. Endelig er der i de seneste år fremkommet en relativ ny variant af denne deltagerorienterede tilgang i evalueringsverdenen, kendt under betegnelsen 'empowerment evaluation', som ikke mindst i USA har fået kraftig fremvækst de seneste år.

⁸ Aktionsforskning kan føres tilbage til Kurt Lewin i 1940erne, se den programmatisk beskrivelse i Lewin (1945), men har især udviklet sig i Norden i efterkrigstiden, se Toulmin og Gustavsen (1996).

Evalueringer og NPM-reformpolitik

Et andet centralt kendetegn ved de moderne markedsøkonomiske samfunds styring og regulering er den tiltagende anvendelse af eksperter og ekspertrådgivning (Beck, Lash og Giddens 1994). Dette kan ligeledes betragtes som en konsekvens af den tiltagende kompleksitet og gensidige afhængighed, der kendetegner alle sider af disse moderne eller sen-moderne samfund og som derfor indgår som en væsentlig forudsætning for alle forsøg på styring og regulering. Kompleksiteten kombineret med hastige sociale forandringer, der skaber vanskeligheder med at frembringe sociale data, der er gyldige over længere tid (Beck 1997), er forhold, der gør sig gældende ikke mindst ved den åbenbare og ofte politisk offensive brug af videnskabelige resultater og eksperter. Denne 'politiske' brug af eksperter medfører, at hvad der i tidligere tider højst ville have været videnskabsinterne diskussioner og konflikter nu kommer ud i fuld offentlig belysning. Ekspertanvendelsen er en meget modsætningsfyldt proces (Beck 1994, Stehr 1994, Aagaard Nielsen 1997), som også er præget af førnævnte orenteringsusikkerhed og fremviser sprækker og revner i det ikke længere så entydige fundament for anvendelse af videnskabelig viden i takt med ændringer og skærper af samfundsmæssige modsætninger og interessekonflikter.⁹⁾

Evalueringsrådets historie fra traditionel anvendt samfundsforskning til en videnproduktion, hvor deltagere indgår aktivt i frembringelse og brug af viden som i dialogmodellen og aktionsforskning, har tydeliggjort, hvor kompleks denne nye form for kombineret produktion og anvendelse af viden i realiteten er. Beck og Bonss' (1989) har i den forbindelse en vigtig pointe fra deres omfattende analyse af betingelserne for anvendt samfundsvidenskab og evalueringer i det moderne samfund. Samfundets komplekse strukturer og hastige ændringer indebærer, at grænser og roller mellem samfundsforskere og brugere af forskningsresultater ikke længere er så klare og entydige, som det i mange år har set ud. De formulerer i nedenstående citat meget klart, hvor grundlæggende forandret produktion og anvendelse af samfundsvidenskabelig viden er blevet:

"stillet overfor den institutionelle, professionelle og hverdagsmæssige videnskabeliggørelse kan anvendelsen af

⁹ Der tænkes her naturligvis først og fremmest på de seneste års voldsomme videnskabssociologiske kritik af objektivitet og neutralitet i videnssammenhænge, som også kan ses som en videreførelse af den tidligere positivismekritik.

samfundsvidenskabelig viden ikke længere tænkes deduktivt ud fra en 'missionærholdning'. De 'sociologiløse indbyggere' er for længst blevet 'missioneret', visse steder begynder de selv at producere videnskab... ." (Beck and Bonss 1989)¹⁰

Denne konklusion er også på linie med hovedteserne i den efterhånden ganske omfattende litteratur om fremkomsten af nye former for videnskabelig produktion af viden, hvor videnskabssociologien har introduceret begreber som Mode 1 og Mode 2 viden¹¹ (Gibbons 1994, Nowotny 2001) eller 'Triple Helix' (Etzkowitz og Leydesdorff 2000), som netop betegner en opløsning af traditionelle roller og former for arbejdsdeling mellem producenter af videnskabelig viden og brugere.

Det er en gennemgående tese i denne undersøgelse af interne evalueringer, at en af de vigtigste kendetegn ved den New Public Management - dominerede udvikling af den offentlige sektors reguleringspolitik, som Danmark såvel som en række andre OECD lande har gennemlevet, er en omfattende brug af alle mulige former for evalueringer og tilsvarende vurderings- og kontrolsystemer til styring og regulering af offentlige institutioner og aktiviteter.

Den overordnede samfundsmæssige baggrund for denne udvikling er skitseret ovenfor og består af en sammenkædning af grundlæggende ændringer i forholdet mellem videnskab og samfund, udvikling af mere og mere komplekse reguleringsystemer og måske vigtigst, hastige ændringer i samfundenes værdigrundlag (individualisering, se Beck 1999) – alt sammenholdt med voksende økonomiske problemer for den offentlige sektori en globaliseret økonomi.

I første omfang fremstod den nødvendige modernisering af den offentlige sektor som et spørgsmål om New Public Management modernisering af den offentlige sektor, med privatisering, og deregulering i fokus - fremkaldt af krav om økonomiske besparelser og større ydelser. Under betegnelsen NPM er denne udvikling blevet en sammenfattende betegnelse for de sidste 10 års modernisering af den offentlige sektor i de vestlige OECD-lande, ofte udviklet i ganske ideologiske former, som det f.eks. fremgår af titlen på Clinton-regeringens rapport om modernisering af forbundsregeringen, *Creating a Government that Works Better and Costs Less* (Peters m.fl. 1993).

Nogle centrale træk ved NPM er de gennemgående forsøg på at opløse traditionelle bureaukratiske organisationsformer, kombineret med oprettelse

¹⁰ Zum Strukturwandel von Sozialwissenschaft und Praxis, min oversættelse.

¹¹ Se også diskussionen i Hansson 2002

af selvstændige enheder, i form af decentralisering kombineret med kontrol. Endvidere introduktion af markeds- og kvasi-markedsmekanismer, fremhævelse af servicekvalitet og brugerorientering, krav til medarbejderne om at arbejde efter fastsatte produktionsmål, indikatorer og resultater (performance management), ændring af ansættelsesvilkårene for offentlig ansatte til tidsbegrænset kontraktansættelse, produktivitetsløn og lokal fastsættelse af løn og arbejdsvilkår,

Pollit har i flere analyser har undersøgt hovedlinierne i denne modernisering af de vestlige staters offentlige sektorer og påpeger i en kritisk gennemgang af disse udviklingstendenser (1995, 134), at et gennemgående karakteristika ved NPM er, at trods politikens klare orientering mod 'accountability', dvs. systematisk opfølgning og tilpasning, har disse forskellige tiltag aldrig været underkastet systematiske evalueringer som sammenhængende politik. Der er ifølge Pollit således ikke ud fra systematiske empiriske studier grundlag for at tage stilling til spørgsmålet om disse ofte meget omfattende politiske tiltag fremmer den modernisering og omkostningsreduktion i den offentlige sektor, som de antages at gøre. En del større evalueringer har været gennemført, ikke mindst af OECD, men ingen af disse eksterne evalueringer tager ifølge Pollit fat på de store vanskeligheder, ikke mindst af metodisk karakter, der er forbundet med at evaluere så store og komplekse politiske forandringer, som NPM er - (mangesidede samtidige forandringer, samtidige ændringer i ydre forhold - ikke mindst betydningen af internationaliseringen af de nationale økonomier via globaliseringsprocesser, ændrede politiske rammer, usikre evalueringskriterier og målsætninger mm.). Pollits konklusion på gennemgangen af en række store evalueringer af NPM er, at de resultater der er fremkommet i disse evalueringer er yderst skrøbelige pga. disse evalueringers metodiske begrænsninger og meget usikre eller upålidelige data mht. fastlæggelse af målopfyldelser, bedømmelseskriterier og beregning af omkostninger ol.. Den politiske begejstring for NPM-redskaberne bygger således ikke på et solidt grundlag af viden om hvad der fungerer og hvad der ikke fungerer. Men samtidig noterer Pollit også, at denne manglende viden ikke har haft nogen nævneværdig indflydelse på udbredelse af NPM. Den manglende anvendelseskobling mellem viden fra nævnte evalueringer og den videre politikudvikling af NPM er næppe overraskende – dette har været gængs viden siden Weiss og Bucuvalas' (1980) klassiske studie af den politiske (mangel på) anvendelse af forskningsresultater¹².

Pollits analyse nærmest har karakter af at være en metaevaluering, som ser på omfanget og med hvilke resultater evalueringsforskningen er blevet anvendt til at undersøge og vurdere NPM-redskaber og metoder, dvs.

¹² Angående anvendelse af evalueringsresultater se Weiss (1999).

vurderer NPM ud fra en overordnet sammenfatning af evalueringer af moderniseringsplaner og forsøg. Men der findes der også en anden, mere mikroorienteret anvendelse af evaluering i forbindelse med NPM, brug af evalueringer som et internt instrument i moderniseringen af offentlige organisationer og institutioner.

Et af de mest karakteristiske træk ved NPM i forhold til tidligere politiske strategier for udvikling af den offentlige sektor er den meget omfattende og i nogle sammenhænge systematiske anvendelse af evalueringsmetoder, af refleksive kontrol og reguleringssystemer, det som Power (1997) sammenfattende betegner som 'the audit society'. Overgangen fra styring af offentlige institutioner og organisationer via generelle politiske målsætninger og økonomisk rammestyring til en moderniseringsproces, hvor udviklingen er styret af et quasi-marked¹³⁾ i form af brugerefterspørgsel og resultaterne skal dokumenteres i form af produktbeskrivelser, kontrakter (contract management) og opfyldelse af kvalitetsmål, har krævet ibrugtagning af en række metoder og teknikker til organisationsudvikling og styring af produkterne kvalitet, som tidligere kun yderst sjældent har været anvendt af offentlige institutioner og organisationer. I denne proces er det tydeligt at se, hvordan man forsøger at opfinde en mekanisme for markedsstyring af de statslige 'markeder' ved at forsøge at beskrive disse som pseudomarkeder. Disse markeder mangler en 'pengeform', dvs. en almen målestok, der på samme måde som markedets kvalitets-kvantitetsmålestok, penge og priser, kan gøre sammenligninger af både kvantitet og kvalitet mulige.

Mens NPM strategien med privatisering og deregulering viste sig at have

¹³ Denne betegnelse er i virkeligheden ikke helt præcis. Et helt afgørende problem i hele NPM-filosofien er forestillingen om, at metoderne der har udviklet og effektiviseret den private markedsøkonomi, baseret på konkurrence og pris, uden større vanskeligheder kan

anvendes i den offentlige sektor. Enhver der husker den marxistiske statsdiskussion fra 70'erne ved, at dette er en ganske problematisk antagelse af den grund, at langt de fleste af de statslige opgaver netop er fremkommet som et resultat af at markedet/markedsøkonomien af historiske, principielle (uddannelse, socialpolitik) eller praktiske (infrastruktur) forhold ikke eller kun utilstrækkeligt har kunnet løse disse opgaver, som derfor er blevet udskilt fra markedsøkonomien, statsliggjort.

De seneste års globalisering og internationalisering af kapitalen har endvidere gjort pengeformen i hidtil uhørt omfang fri fra sine fysiske, stofflige (brugsværdien) og historiske bindinger, således at den som finans- og pengekapital i hidtil uhørt omfang har frigjort sig fra nationalstatens reguleringsforsøg -og muligheder (Altvater og Meinkopf 1996). På det frie internationale marked fungerer pengeformen mere og mere uafhængig af lokale forhold og begrænsninger. Men de statslige aktiviteter har netop ikke en sådan af de stofflige former uafhængig ækvivalent som pengeformen, der 'bag om ryggen på agenterne' kan regulere og styre udviklingen. De er og vil altid være bundet til de lokale omgivelser, de er stofflige, yderst stationære og meget konkrete.

begrænset rækkevidde, jf. Pollits kritik, så efterlod den omfattende brug af evalueringer en arv, som i første omgang blot fremtrådte som et teknisk hjælperedskab i udvikling og omlægning af offentlige institutioner, den systematiske brug af evalueringer som kontrol og styringsredskab, men som efter en periode viste sig at blive til noget, der er mere end et teknisk hjælpemiddel. Evalueringer er kommet for at blive og har i denne proces fået en langt mere og mere selvstændig rolle og funktion – de er ikke længere blot et redskab til justering og feedback – men fremstår mere og mere som den nye centrale metode for organisationsudvikling i store såvel som små offentlige organisationer.

De usynlige interne evalueringer i evalueringslitteraturen

I den foran skitserede udvikling tegner der sig flere spor eller linier i offentlige organisationers brug af evalueringer. Brugen af evalueringer kan opdeles i to principielle retninger afhængig af den måde, evalueringer indgår i organisationer på. Der tegner sig i dag et billede af de mange eksterne evalueringer, som i stort omfang rummer kontrolfunktioner, som ofte er en del af selve programudviklingen og som kan fungere som metode til sammenfatning af udviklingen på et område (f. eks. DANIDA-evalueringerne, uddannelsesevalueringerne, institutions- og områdeevalueringer). Disse ofte store og professionelt styrede evalueringer udføres af professionelle, eksterne evaluører, fra forskningsinstitutioner og konsulentfirmaer. Disse evalueringer kan antage en række forskellige former, hvoraf de vigtigste ifølge Vedung (1991, 43) er:

- målopfyldelsesevaluering,
- sideeffektevaluering,
- klientorienteret evaluering,
- stakeholder evaluering,
- professionel evaluering (peer review),
- professionel evaluering (selv evaluering),
- produktivitetsevaluering,
- cost-effectiveness evaluering
- cost-benefit evaluering

Evalueringer har længe været kendt - især fra udlandet, USA - som en metode til undersøgelse og vurdering af især offentlige projekter og programmets udvikling og opfyldelse af målsætninger¹⁴. Bruger-og

¹⁴ At dette stadig er yderst aktuelt kan ses af følgende kommentar til Bush-administrationens nyeste budgetforslag: "The President's '06 budget proposal delivers on a promise made by the Administration to subject federal programs to rigorous performance evaluations and tie agency budgets to those evaluations. "Performance budgeting is finally paying dividends to taxpayers," notes Performance Institute President Carl DeMaio through the continued use of the PART. The '06 budget contains PART "scores" for over 600 programs this year and "most importantly, the budget request uses performance information to terminate or make targeted reforms in poor-performing programs while expanding programs with solid results," DeMaio says. The FY '06 federal budget proposed by President Bush proposes some of the toughest spending restraint in years. The budget proposal will attempt to make a dent in the deficit by holding overall discretionary spending to a 2.1% increase and by cutting or significantly scaling back more than 150 federal programs deemed ineffective, obsolete or duplicative." The Performance Institute, White Paper Analysis of President's '06 Budget <http://www.performanceweb.org/>

interessentundersøgelser har jo været en del af evalueringsredskaberne i en årrække. Men hvor mere traditionelle evalueringer typisk har været udført ved brug af eksterne, professionelle evaluatore så er en mere og mere fremtrædende evalueringsform i dag de *interne evalueringer*.

For Danmarks vedkommende ses denne udvikling tydeligt i Finansministeriets forskellige offensiver for modernisering af de offentlige institutioner og organisationer gennem de sidste 10 år som en dansk udgave af NPM. Det bliver ikke mindre interessant i det øjeblik den officielle politik ændrer sine signaler og nu ikke længere kræver (eller forventer) samme omfang af traditionelle evalueringer som før, men derimod propagerer for, at de offentlige organisationer begynder at bruge evalueringsværktøjerne i en løbende intern udviklingsproces, nemlig som interne evalueringer.¹⁵

Ser man på førende lærebøger i evaluering fra de senere år er det tydeligt, at begrebet interne evalueringer nu er dukket op og er blevet en mere og mere selvstændig form for evaluering, som nu kan og skal behandles som et område med særlige træk og problemer, om end det stadig er begrænset hvor omfattende interne evalueringer behandles. På mange måder følger interne evalueringer ikke samme traditioner, praksis'er og metodologier som de professionelle eller eksterne evalueringer har udviklet over en årrække (Sonnichsen 1994). I den nyeste udgave af en af de mest udbredte lærebøger i praktisk evaluering, Patton's Utilization-Focused Evaluation (2002) bruges der mere end 6 sider på at diskutere forskelle og ligheder mellem eksterne og interne evaluatorens roller. En mere dybtgående behandling får interne evalueringer hos Love (1991) som ser det forhold, at den interne evaluator er en del af organisationen som en særlig styrke ved interne evalueringer trods de mange mulige problemer der kan udspringe af samme forhold.

En særlig problematik gør sig gældende for interne evalueringer i forhold til andre former for evalueringer der er gennemført med eksterne evaluatore. Det er det usynlighedsforhold som gælder for interne evalueringer alene af den grund, at de gennemføres internt i organisationen og yderst sjældent kommunikeres ud til eksterne sammenhænge, det være sig til andre organisationer, til faglige arrangementer som konferencer og til offentligheden generelt i både medier og i faglige tidsskrifter. Dette usynlighedsforhold er med til at gøre det ganske kompliceret at foretage

¹⁵ I den sammenhæng har finansministeriet udgivet en række håndbøger og praktiske vejledninger om anvendelse af metoder til bruger- og personaleundersøgelser. Først og fremmest spørgeskemametoder er faldet i finansministeriet smag. Denne tilgang til anvendelse af især kvantitative samfundsvidenskabelige metoder til intern evaluering har vundet stor udbredelse i offentlige organisationer.

mere systematiske undersøgelser af omfang, kvalitet og udvikling i interne organisationer i eksempelvis staten. Foss Hansen (2003) har gennemført en større spørgeskema-baseret undersøgelse af evalueringsanvendelsen i den statslige sektor og beskriver udbredelse, organisering, metoder og tilgange samt en række andre forhold i den danske centraladministration. Der er dog næppe tvivl om, at der er et stort gråzoneområde bestående af interne evalueringer, som ikke er medtaget i denne undersøgelse. Det kan være mindre interne evalueringer, som udføres decentralt som led i løbende udviklingsprogrammer, evalueringer som tager form (og ofte også navn) efter deres formål, f. eks. personaleundersøgelser eller tilfredsundersøgelser eller som led i beslutninger, hvor den interne evaluering indgår som del af beslutningsgrundlaget, f. eks. ved tildeling af forskningsmidler i forbindelse med store programmer.

Det er det forhold, at interne evalueringer i langt mindre omfang indgår i og er en del af den politiske offentlighed, for slet ikke at tale om evalueringsforskningens faglige offentlighed, understøtter denne usynliggørelse. Interne evalueringer offentliggøres sjældent og interne evalueringer er også sjældent professionelt uddannede indenfor evalueringsfeltet og deltager i evalueringsfaglige aktiviteter og er tilsvarende yderst sjældent uafhængige af den evaluerede organisation. Interne evalueringer bliver af disse grunde også sjældent konfronteret med de diskussioner om evaluatorers politiske og moralske dilemmaer og andre faglige spørgsmål, som den professionelle litteratur i stigende grad er optaget af (se f. ex. Schwandt 2003).

Hvilke konsekvenser dette kan få for både evalueringsforskningen og for interne evalueringer har i de senere år været diskuteret heftigt blandt professionelle evalueringer i midten af 90'erne. House (1993) er en af de evalueringsforskere, som gennem årene har rejst en række fundamentale spørgsmål om evaluatorers uafhængighed i interne evalueringer, herunder spørgsmålet om organisatoriske magtforhold vil tendere til at underlægge resultater af interne evalueringer andre overordnede organisationsinteresser, eller blot betyder, at de bruges som rent symbolske handlinger (Albæk 1996). En undersøgelse af den interne evalueringsfunktion i en stor amerikansk offentlig organisation, National Science Foundation (House, Haug og Norris 1966), har vist, at en synlig og selvstændig intern evalueringskultur og en organisationsstruktur som giver uafhængighed til den interne evalueringsenhed i visse tilfælde kan sikre både uafhængighed og kvalitet i interne evalueringer. Men en afgørende forudsætning er, at organisationen har en tilpas størrelse, der muliggør eksistensen og udviklingen af en selvstændig og professionaliseret organisatorisk enhed, som i ovennævnte undersøgelse og måske også, at det der evalueres kræver en særlig professionel kompetence hos evaluatorerne, som i dette tilfælde, hvor der var tale om evaluering af forskningsprojekter.

Blandt problemer for interne evalueringer peger Sonnichsen (1994) på, at er

den interne evaluerings usynlighed ofte gælder både i organisationen og udenfor. Den interne usynlighed kan hænge sammen med at interne evalueringer ofte er tæt sammenvævet med andre elementer og redskaber til organisationsudvikling, eller har fået tildelt en rituel rolle der isolerer evalueringen fra organisationens udvikling (Dahler-Larsen 1998, Albæk 1996). Det sidstnævnte, den eksterne usynlighed, hænger som nævnt nøje sammen med den manglende tradition og ikke mindst en ofte politisk-organisatorisk interesse for ikke at offentliggøre resultater af interne evalueringer. Uden offentliggørelse kan interne evalueringer kun i meget begrænset omfang indgå i en eksisterende offentlig og faglig debat. En åbenlys konsekvens heraf er, at der hersker der en udbredt mangel på viden om interne evalueringer omfang og udbredelse. Foss Hansen (2003, 143) fandt, at ca. 40% af de undersøgte statslige evalueringer var gennemført af "andre evalueringsprofiler", dvs. ikke af de eksterne professionelle.

Den positive side af at bruge interne evalueringer til at udvikle og ændre organisationer, til organisatorisk læring, fremhæves bl.a. Leeuw, Rist og Sonnichsen (1994) og Love (1991), som peger på, at en systematisk brug af evalueringsredskaber positivt kan bidrage til at en organisation selv kan gennemføre moderniseringsprocesser, som måske ellers ville være påtvunget udefra med højere omkostninger. Et godt eksempel på brug af evalueringsredskaber i interne evalueringer giver Peiperl (1999), som har undersøgt konsekvenserne af at indføre brug af peer evaluering, en variant af peer review'et kendt fra forskningsevalueringer, som metode til personalevurdering i stedet for mere traditionelle evalueringssystemer baseret på skemavurderinger med skalaer osv., metoder der ofte fører til inflation i vurderingen. En af konsekvenserne af denne peer evaluering var en mere horisontal og 'demokratisk' evaluering, i og med at alle blev inddraget i processen.

Interne evalueringer og organisationsudvikling

Udover disse problemer som især vedrører interne evalueringers relationer til evalueringsforskningen og berører forhold om metodologier, faglig udveksling og kritik, brug og misbrug af resultater mm., så rummer interne evalueringer også en funktion som udviklingsredskab for organisationer. Her viser spørgsmålet om omfanget af ledelsesstyring sig som et tydeligt dilemma. Interne evalueringer kan være et ledelsesredskab, hvis anvendelse på forhånd er begrænset til at producere viden om visse og kun visse aspekter af organisationen som f. eks. økonomistyringsredskaber, eller enkelte projekter. Der er også den mulighed, at der er et klart moment af både medarbejderindflydelse i form af brug af medarbejdernes opsamlede erfaringer og 'tavse viden' til en aktiv deltagelse i udformningen af organisationens udvikling og i det hele taget til at være et redskab til organisatorisk læring og udvikling, der ikke på forhånd er begrænset til kun

at berøre på forhånd definerede problemstillinger og om råder.

For den interne evaluering som organisatorisk system til produktion af viden om organisationen handler i denne forbindelse om hvilke typer viden der skal indsamles, hvordan den skal indsamles, hvordan den skal tolkes og anvendes, med andre ord, betingelserne for vidensproduktion og -anvendelse i organisationerne.

Interne evalueringer er på den ene side 'fanget' i en anvendt samfundsvidenskabelig ekspertkultur, da evalueringer har en historisk oprindelse i den anvendt samfundsforskning. Det kan let føre til, at man begrænser sig til at indsamle og bearbejde den viden, der kan frembringes ved brug af traditionelle videnskabelige metoder, og det er oftest kvantificeret og objektiviseret information (Porter 1992, 1995). På den anden side er interne evalueringer i høj grad rettet mod organisationen, evt. de dele af en organisation, hvor der foregår forandringsprocesser og her kommer traditionelle videnskabelige tilgange med krav om objektivitet og gentagelighed oftest til kort. Kravet om at undersøge effekter af politiske programmer ved brug af kontrollerede eksperimenter holder sjældent i en organisationsverden – også selvom en variant heraf var dominerende i evalueringsforskningens tidlige periode ikke mindst i Donald Campbells arbejder (1968, 1988).

Breslau (1997) har i en Bourdieus' sk inspireret analyse undersøgt et på flere måder meget interessant forløb i ændringer i den rolle og placering, som evalueringer har haft amerikansk arbejdsmarkedspolitik. Gennem en årrække havde der udviklet sig nogle interne evalueringsprocedurer, baseret på case studier og brug af quasi-eksperimentelle metoder med klar inspiration fra Campbell (1968), som frembragte evalueringsinformation om en række beskæftigelsesprogrammer finansieret af arbejdsministeriet (DOL). Hele dette evalueringsforløb var ret lidt formaliseret og i høj grad baseret på et løbende uformelt samarbejde mellem konsulenter og embedsmænd (fra DOL). Aktørerne var her evalueringskonsulenter på kontrakt og embedsmænd. Hjulpet på vej af politiske ændringer, herunder et ministerskift, og ikke mindst af de voksende politiske krav om kvantificerbare evalueringsresultater, eller som det i de seneste år er blevet kaldt 'evidence-based decision making', lykkedes det en gruppe universitetsøkonomer med økonometriske kompetencer at overtage evalueringen af beskæftigelsespolitikken. Dette introducerede en helt anden ekstern evalueringsmodel, baseret på økonometriske modeller, dvs. på aggregerede kvantitative data, skulle måle effekten af de forskellige beskæftigelsesprogrammer. Breslau analyserer dette forløb som en kamp mellem to ekspertkulturer, mellem konsulenter og universitetsøkonomer, da der ikke var uenighed om hvad der var den centrale indikator på succes, nemlig højere løn efter jobtræning. Men det er lige så relevant at se eksemplet som udtryk for et skift fra en integreret intern evalueringsmodel

til en helt ekstern, hvor skiftet også markerer grundlæggende forskelle i de data og evaluerings- og beregningsmetoder, der bruges i evalueringerne. Eksemplet peger også på at indsamlingsmetoder og modeller ikke helt tilfældigt kan anvendes i den ene henholdsvis den anden form for evaluering, men at indsamling og behandling af data kan producere meget forskellige former for viden til brug i beslutningsprocesser. Men den mere abstrakte og kvantitative dataproduktion til brug for evalueringerne, som skiftet til universitetsøkonomer indebærer, behøver ifølge Breslau ikke at være en endelig udvikling, idet selve det forhold, at de første former for evalueringer kunne foregå så længe de gjorde det tegner en åbning af kontrol med vidensproduktionen. "Contract shops have become the institutional base for a break with academic economists' conception of what counts as scientific knowledge." (Breslau 1997)

I det meste af evalueringsslitteratur diskuteres interne evalueringer overfor eksterne som to afgrænselige og mht. organisatoriske og praktiske forhold meget forskellige former for evalueringer, mens der sjældent stilles spørgsmål om der foregår den samme eller en forskellige vidensproduktion de to steder. Patton (1996) diskuterer i sin meget udbredte lærebog forskellen som spørgsmålet om hvilken indflydelse på godt og ondt forskellen i organisatoriske tilknytning har og House (1993) diskuterer mere principielt problemerne om evalueringernes principielle uafhængighed, og tager senere (House m.fl. 1996) dette op i relation til en case med intern evaluering i en meget stor organisation. Den generelle tendens i evalueringsforskningen må ud fra dette antages at være, at videnskabelig-metodiske 'kerne' og tilhørende krav også må gælde for interne evalueringer. Bag denne antagelse om en 'ligestilling' af in- og eksterne evalueringer ligger nogle implicite antagelser om interne evalueringers organisatoriske tilhørsforhold, dvs. hierarkiske indplacering i en organisation. Et par undersøgelser af interne evalueringer i USA har dog sat spørgsmålstegn ved hvor betydningsfuld og reel ovennævnte forskelle er. Sonnichsen (1994) og House m.fl. (1996) har påvist at interne evalueringer i store offentlige organisationer i stor udstrækning kan gennemføre evalueringer med samme grad af frihed hhv. kontrol som eksterne evalueringer. Disse undersøgelser tager udgangspunkt i offentlige organisationer af en sådan størrelse (FBI og NSC)¹⁶, der sjældent kan findes i lande udenfor USA. Det vigtigste forhold der direkte udspringer af organisationernes størrelse er uden tvivl, at de interne evalueringer udføres af relativt store og selvstændige enheder i organisationen, bemandede med fuldtidsevaluere med en ofte professionel evaluering baggrund (House m.fl. 1996).

¹⁶ Federal Bureau of Investigations og National Science Council.

I europæiske og ikke mindst danske sammenhænge findes så store organisationer sjældent. EU's interne evalueringsenheder kan nok i størrelse sammenlignes med de amerikanske eksempler, men er pga. EU's administrative tradition for lukkethed meget vanskelige at få indblik i. I Danmark har undervisningsministeriets evalueringsinstitut udført evalueringer af uddannelserne fra folkeskoler til universiteterne. Sådanne evalueringer kan dog næppe betegnes som interne, selvom der indgår elementer af selvevaluering, da den samlede evaluering udfærdiges af evalueringsinstituttets professionelle evaluatore og er offentlig tilgængelig. I forhold til den samlede uddannelsessektor kan man måske betragte dem som interne evalueringer, men dels indgår der en vis anvendelse af eksterne eksperter og dels fungerer og opfattes disse evalueringer af langt de fleste som eksterne overfor de evaluerede universiteter og institutter. Det sidste er der ingen tvivl om efter de seneste års heftige debatter. I den svenske undervisningssektor kan man finde en tilsvarende central evalueringsenhed. Heroverfor står de om man så må sige regulære interne evalueringer, som i NPM's kølvand i stigende omfang er blevet en del af hverdagen i danske offentlige institutioner og organisationer. De er mere og mere direkte blevet en integreret del af organisationens/institutionens egen udviklingsproces. Disse evalueringer udføres typisk af allerede ansat personale uden at disse nødvendigvis har en særlig baggrund i evalueringsundersøgelser¹⁷ og som ved siden af dette udfører andre arbejdsopgaver. Sådanne evalueringer omfatter typisk en lang række mindre undersøgelser, brugerundersøgelser, tilfredshedsundersøgelse ol., og inddrager ofte allerede eksisterende eller nyproduceret produktions- og aktivitetsstatistik, men kan også udvides til mere omfattende evalueringer.

I nedenstående tabel har jeg forsøgsvis søgt at opstille de vigtigste forskelle på eksterne og interne evalueringer, vel vidende at der i den konkrete praksis i organisationer eksisterer mange varianter og mellemformer, eksempelvis når interne evaluatore forbereder materiale for eksterne evalueringer. Hovedsigtet med tabellen er at belyse spørgsmålet om de forskellige måde disse evalueringer fungerer på i organisationer – og ikke mindst hvilken mere indirekte virkning og indflydelse i organisationen udbredelsen af interne evalueringer har fået.

¹⁷ Systematiske uddannelsesforløb i evalueringsteori og metode har været en sjældenhed i det danske universitetssystem.

TABEL 1**Centrale forskelle på eksterne og interne evalueringer:**

	Eksterne evalueringer	Interne evalueringer
Evaluatorer	Professionelle, eksterne evaluatorer	Organisationens egne medarbejdere
Stor afstand til organisationen	Traditionel ekstern evaluering rettet mod proces og resultater	Selvstændig enhed i organisationen – ofte med egen kultur
Lille afstand til organisationen	Typisk udviklingskonsulenter ol.	Udgør selve kernefunktionen i interne evalueringer – som udføres på en række områder og niveauer i organisationen
Governance/social kontrol funktion	Evalueringer udføres for tredjepart (myndighed) med sigte på kontrol og dokumentation af processer og forløb	Evalueringer fungerer som redskab til organisationsudvikling, en særlig feedback mekanisme og som redskab til intern organisationspolitik. Meget få erfaringer med mere langsigtede konsekvenser

De fordele en organisation får ved at bruge de interne evalueringer er den tætte kobling til organisationens hverdag og muligheden for brug af og direkte inddragelse af medarbejdererfaringer og anden 'tavs viden' i organisationen. Endvidere kan det også have en positiv betydning for medarbejdernes holdninger til organisationen generelt, at de kan få mulighed for at deltage i arbejdet med at præge både organisations hverdag men også udviklingsretning.

Bagdelen er, at organisationen som helhed, men først og fremmest ledelse eller ledere på mellemniveauer, let får en uforholdsmæssig stor direkte indflydelse på udformning af evalueringerne. Det kan betyde at fokus i evalueringer, temaer og problemstillinger som kan tages op og vægtningen i analysen let kan komme til at virke som en slags selvopfyldende profeti i forhold til de af ledelsen ved siden af fastsatte mål for organisationen. Tæt ledelsesdeltagelse og dominans kan ligeledes blokere for at 'erfaringer fra

gulvet' og 'tavs viden' om dysfunktioner og interne vanskeligheder i organisationen kan indgå som relevant og ligestillet viden i evalueringen. Desuden kan det være et reelt problem, at der ofte ligger forventninger om brug af traditionelle samfundsvidenskabelig metodikker og helst metoder der kan producere kvantificerbare resultater (Porter 95), som kan indgå i en 'evidence-based decisionmaking' (Shadish m.fl. 2005). Metodernes krav til teknisk viden og tilsvarende til at kunne analysere de objektiverende resultater de frembringer som tabeller og enkle statistiske analyser, udelukker let metodisk utrænede medarbejdere fra at være aktive i evalueringsprocessen. Brug af kvalitative metoder er i sig selv ikke en løsning, da disse er forudsætningskrævende på andre vilkår. Dette dilemma gør sig gældende for alle typer evalueringer, interne såvel som eksterne, men da de eksterne i et vist omfang indgår i en evalueringsoffentlighed og dermed kan udsættes for fagfolks kritiske blikke, er der så meget mere grund til at antage at dilemmaet især gør sig gældende for interne evalueringer. Problemets kerne er, at ønsket om at producere objektiv, kvantificerbar viden med den interne og eksterne troværdighed, som ofte tillægges denne form for viden, kommer i konflikt med et andet ønske til interne evalueringer, nemlig at de skal forsøge at 'italesætte' eller synliggøre problemer og vurderinger også fra de grupper, der normalt ikke deltager i en organisations politikudvikling. At problemet med kvalitetssikring af evalueringer er påtrængende kan illustreres med henvisning til en rapport om kvalitetssikring af kvalitative evalueringer fra en forskningsenhed nedsat under premierministerens kontor, og som netop er udformet som et hjælpemiddel med 18 spørgsmål eller dimensioner som skal besvares for at kunne vurdere evalueringresultater (Spencer m.fl. 2003).

Vidensproduktion og interne evalueringer.

I dette afsnit er det hensigten at se nærmere på den viden der frembringes i forbindelse med interne evalueringer i organisationer og de anvendelsesammenhænge denne viden indgår i. Vidensproduktionen og selve vidensprocessen er på det seneste igen blevet sat i fokus i evalueringssammenhænge. Det drejer sig i virkeligheden om flere delvis overlappende aktuelle diskussioner i den internationale (amerikanske) evalueringforskning, dels diskussionen af evalueringer med en klar deltagerorientering og et demokratisynspunkt overfor deltagerne, dels diskussionen om 'theory of evaluation', hvor hovedproblemstillinger drejer sig omkring sikring af kvaliteten i dataproduktionen og i tolkningen af data, dvs. vidensproduktionen i evalueringen ved brug af specifikke procedurer. Begrebet 'evidence-based decision making' som har været lanceret af den amerikanske regering som et ufravigeligt grundlag for politikudvikling, er i evalueringssammenhæng blevet taget op og udformet som en form for meta-evaluering med tydelig vægt på metodiske aspekter ved evalueringer (Shadish m.fl. 2005). Endvidere indgår diskussionerne om evalueringer og ikke mindst evaluators sociale rolle og etiske forpligtelser i diskussionen om tolkning af evalueringresultater (Schwandt 2001, 2003, Greene 1996). Baggrunden for disse diskussioner kan findes i en del af den problematik om vidensproduktion i samfundsvidenskaberne, som er blevet aktualiseret i de senere år i forbindelse med diskussionerne om ekspertrollens ændrede karakter. Dette finder vi hos med direkte med fokus på en diskussion af anvendt samfundsvidenskabs rationalitet overfor andre former for rationalitet hos Beck og Bonss (1989) eller i mere generelle vendinger som et spørgsmål om ekspertviden og ekspertfunktionen overfor lægviden hos Beck, Giddens og Lash (1994) og Beck (1999).

De to problemstillinger fra disse diskussioner er særlig vigtige i forhold til evalueringer. Det ene er spørgsmålet om hvem der producerer anerkendt viden, og hvor, hvorfor og som Beck og Bonss så klart viser, så er institutionelle grænser som i århundreder har differentieret mellem dem der producerer viden der er anerkendt af samfundet og andre vidensproducenter ved at brydes ned – universiteter er ikke længere i en monopolsituation mht. vidensproduktion, måske snarere tværtimod. Dette er blevet meget synligt i evalueringssammenhænge, når man ser nærmere på hvilke institutioner der er aktive i de større eksterne evalueringer samt naturligvis for hele området interne evalueringer. Det andet spørgsmål er i virkeligheden først og fremmest taget op af den nye organisationsteori, som i høj grad har beskæftiget sig med hvorledes viden organiseres og produceres i vidensorganisationer. I denne diskussion dukkede Michael Polanyis (1983) genopdagede begreb fra videnskabsfilosofien 'tacit knowledge' op i organisationslivet, som det kan ses i en række sammenhænge, f. eks., hos

Nonaka og Takeuchi (1995) og kritisk hos Tsoukas (2001) og Gourlay (2004). Spørgsmålet om tavs viden rejser helt særlige problemer i evalueringsforskningen på baggrund af denne forsknings meget eksplicite anvendelsesorientering og tilhørende krav om dokumentation af effekt. Dette bliver, som jeg vil vise senere, ikke mindre aktuelt af den tiltagende udvikling hen mod krav om synlig, kvantitativ og 'evidence-based' evalueringsviden.

Et af de centrale træk ved instrumenter der har været anvendt af de forskellige regeringer i de lande, hvor NPM i større omfang har været brugt til at reformere og modernisere den offentlige sektor er, at disse i høj grad består af metoder og procedurer til forvaltning og kontrol af de enkelte offentlige institutioners økonomi og aktiviteter, som var der tale om en privat virksomhed. Denne eksplicite virksomhedsorientering gør sig tydeligst gældende i den nyeste styringsteknik, mål- og rammestyningen.

Hvad der indtil i dag må siges at være et ret entydigt kendetegn ved brugen af disse metoder, hvor interne evalueringer indtager en vigtig plads, i offentlige organisationer i Vesteuropa, er fokuseringen på frembringelsen af en stor mængde kvantitative data, som bruges til vurdering af de interne processers og programmets forløb og resultater. For de offentlige institutioners produkter og ydelser gør sig endvidere det forhold gældende, at de ikke er underlagt en kapitalistisk, markedsøkonomisk styring via en markedsefterspørgsel, hvor den altafgørende 'kvalitetsvurdering' viser sig i den private producents evne til at fastholde og udvikle sin markedsposition og dermed indtjening i form af profit. De interne økonomiske styringsredskaber i virksomhederne har udviklet sig siden handelskapitalismens opblomstring i 1400-tallets italienske bystater og har som formål at sikre en ekstern kontrol med virksomhedens aktiviteter for dermed at skabe tillid imellem de deltagende købmænd. Systemerne har udviklet sig fra købmændenes dobbeltbogholderiet, standarder for mål og vægt til vor tids finansielle styringsinstrumenter og ledelsesinformationssystemer. Disse systemer har kun i begrænset omfang kunnet kopieres, men ikke fuldt ud implementeres i offentlige institutioner pga. disses afhængighed af en politisk og ikke en markedsøkonomisk baseret styring. For offentlige institutioner er der typisk en særlig relation mellem ydelser, ansatte og produktet i den forstand, at ydelserne som service i bred forstand formidles via de ansatte i institutionen men efter behov, regler og krav, der er politisk fastlagt og ikke markedsstyret. Således skriver Finansministeriet i Budgetredegørelsen fra 1995,

"Regeringen vil i de kommende år arbejde for en løbende udvikling og fornyelse af den offentlige sektor. .. Kodeordene i den proces er

forandringsparathed, fleksibilitet og resultater".¹⁸⁾

Heraf følger en fokusering på udformning af strategi og planlægning i disse offentlige institutioner med fokus på offentlig service overfor kunden og kundetilfredshed, budgetoverholdelse og produktivitet (Mouritsen, 1997, 24). Det betyder en radikal ændring i principperne for hvorledes offentlige institutioners udvikling nu skal styres i forhold til tidligere, hvor der blev forsøgt styring - ofte i detaljer - fra central politisk hold. Den tidligere centralt styrede udførelse af fastlagte politiske mål for offentlige institutioners aktiviteter afløses nu i Danmark i 90'erne af overgangen til en styring, hvor ansvaret for den enkelte institutions varetagelse af sine opgaver decentraliseres til institutionens ledelse, mens den overordnede kontrol nu udformes gennem markedslignende organisationsformer, hvor resultatvurderinger, produktivitet, kvalitet i form af brugertilfredshed alle er parametre der indgår i de strategiske planer, som nu bliver af afgørende betydning for den overordnede vurdering og kontrol med de enkelte institutioner.

De nærmere omstændigheder omkring udformningen af disse overordnede mål- og rammestyringsmekanismer skal ikke behandles her¹⁹⁾, hvor hovedsigtet er at undersøge betingelserne for og udviklingen intern evaluering i den offentlige sektor i Danmark. En gennemgående synsvinkel er, at evalueringslitteraturen med dens dominerende vægt på amerikanske erfaringer og debatter, mangler at medtænke centrale institutionelle forhold omkring udviklingen af interne evalueringer. Vigtigst er nok de offentlige institutionernes afhængighed af en udviklet velfærdsstat, de offentlige institutioners fysiske størrelse, den politiske forhistorie (skandinaviske velfærdsstater overfor USA) - alle forhold, som ikke har fået en tilstrækkelig plads i debatten om interne evalueringer. Denne debat har mere været præget af at diskutere forskelle og ligheder mellem den eksterne professionelle evaluering overfor den interne end førnævnte forhold. Yderligere gør det forhold sig gældende, at evalueringsfeltet har en lang historie i USA (Albæk 1988), mens den systematiske anvendelse af både

¹⁸ Finansministeriet, Budgetredegørelse 95, København 1995, s. 25.

¹⁹ Mouritsen (1997) giver en skitse til en historisk analyse af udviklingen i forskellige former for økonomistyring af danske offentlige institutioner i kapitel 1. Hvorledes en talbaseret styringsform som mål- og rammestyningen vinder frem som styringsredskab i forbindelse med de tidligere omtalte moderniseringsprocesser kræver en mere videnskabssociologisk undersøgelse af hvordan økonomisk nyttetænkning som den kommer til udtryk i mikroøkonomien får en så omfattende praktiske betydning har Porter søgt at eftervise historisk (1995). Mulkey, Pinch og Ashmore har i en artikel fra 1987 forsøgt at belyse nogle af de processer der er i spil når en økonomisk kalkulation i form af cost-benefit analyser bliver afgørende for beslutninger på alle led i sundhedssektoren.

eksterne og interne evalueringer kom relativt sent til de fleste europæiske lande (Hansson 1997, Chelimsky og Shadish 1997). Den voldsomme vækst i brugen af interne evalueringer i den offentlige sektor hænger i Vesteuropa nøje sammen med de moderniseringsprocesser som er gennemført i denne sektor de seneste 10 år. Et centralt træk ved disse moderniseringsprocesser er forsøg på via NPM metoder at forøge produktiviteten i sektoren først og fremmest ved at fokusere på forandringer i ledelsesfunktionen, konkret som det kommer til udtryk ved brug af mål- og rammestyring.

Koblingen mellem brug af, især interne evalueringer til løbende organisationsudvikling og den overordnede styring af samme organisationer ved brug af mål- og rammestyring etablerer imidlertid nogle nye vilkår for organisationen. Hele organisationens opbygning søges rettet mod frembringelse af sådanne data, som overgangen til mål- og rammestyring er et klart udtryk for.

"Tællighed indebærer ikke blot at der sættes et tal på et eller andet organisatorisk fænomen. Den indebærer især, at tallet er gjort til den *centrale karakteristik* heraf." (Mouritsen 97, 152)

Frembringelsen af en række såkaldte nøgletal for organisationen i forbindelse med mål- og rammestyring kan siges at være en forudsætning for udviklingen af kvantificerede interne evalueringer.

"Mål- og rammestyriings tale er rettet mod at gøre organisationer og samfund *interventionsmodne* og dermed egnede til at blive påvirket på distancen. Hermed gøres institutioners opgaver, deres måder at forvalte sig selv på, deres rolle i samfundet og deres økonomiske styring så endimensionelle, at de kan evalueres på simple snarere end komplekse evalueringskriterier... Denne sammenligning (nøgletal) muliggør intervention, idet organisatoriske enheder kan *benchmarkes* mod hinanden,.."(Mouritsen, 97, p. 24)²⁰.

Mouritsen fastslår endvidere, at

"mål- og rammestyriingen fordrer og producerer på denne vis en ny form for disciplin, hvor ledere identificerer sig med dens idé, retorik og legitimitet og dermed gør den sand. Mål- og rammestyriingen disciplinerer ikke blot medarbejderne. Den disciplinerer også lederne."(Mouritsen 1997, 104).

En særlig ledelsesform, produktionsstilen, fremhæver Mouritsen som et helt

²⁰ Se også Porter eksemplariske analyse af udviklingen af cost-benefitmetoden og dens organisatoriske konsekvenser i Porter 1995.

centralt træk, der vokser frem gennem denne udvikling.

Empowerment evaluation - deltagelse i vidensproduktion

En særlig variant af eksterne evalueringer er i de senere år dukket op ikke mindst fra USA, hvor de samles under betegnelsen empowerment evaluation, som omfatter hvad der også kaldes collaborative, participatory & empowerment evaluation²¹. Dette begreb skal nærmere diskuteres i dette afsnit fordi det kan tolkes som et af den professionelle evalueringssvar på problematikken omkring opblødning af grænserne for vidensproduktion, kombineret med erkendelsen af politiske krav om 'demokratisk' indsigt og deltagelse. Begrebet tegner eller markerer en nyorientering i evalueringfeltet med en langt større opmærksomhed på spørgsmål omkring deltagere og deltagelse i udformning og gennemførelse af evalueringer end tidligere tiders langt mere autoritative ekspertbaserede evalueringer. Begrebet er indtil i dag oftest brugt på eksterne evalueringer i det sociale, uddannelsesmæssige og socialpolitiske felt, men det er ikke vanskeligt at tænke det ind i forhold til interne evalueringer. Her rejser der sig hurtigt et par meget interessante spørgsmål. Interne evalueringer og de evaluerede i organisationer er ifølge sagens natur tæt involveret i de interne evalueringer, selvom ikke alle i en organisation nødvendigvis er vidende om disse. Samtidig har de fleste organisationer et ledelseshierarki som let kan give anledning til konflikter om påvirkning af 'empowerment evalueringer'. Empowerment i evalueringssituationen synliggør i evalueringsspektivet en række uklare spørgsmål omkring ledelse, medindflydelse og medbestemmelse i organisationen.

Selve begrebet er dukket op i de seneste 8-10 år som betegnelse for evalueringer, der meget aktivt søger at involverer de personer, der deltager i et program eller lignende i tilrettelæggelse, gennemførelse og tolkning af evalueringen. David Fetterman har i en række bøger (Fetterman m.fl. 1996, Fetterman 2001) placeret sig som en central figur i begrebets fremkomst i amerikansk evalueringsforskning, ikke mindst gennem den periode han var præsident for AEA (1993). Begrebet fokuserer som nævnt meget på de processer hvorigennem deltagerne i et program oplæres til at evaluere og formidle resultaterne. Empowerment evaluation har især i USA fået en stor udbredelse som en kritisk radikal tilgang til evalueringer modsat de

²¹ Dette er efterhånden blevet den korte 'officielle' betegnelse for Collaborative, Participatory & Empowerment Evaluation, se eksempelvis den amerikanske evalueringssituationens hjemmeside, eval.org

ekspertbaserede, men er ofte ret begrænset i det deltagerperspektiv som der lægges op til. På mange måder er det politiske indhold i dette bergeb langt mindre radikalt end den form for medinddragelse, som vi finder i den nordiske model for aktionsforskning i arbejdslivet som rummer en mere omfattende og forpligtende deltageropfattelse (Toulmin og Gustavsen 1996) og som i evalueringssammenhænge er taget op af Greene (1997) og Schwandt (1997).

For at give et indtryk af denne deltagermodels kvaliteter og begrænsninger i evalueringssammenhæng følger nedenfor en sammenfatning af debatten fra et førende evalueringstidsskrift, *Evaluation Practice*, hvor to førende evalueringsskribenter, Michael Patton og Michael Scriven kritisk kommenterer den bog, som if. Fetterman i dag må læses som den mest omfattende præsentation af empowerment evaluation, (Fetterman m. fl. 1996). Scrivens kritik (1997) er uden tvivl den mest principielle og vidtrækkende. Scriven finder den mest omfattende definition i ovennævnte bog. Empowerment evaluation, hvor forfatterne lægger følgende definition frem af empowerment evaluation:

"(it) is designed to help people help themselves and improve their programs using a form of self evaluation and reflection. Program participants conduct their own evaluations and typically act as facilitators; an outside evaluator often serves as a coach or additional facilitator.." (Fetterman m.fl.1996, p. 5)

Scriven påpeger, at vi her har at gøre med et begreb om evalueringer, der først og fremmest skal forstås som interne evalueringer, som en teknik til at forbedre disse. Det er ifølge Scriven i den sammenhæng overraskende, at interne evalueringer som begreb slet ikke diskuteres i ovennævnte bog, til trods for at stort set alle de enkeltcases der behandles i de forskellige kapitler er interne evalueringer! Her har vi ifølge Scriven det helt afgørende problem ved bogen. Den skelner ikke mellem professionelle eksterne evalueringer og interne og kommer derfor let til at fremstå som et program for fornyelse af alle slags evalueringer i tidens aktive deltagelsestegn. Men for Scriven kan bogen kun argumentere for en opfattelse af empowerment evaluation som en strategi for udvikling af interne evalueringer hvor personalet er aktivt deltagende. En stor del af diskussionen går på hvordan den udefrakommende evaluator-coach skal oplære de personer, personale og lokale aktører, der deltager i et program til selv at kunne gennemføre evalueringer. Scriven slår fast, at

"at best, this failure to focus clearly and precisely on who should be empowered, and when, is one of the most unclear parts of the concept of empowerment evaluation, and that lack of clarity tells us that we cannot casually welcome this approach to evaluation as an obvious improvement. It has some advantages for staff, indeed, but

we know little about what it does to, or for, the consumers."
(Schriften 1997, 169)

Der mangler ifølge Schriften klare refleksioner om vidensproduktionen i empowerment evaluation. Schriften viser hvorledes denne fokusering på aktivisering af personalet i en fortløbende evalueringsproces nok er vigtig og kan rumme en række fordele for interne evalueringer - men sandelig også principielle problemer. Det ene er de evalueringsblokeringer der kommer, når problemer og ubehageligheder kommer for tæt på, og også den bias der er et resultat af at være en del af processen. Og det andet er den utroværdighed eller i hvert fald usikkerhed mht. evalueringresultater produceret på denne måde set udefra, fra andre og især forbrugere. Til spørgsmålet om hvem der skal nyde godt af empowerment evaluation hælder Fetterman uden nærmere diskussion mod dem der deltager i programmer og projekter. Denne nærmest refleksemæssige aktivisme har Schriften sat nogle klare spørgsmålstejn ved, fordi den i ringe grad medtænker spørgsmål om vidensproduktion og anvendelse, selvom den naturligvis må ses som en positiv udvikling i forhold til helt traditionelt og entydigt ledelsesstyrede evalueringer!

Patton's kritik tager denne ende op og kritiserer den meget ideologiske lancering af begrebet empowerment evaluation, som slet ikke for alvor placerer dette begreb i forhold til andre og langt ældre traditioner for at inddrage andre end forskere og eksperter i evalueringsprocessen; dels de europæiske traditioner for frigørende socialforskning såsom kritiks teori, kommunikativ handlen, frigørelsespædagogik mm., men også deltagerforskning og aktionsforskning. Men Fetterman har ikke i de forskellige bøger forsøgt at udvikle begrebet eller placere det i en evalueringsteoretisk sammenhæng og ifølge Patton efterlader det begrebet i en rodet eller mudret sammenblanding med andre evalueringstyper.

"Moreover, because empowerment evaluation overlaps participatory, collaborative, stakeholder-involving, and utilisation-focused approaches to evaluation in its emphasis on attending such issues as ownership, relevance, understandability, access and involvement, a critical question becomes how to distinguish empowerment evaluation from these other approaches". (Patton 1997, 148)

Nok fokuserer empowerment evaluation meget på de læreprocesser, som sker når deltagere/personale skal inddrages i et program/projekt og skal lære at gennemføre evalueringer - men dette diskuteres ikke på det mere overordnede plan og slet ikke i forhold til mere principielle forandringer, som den professionelle evaluators rolle også gennemgår, når den ændres til at være programadvokat.

"Program advocacy may undercut the evaluator's credibility, not

only with external audiences, but even with program participants who may sense that the evaluator has been co-opted and that the very capacity to do meaningful reality-testing that the evaluator was supposed to be teaching is not being practiced by the evaluator." (Patton 1997, 159)

Ryan og Johnson (2000) har i en artikel i *New Directions for Evaluation's* temanummer om *Evaluation as a Democratic Process*²², gennemgået litteraturen med særlig vægt på den interne evaluators problemer med på den ene side at have tætte relationer til organisationens ledelse og på den anden i nogle tilfælde med klare sympatier for en bredere deltagelse i evalueringsprocessen. Ryan og Johnson peger på House (1996) og House og Howe (1999) som de bidrag i denne debat, der er mest lovende for dialog og inddragelse i evalueringer, selvom de efterlyser en klarere diskussion af deltagelsesbegrebet.

²² Dette temanummer fra 2000 rummer en række bidrag fra fremtrædende evalueringsforskere.

Kvalitet i interne evalueringer – hvem kontroller viden og hvordan

Normalt vil anvendelsen af evalueringsviden, dvs. viden frembragt gennem anvendelse af evalueringsprocedurer og metoder, være underlagt samme rammer og betingelser både i frembringelse og anvendelse som anden anvendt samfundsvidenskabelig viden.

Tidligere tiders optimistiske evalueringsforskere så ofte evalueringsforskning baggrund af en social-engineerings forestilling, hvor evalueringer fremstod som et særligt redskab til at 'forbedre verden' gennem direkte påvirkning af politiske beslutningsprocesser. De sidste 15 års forskning i implementeringsprocesser, evalueringers rolle og politisk styring har grundigt belært også de optimistiske evalueringstilhængere om evalueringsforskningens mere ydmyge rolle i forhold til politikformulering (Weiss 1999, Albæk 1996, Lindblom og Cohen 1979, Vedung 1994, 1997, Schwandt 2001, 2003). Omvendt har New Public Management bølgen i et ikke ringe omfang inddraget evalueringer på en ganske håndfast og instrumentel måde som redskab til politikformulering (Pollitt 1995).

For så vidt angår den professionelt udførte eksterne evalueringsforskning kan diskussionen om anvendelse af videnskabelig viden og ekspertroller i denne sammenhæng læses som et led i den løbende diskussion af disse problemstillinger. I Danmark, hvor offentlige organisationer mht. størrelse aldrig er kommet i nærheden af de førnævnte amerikanske organisationer mht. størrelse, finder vi kun noget der i struktur ligner de amerikanske forhold med selvstændige evalueringsenheder i offentlige organisationer på ministerieplan, f.eks. som undervisningsministeriets evalueringsenhed (Evalueringsinstituttet EVA) eller det i Sverige tilsvarende Statens Skolverks Utvärderinger. I første omgang er det måske mere en empirisk konstatering end en egentlig forklaring på forholdet, når der i mindre lande som Danmark og Sverige ikke findes samme udbredelse af professionelt bemandede og organiserede, delvis selvstændige, evalueringsenheder i den offentlige sektor som i USA. Her findes en årelang tradition (Albæk 1988) for brug af program- og projektevaluering i forbindelse med offentlige foranstaltninger, politikker, programmer osv., og en politisk anvendelse af evalueringresultater samt ikke mindst en specialiseret akademisk profession.²³⁾

²³ Omfanget af selvstændige evalueringsenheder i offentlige organisationer i USA er vanskelig at opgøre. En række artikler i tidsskriftet *Evaluation Practice* i de senere år giver dog baggrund for at antage, at sådanne enheder findes i betragteligt omfang. I en diskussion af i hvilket omfang evalueringsaktiviteterne på forbundsplan er blevet involveret i Clinton-administrationens offentlige moderniseringsplan fra 1993 (*Government performance and*

Men trods mangel på en tidlig udbredelse af de mere professionelt organiserede og udførte interne evalueringer i forhold til USA, finder vi i inden for de seneste 10 år også i lande som Danmark og Sverige en stor udbredelse af en anden form for interne evalueringer, nemlig interne evalueringer som planlægges og udføres af den enkelte offentlige institution/organisation og som mere eller mindre integreret i den almindelige drift som en del af udviklings- og kontrolfunktionen, ofte begrundet i den overordnede statslige moderniseringspolitik.

I interne evalueringer organiserer institution eller organisation selv evalueringen og fastlægger selv hvad der skal evalueres, hvilke metoder der skal bruges, hvordan data skal analyseres samt den videre anvendelse af resultaterne - og ganske ofte uden brug af professionel evalueringsbistand. På den måde bliver sådanne interne evalueringer nærmest at sammenligne med andre dele af den interne organisationsudvikling. Den evalueringprofessionalisme som må antages at være en forudsætning for at sikre et vist niveau af faglig evalueringskompetence til kritisk vurdering af metoder og resultater samt ikke mindst at få stillet de kritiske evalueringsspørgsmål, kan ikke uden videre antages at være til stede i disse interne evalueringssammenhænge. Professionaliseringen som kendetegner de tidligere nævnte interne evalueringer i de store amerikanske organisationer er kun undtagelsesvis til stede når vi ser på interne evalueringer udenfor ovennævnte store organisationer. Det kan lede frem til situationer, hvor kombinationen af et ydre pres for at benytte evaluering og andre kvalitetsstyringsmetoder, som det hyppigt ses formuleret bl.a. i resultatkontrakter for offentlige institutioner, og de samtidig begrænsede økonomiske og personalemæssige ressourcer, ikke skaber de bedste vilkår for en udvikling af en egentlig professionel evalueringseenhed i organisationen. Der er også grund til at antage, at organisationens ledelse kan have en ambivalent holdning til en professionalisering af evalueringsfunktioner, da det dels alt andet lige vil blive dyrere at gennemføre evalueringer og måske vigtigst, de vil opnå en højere grad af selvstændighed i organisationen, i og med at de kan referere til eksterne fagfæller. Brug af eksterne konsulenter og Finansministeriets mange vejledninger er ikke altid tilstrækkelig hjælp til at sikre en rimelig kvalitet som det vil fremgå lidt senere.

I Danmark er situationen for evalueringsforskningen den, at fremkomsten af et fagligt, professionelt evalueringsmiljø - trods mængden af evalueringer der er produceret de senere år – skete ret sent, senere end i de fleste

Result Act) har Wargo (1995) samlet data fra de 14 største interne evalueringseenheder på forbundsplan, som står for 98% af den interne. I en spørgeskemaundersøgelse af et tilfældigt udvalg af medlemmer af American Evaluation Association, hvor Torres m.fl. (1997) undersøgte holdninger til forskellige kommunikationsstrategier angav 52% at de opfattede sig som interne evaluatører i offentlige institutioner.

Vesteuropæiske lande (Hansson 1997). En stor del af de eksterne evalueringsaktiviteter er gennemført af forskningsinstitutioner for anvendt forskning, som indtil nu har udvist en overvældende forkærlighed for mere traditionelle metodiske tilgange til evalueringer, først og fremmest surveymetoder, eller konsulentfirmaer, hvis metodebrug ofte er mindre veldokumenteret, men baseret på mere kvalitative og case baserede tilgange. En faglig specialisering i en egentlig professionel og selvstændig evalueringsforskning, som den tydeligt kan findes i USA (AEA), men også ses i f. eks. Sverige og England, er først ved at udvikle sig i Danmark i de seneste år.

Det har flere konsekvenser udover det rent praktiske, at det er vanskeligt at rekruttere professionelle evaluører. En måske større vanskelighed er at fastholde og understøtte den faglige og metodiske kvalitetsudvikling og standardisering, som normalt foregår indenfor en videnskabelig disciplin, når den udvikler sig som profession (Whitley 1995). Mulighederne for at opsamle og fastholde den lokale og internationale faglige debat i evalueringsforskningen er således svagere i Danmark, og dette rammer første omgang den eksterne evalueringsforskning. Men også intern evaluering rammes, idet både formidling af teorier og metoder, samt personer med særlige kvalifikationer til at dække det store behov, der opstår i begyndelsen af 90'erne, bliver vanskelig. Der er således en række gode forklaringer på den voldsomme praksisorientering og meget ukritiske anvendelse af forenkede og nogle gange fejlagtige metoder som er et kendetegn for mange af 90'ernes interne evalueringer. Evalueringer i en organisation bliver under sådanne betingelser let begrænset til at rette sig mod hvad det er let at skaffe data om og ikke så meget, hvad det er relevant for organisationen at evaluere!

En undersøgelse som det svenske Finansdepartement iværksatte i 1996 af netop disse spørgsmål, rejser spørgsmålet direkte i rapportens titel "Kan myndigheter utvärdera sig själva?" - og Barkman og Fölster (1996) konkluderer efter en gennemgang uafhængige ekspertpanelers vurderinger af 23 svenske offentlige institutioner og organisationers interne evalueringer (kaldet selvevalueringer i rapporten), "at de fleste självutvärderingarna är av så dårlig kvalitet att de faktisk inte bör användas som beslutsunderlag" (s. 8) og at der i høj grad er tale om lav kvalitet i dokumentation, metodevalg og analyse og at der drages slutninger der ikke er belæg for i det empiriske materiale. Deres forslag til forbedringer af kvaliteten i interne evalueringer omfatter brug af uafhængige sagkyndige til at kommentere evalueringerne. En sådan kvalitetssikring vil nok forbedre kvaliteten i de enkelte evalueringer, men vil især understreger forfatterne have betydning ved at understøtte udviklingen af den nødvendige evalueringskompetence i den enkelte organisation (s. 74-75). I Danmark har finansministeriet ikke gennemført sådanne omfattende metaevalueringer, men der er intet grundlag for at antage, at resultatet ville have været meget anderledes end i den

svenske undersøgelse. Finansministeriet har godt nok forsøgt sig med en 'discount' model for at opruste den interne evalueringskompetence i offentlige organisationer, som fra politisk hold bliver pålagt et stadig mere omfattende arbejde end at kvalitetssikre og evaluere sine aktiviteter. Som et led i denne evalueringsoprustning har finansministeriet udsendt en række vejledninger i hvordan interne evalueringer kan gennemføres ²⁴⁾. Hvis man ser nærmere på disse, må man nok erkende at de nok ikke ville få en særlig positiv vurdering af evalueringseksperter. Hovedlinien i vejledningerne ligger på en meget traditionel brug af meget enkle spørgeskemaer som kilde til indsamling af empiri. Andre kvantitative metoder - for slet ikke at tale om kvalitative metoder - eksisterer ikke i dette univers. Analysetilgangen til det indsamlede materiale er tilsvarende forenklet, der skal blot laves procentfordelinger på tabellerne på de enkelte spørgsmål og enkle fordelingsanalyser. Vejledningerne er tilsvarende rensset for henvisninger til evalueringsforskningens store arsenal af metodelitteratur ²⁵⁾

Der findes et par danske studier af kvalitet og metodeanvendelse i lokale evalueringer. En undersøgelse gennemført af Bjørn Lomborg (1998) af primært interne evalueringer i 26 kommuner i Århus Amt. Denne metaevaluering ser nærmere på de metoder og analyseteknikker der blev anvendt i disse undersøgelser viste, at 95% af disse evalueringer brugte udelukkende kvantitative metoder og selvom Lomborg har valgt at stille ganske beskedne metodologiske krav til dataindsamling og analyse er konklusionen,

"..at borger/brugerundersøgelser i Danmark gennemgående er ringe. Kun 45% af alle datasæt lever op til nogle meget minimale krav, og for analyse og teori ser det endnu værre ud: kun 12,5% lever op til de minimale metodekrav og kun 5% til minimale teorikrav. Samlet er der ikke én analyse, der lever op til alle de krav om minimalt acceptabel teori, data og analyse."

²⁴ Eksempler på disse publikationer er:

Finansministeriet (ed.) Medarbejderne i fokus. Om interne spørgeskemaundersøgelser. København: Schultz. 1995, Focus på spørgeskemaundersøgelser. København: Schultz. 1995, Medarbejdernes syn på den offentlige sektor. København: Finansministeriet. 1995, Evaluering og effektsikring. København: 1996, Brugere i fokus. effektive institutioner. København: 1996.

²⁵ Det vil med stor sandsynlighed lede frem til sådanne evalueringer, som de svenske eksperter kritiserer for ikke at gennemføre en egentlig evaluering men alligevel drage kontante slutninger ud fra materialet, " där man i syftet "friskriver sig" från at göra en "riktig" utvärdering men ändå med bestämdhet slår fast olika resultat. I detta fall bör givetvis resultatet i studien bedömmas utifrån om det i studien finns empiri och metod som håller för sådana slutsatser". (Barkman og Fölster, op.cit. s.55.

Hvorfor man finder denne dominans eller ensidighed i valg af kvantitativ metodologi til måling og analyse af evalueringsdata, endda dårligt og uprofessionelt udført, er ikke en problemstilling, som er blevet undersøgt på nogen systematisk måde. Det forekommer rimeligt at antage, at den dominans af kvantitative metoder, som gør sig gældende, og stadig præger samfundsvidenskaberne og især de anvendte videnskaber, trods væksten i brug af kvalitative tilgange i de senere år, kan forklare dette forhold. Videnskabshistoriske studier har overbevisende påvist kvantificeringens dominans og især politiske gennemslagskraft, som f. eks. hos Porter (1995) og Rescher (1995), men vi finder kun meget få og spredte tilløb til at fastholde en metodekritik i de få metaevalueringer der har fokuseret på metodeaspekter²⁶. Balancen mellem at kunne gennemskue gyldighed i det empiriske materiale og relevansen af de analysemetoder, der er relevante overfor det tids- og økonomipres for at 'levere varen' som interne evalueringer formodentlig er underlagt i endnu større omfang end de eksterne evalueringer, er uden tvivl vanskelig. Det kan dog ikke bruges som undskyldning for den udbredte mangel på kvalitet i interne evalueringer, som ovennævnte undersøgelser påpeger og som viser sig i manglende overholdelse af helt grundlæggende krav til produktion og analyse af samfundsvidenskabelige data. Der kan ikke herske tvivl om, at der på dette foreløbige grundlag kan rejses alvorlig kritik af både kvalitative og kvantitative tilgange i interne evalueringer.

Formel kvalitetskontrol med evalueringer

Kritikken af metodologiske og analytiske fejl og svagheder i både in- og eksterne evalueringer er dog ikke gået evalueringsverdenen ubemærket forbi. Med afsæt i erfaringer fra en succesfuld organisering af kvalitetskontrol for kræftforskning, som er gennemført af Cochrane Collaboration, har en gruppe evalueringsforskere forsøgt at få etableret et tilsvarende kvalitets- og autorisationssystem under betegnelsen Campbell Collaboration, som eksplicit trækker på nu afdøde Donald Campbell's velkendte quasi-eksperimentelle tilgang til evalueringer fra 1960'erne. Til at understøtte behovet for en sådan kontrol eller autorisation af kvaliteten i evalueringer henviser tilhængerne af dette initiativ også til det i den amerikanske forbundsregering meget udbredte krav om 'evidence-based decision making' – forstået på den måde, at der skal fremskaffes et solidt grundlag for at forskellige politikker og programmer på forbundsniveau. Campbell Collaboration tænkes i den forbindelse som en uafhængig institution, der kan garantere kvaliteten i evalueringsresultater og som primært skal baseres på en omfattende brug af metaevalueringer for at

²⁶ Henning Olsen har i flere undersøgelser beskæftiget sig med metodeaspektet i anvendt samfundsforskning, se Olsen 2002.

belyse effekter af politikker og programmer på forskellige områder. Campbell-initiativet er etableret i USA (Shadish m.fl. 2005) og har fået følge af en nordisk Campbell organisation, som Socialforskningsinstituttet aktivt er medorganisator af. Shadish (2005) refererer en undersøgelse af europæiske evalueringer og fremlægger i den forbindelse en række af de metodekrav, som for dem er uomgængelige hvis evalueringer skal indgå i meta-evalueringer og godkendes af Campbell Collaboration. Blandt disse metodekrav er krav om at brug af randomiseret udvælgelse af undersøgelsespopulationen, bruge quasi-experimentelle metoder, standardisering af afhængige variable, lavt frafald, tilstrækkelig samplestørrelse Shadish m.fl. (2005). Til at underbygge argumentet for at udvikle en sådan kvalitetskontrol, har de fra en række databaser fremfundet et større antal europæiske evalueringstudier og analyseret disse ud fra de førnævnte metodekrav. Her har de fundet en række kritikpunkter i de fleste europæiske evalueringer – og det skal her fremhæves at dette initiativ nok først og fremmest er rettet mod eksterne evalueringer, selvom en række interne evalueringer sikkert i endnu større omfang vil falde for deres kritik.

”These data suggest some potential reasons why past meta-analyses have included relatively few experiments that were conducted in Europe. To begin with, relatively few such experiments exist from Europe compared to the many thousands that have been conducted in the US. In addition, without the data to calculate effect sizes, a study cannot be included in a meta-analysis. And perhaps most important, the fact that most studies used non-experimental or quasiexperimental methods would often result in the elimination of such studies from systematic reviews using strict inclusion criteria. A real paucity of randomized experiments exists in the European context.” Shadish m.fl. (2005)

Denne meget bastante kritik af europæiske evalueringer illustrerer ganske godt den metodiske snæversynethed, rettet alene mod eksperimenter som metode og kvantitative data som resultat, som hele dette projekt er gennemsyret af. At kræve bedre kvalitet i dataindsamling, redegørelse for gyldighed og for valg af analysemetode bør ikke sættes lig at tilslutte sig denne neofundamentalistiske metodeskole. Det er ikke uden ironi at kunne konstatere, at den Donald Campbell, som i tide og utide bringes frem for at legitimere hele projektet, næppe selv ville have accepteret at blive (mis)brugt til et sådant dogmatisk og fundamentalistisk projekt. I 1978 skrev han i en artikel om kvalitative metoder i programevaluering og fremhævede, at kvalitativ viden altid har haft og altid vil have en central plads i videnskabeligt arbejde – ikke mindst fordi de kvalitative dele af videnskabeligt arbejde kan medtage alle de ikke-udtalte, ikke formaliserede elementer, den tavse viden ifølge Michel Polanyi (1983). Det er fristende at citere Campbells afsluttende kommenter fra artikel:

”More than that, I have sought to remind my quantitative colleagues that in the successful laboratory sciences, quantification both builds upon and is cross-validated by the scientist’s pervasive qualitative knowledge. The condition of mass-produced quantitative sociale science in program evaluation are such, that much of this is apt to be lost. If we are to be truly scientific, we must reestablish this qualitative grounding of the quantitative in action research.” (Campbell 1978)

Vægten henimod brug af mere traditionelle kvantitative metoder og quasiekperimentet som undersøgelsesdesign, pakket ind i en retorik om ’stringente metodiske regler’ og højeste videnskabelige standard’ vækker minder om den værdifri positivistiske samfundsforskning fra 1960erne²⁷. Positivismekritikken og den efterfølgende åbning mod brug af et langt bredere arsenal af samfundsvidenskabelige metoder og analysetilgange, ikke mindst kvalitative, samt erkendelsen af metodernes ’teoriladethed’ (Quine 1964, 1996). Samfundsforskningens iboende afhængighed af sociale og politiske værdier og dens historiske evne til at stille relevante kritiske spørgsmål udsættes med initiativer som Campbell Collaboration for, hvad House (2005) meget rammende kalder en ny fundamentalisme, hvor det også i samfundsforskning, herunder evalueringsforskning, handler om at vende tilbage til det sikre og afprøvede – væk fra de kritiske. Det skulle fremgå, at der i høj grad er tale om kampen om magt i forhold til evalueringer, om magten til at definere hvad der er anerkendt videnskabelig viden og dermed den viden der skal anvendes samt om hvem der skal stå for denne definitions magt. Man kan have megen sympati med forsøg på at hæve det metodiske og analytiske niveau i både in- og eksterne evalueringer, og det skulle have fremgået tidligere i dette afsnit, at behovet herfor er stort. Problemet er blot at fremgangsmåder som Campbell Collaboration og tilsvarende forsøg på at sætte standarder berettiget skabe uro i evalueringsverdenen, fordi ensliggørelse og standardisering her – eller som House formulerer det, en ny fundamentalisme – først og fremmest vil ramme kritiske, nye og eksperimenterende metoder og tilgange til evalueringer.

Mens ovenstående forsøg på at standardisere eller regulere evalueringer er udsprunget af et mere eller mindre privat initiativ, så er der også initiativer der har en direkte politisk baggrund. Den britiske regering har udsendt en rapport om kvalitet i kvalitative evalueringer (Spencer m.fl. 2003) til alle

²⁷ Se eksempelvis diskussionen af positivismekritikken ud fra et nordisk hhv. tysk perspektiv i Mjøset 2003 og Dahms 1994. I dansk sammenhæng se Hansson og Aagaard 1996.

regeringskontorer og organisationer. Rapporten baserer sig på litteraturstudier og interviews med forskere og behandler et stort antal dimensioner. I en kort sammenfatning af rapporten er der et skema med 18 nøglespørgsmål, alle med en række underspørgsmål, som kan bruges til at vurdere en kvalitativ evaluering. Det er interessant og ganske tankevækkende, at rapportens udgangspunkt er kvalitative metoder i evalueringer og ikke som hos Campbell Collaboration, en mere eller mindre forklædt orientering mod kvantitativ metodologi. Kushner (2005) fremhæver i sin anmeldelse, at rapporten trods en række begrænsninger i sin forståelse af forskelle mellem evalueringsformer, kan være med til at hæve kvaliteten i interne evalueringer og i brug og tolkning af evalueringer, som trækkes ind i organisationer. Af mere alvorlige mangler er fraværet af diskussionen af forholdet mellem professionelle evaluatore og den rekvirerende organisation - måske fordi rapporten slet ikke behandler det centrale spørgsmål om "the democratic function of evaluation and its sponsorship" (Kushner 2005, 121), men dette er måske en mangel der afspejler at rapporten er udarbejdet direkte under premierministerens kontor.

I stedet for en slutning: Lake Wobegon effekten i evalueringer

Evalueringssystemer, der systematisk bringes ind og anvendes i en organisations indre liv bliver udenfor enhver diskussion et element i en disciplinering af organisation, omgivende samfund såvel som af de enkelte ansatte.

Helt på samme måde som en organisation i evalueringssystemer ofte sætter sine indsatsområder netop på de felter, hvor der evalueres for at synliggøre sig i det politiske landskab, så handler de enkelte medarbejdere ud fra samme strategi. Dette er en side af evalueringer, der er uhyre vanskelig at belyse empirisk, den er kompleks og sammenvævet med en række andre led i organisationen, men må netop derfor fremhæves som et yderst vigtigt perspektiv på evalueringer i organisationer, fordi der er tale om en mere varig og dybtgående virkning af evaluering end de blotte tilpasninger til et resultat. Disse sider af evalueringens anvendelsesadfærdsmæssige konsekvenser ville Foucault have medtaget i begrebet om den skjulte disciplinering, hvor evalueringen måske kan tage panoptikonmodellens plads som det altseende øje. For de enkelte personer betyder det at adfærd og aktiviteter indrettes efter forventninger om en optimering i forhold til de evaluerede mål. For organisationen betyder det at evalueringer mere og mere systematisk indgår i organisationens strategiske tænkning såvel som i den måde organisationen udvikler eller undgår at udvikle og forholde sig til politikområder på. (Brunsson 2003)

Foran har jeg søgt at diskutere en række af de dilemmaer og modsigelser, som interne evalueringer støder på. En hel del er fælles for alle former for evalueringer, såsom spørgsmål om kvalitet i de data der produceres og analyseres. En hel del af disse dilemmaer og problemstillinger har at gøre med spørgsmål om i hvilket omfang en evaluering overhovedet bliver synlig, dvs. om den kommunikeres udenfor den helt snævre kreds af opdragsgivere. Torres, Preskill og Piontek (1997) har gennemført en spørgeskemaundersøgelse af et tilfældigt udvalg på 200 evaluators rapport- og kommunikationserfaringer. Lidt overraskende angiver halvdelen at de er interne evaluere. Mht. rapport- og kommunikationspraksis fandt undersøgelsen en række forskelle mellem eksterne og interne evaluere. Eksterne evaluere skriver hyppigere tekniske rapporter, afsluttende rapportering, planer, summaries og tidsskriftartikler, dvs. de opfylder mere eksplicit de forventninger som klienten måtte have eller som er formuleret i en kontrakt.

Både eksterne og interne evaluere er dog ganske utilfredse med den modtagelse deres evalueringer får. Mangel på en effektiv kommunikation i forhold til det publikum, som skulle bruge evalueringen skyldes dels tidsnød mht. afsluttende kommunikation som tilsyneladende især rammer interne

evaluatorer, dels at den afsluttende rapport i sin form følger samfundsvidenskabelige traditioner for præsentationer af forskningsresultater, dels at der hos brugerne fandtes en uklarhed om hvad og hvordan evalueringen skulle bruges og at der i en række tilfælde var en række politiske omstændigheder, der begrænsede modtagelsen af resultaterne. Specielt interne evaluatorer kan have problemer med at kommunikere resultater af evalueringer, især hvis de er negative i en organisation. Interne evaluatorer er ifølge undersøgelsen signifikant mere utilfredse med deres kommunikations- og rapportanstrengelser.

Kommunikation og udbredelse af evalueringer er vigtige for udviklingen af demokratiske processer både internt i organisationer men også udenfor. Man kan næppe tale om at få for megen evalueringer, men den stadig voksende udbredelse og systematiske anvendelse af formaliserede evalueringer (spørgeskemaer) i alle mulige sammenhænge kan føre frem til samme form for 'metaltræthed' som man kender omkring brug af tests til psykologiske og pædagogiske formål.

En parallel til den omfattende institutionaliserede brug af evalueringer er debatten om Lake Wobegon effekten²⁸ som vi finder i pædagogikkens og psykologiens diskussioner om konsekvenserne af en omfattende brug af test og evalueringsmetoder. Kort fortalt er Lake Wobegon effekten en samlebetegnelse for en række testfænomener, som alle har at gøre med det observerbare forhold, at den gyldighed en testmetode (eller en evalueringsmetode) har på et tidspunkt udhules, fordi de testede personer over tid på forskellig vis udviser en adfærd, der ikke svarer til den normalfordeling af adfærd/besvarelser, der er metodens grundlag. Det kan være simple forhold som at svarpersoner systematisk overvurderer (eller undervurderer) forskellige forhold, at testpersoner lærer - eller ligefrem som i forhold til færdighedstests i amerikanske skoler uddannes til at kunne besvare de meget standardiserede testspørgsmål mest effektivt. De skævheder, der på denne måde kommer ind i måleresultaterne, er helt parallelle til de ovennævnte effekter af interne evalueringer i organisationer og er udtryk for den mere skjulte magt i evalueringer (og tests), som disciplinerer aktørerne for at bruge Foucault's terminologi.

Hvis den skjulte magtudøvelse gennem evalueringer på sådanne måder gør os alle til magtstrateger, der søger at påvirke evalueringers udfald gennem at besvare spørgsmål så de fremmer særlige interesser, er vi kommet til en situation, hvor evalueringer – og nok især interne, fordi de fungerer i en sluttet population kan betragtes som en form for 3.ordens observation af de

²⁸ Se f. eks. Maxwell, N. L. & Lopus, J. S. (1994). En mere omfattende diskussion af testeffekter som også giver en udførlig præsentation af Lake Wobegon effekten kan findes i Haney, W. M., Madaus, G. F., & Lyons, R. (1995).

forhold, der undersøges. De der evalueres forholder sig strategisk til de forventninger til resultatet som de antager opdragiverne har med evalueringen. Vi er så kommet en lang vej fra evalueringsfeltets pionerer, som tilbage i 1960'erne havde illusioner om at denne nye form for brug af samfundsforskningen kunne afhjælpe fejl og utilsigtede konsekvenser af politikker og bidrage til social og politiske forbedringer.

Referencer:

- Abma, T. A., Greene, J., Karlsson, O., Ryan, K. E., Schwandt, T. A., & Widdershoven, G. A. M. (2001). Dialogue on Dialogue. *Evaluation*, 7, 164-180.
- Albæk, E. (1988). *Fra sandhed til information*. Copenhagen: Akademisk forlag.
- Albæk, E. (1994). *Hvorfor nu al den evaluering? Teoretiske noter om empiriske observationer om evalueringsforskningens fremvækst og funktion i Danmark*. Aarhus: Institut for Statskundskab, Aarhus Universitet.
- Albæk, E. (1996). Why All This Evaluation? Theoretical Notes and Empirical Observations on the Function and Growth of Evaluation, with Denmark as an illustrative Case. *Canadian Journal of Program Evaluation*, 11, 1-34.
- Albæk, E. (2001). Vidensinteresser og de mange betydninger af evaluering, et udviklingsperspektiv. In P. Dahler-Larsen & H. K. Krogstrup (Eds.), *Tendenser i evaluering* (Odense: Odense Universitetsforlag.
- Altvater, E. & Mahnkopf, B. (1997). The world market unbound. In A. Scott (Ed.), *The limits of globalization - cases and arguments* (pp. 306-326). London: Routledge.
- Barkman, C. & Fölster, S. (1996). *Kan myndigheter utvärdera sig själva? Rapport till expertgruppen för studier i offentlig ekonomi*. Stockholm: Fritzes.
- Beck, U. & Bonss, W. (1989). Zum Strukturwandel von Sozialwissenschaft und Praxis. Ergebnisse und Perspektiven der Verwendungsforschung. *Soziale Welt*, 198-214.
- Beck, U. & Bonss, W. (1989). Verwissenschaftlichung ohne Aufklärung? Zum Strukturwandel von Sozialwissenschaft und Praxis. In U. Beck & W. Bonss (Eds.), *Weder Sozialtechnologie noch Aufklärung? Analyse zur Verwendung sozialwissenschaftlichen Wissens* (pp. 7-45). Frankfurt am Main: Suhrkamp Verlag.
- Beck, U., Lash, S., & Giddens, A. (1994). *Reflexive modernization. Politics, tradition and aesthetics in the modern social order*. Cambridge: Polity Press.

- Beck, U. (1997). *The Reinvention of Politics. Rethinking Modernity in the Global Social Order*. Cambridge: Polity Press.
- Beck, U. (1999). *World Risk Society*. Cambridge: Polity Press.
- Beck, U. (2000). Risk Society Revisited. Theory, Politics and Research Programmes. In B.Adam, U. Beck, & J. van Loon (Eds.), *The Risk Society and Beyond* (pp. 211-229). London: Sage Publication.
- Breslau, D. (1997). Contract shop epistemology. Credibility and problem construction in applied social science. *Social Studies of Science*, 27, 363-394.
- Brunsson, N. (2002). *The organization of hypocrisy. Talk, decisions and actions in organizations*. Copenhagen: Copenhagen Business School Press.
- Brunsson, N. (2003). Organized Hypocrisy. In B.Czarniawska & G. Sevón (Eds.), *The Northern Lights* (Copenhagen: Copenhagen Business School Press).
- Bryant, C. G. A. (1995). *Practical Sociology. Post-empiricism and the Reconstruction of Theory and Application*. Cambridge: Polity Press.
- Campbell, D. T. (1978). Qualitative knowing in action research. In M.Brenner, P. Marsh, & M. Brenner (Eds.), *The social context of method* (pp. 184-209). London: Croom Helm.
- Chelimsky, Eleanor and Shadish, William R. (1997). *Evaluation for the 21st century. A handbook*. Thousand Oaks: Sage.
- Chia, R. (2002). Essai: Time, Duration and Simultaneity: Rethinking Process and Change in Organizational Analysis. *Organization Studies*, 23, 863-868.
- Dahler-Larsen, P. (1998). *Den rituelle refleksion. Om evaluering i organisationer*. Odense: Odense Universitetsforlag.
- Dahler-Larsen, P. (2001). Organisationalisering af evaluering. In P.Dahler-Larsen & H. K. Krogstrup (Eds.), *Tendenser i evaluering* (Odense: Odense Universitetsforlag).
- Dahler-Larsen, Peter and Krogstrup, Hanne Kathrine (2001). *Tendenser i evaluering*. Odense: Odense Universitetsforlag.

- Dahler-Larsen, P. (2004). *Evaluering og magt*. Århus: Aarhus Universitetsforlag.
- Dahms, H. J. (1994). *Positivismusstreit. Die Auseinandersetzungen der Frankfurter Schule mit dem logischen Positivismus, dem amerikanischen Pragmatismus und dem kritischen Rationalismus*. Frankfurt am Main: Suhrkamp.
- Etzkowitz, H. & Leydesdorff, L. 2000, "The dynamics of innovation: from national systems and "Mode 2" to a triple Helix of university-industry-government relations", *Research Policy*, vol. 29, pp. 109-123.
- Fetterman, David M., Kaftarian, Shakeh J., and Wandersman, Abraham (1996). *Empowerment evaluation. knowledge and tools for self-assessment and accountability*. Thousand Oaks: Sage.
- Fetterman, D. (1997). Empowerment evaluation: A response to Patton and Scriven. *Evaluation Practice*, 18, 253-266.
- Fetterman, D. M. (2001). *Foundations of empowerment evaluation*. London: Sage Publications Ltd.
- Foss Hansen, H. (2001). Rejsende praksis. hvilke forhold påvirker udviklingen af evalueringspraksis på forskellige politikområder? In P.Dahler-Larsen & H. K. Krogstrup (Eds.), *Tendenser i evaluering* (Odense: Odense Universitetsforlag.
- Foss Hansen, H. (2003). *Evaluering i staten. kontrol, læring eller forandring?* København: Samfundslitteratur.
- Greene, J. C. (1996). Qualitative evaluation and scientific citizenship. reflections and refractions. *Evaluation*, 2, 277-289.
- Greene, J. C. (1997). Participatory evaluation. In R.E.Stake & L. Mabry (Eds.), *Advances in Program Evaluation. Vol. 3* (pp. 171-189). Greenwich: Jai Press.
- Habermas, J. (1972). *Knowledge and Human Interests*. (*Erkenntnis und Interesse*). London.
- Habermas, J. (1982). *Theorie des kommunikativen Handelns. 1.Handlungsrationalität und gesellschaftliche Rationalisierung*. Frankfurt am Main.
- Haney, W. M., Madaus, G. F., & Lyons, R. (1995). *The Fractured Marketplace for Standardized Testing*. Boston: Kluwer Academic

Publishers.

- Hansson, F. & Aagaard Nielsen, K. (1996). Dansk sociologis historie. En historie om brud og lidt kontinuitet. In F.Hansson & K. Aagaard Nielsen (Eds.), *Dansk sociologis historie* (pp. 9-36). København: Forlaget Sociologi.
- Hansson, F. (1997). Evaluation Traditions in Denmark. Critical Comments and Perspectives. *Evaluation*, 3, 85-96.
- Hansson, F. (1998). Evaluation Research and Sociology in Denmark. A Tale of Two Cultures. *New Directions for Evaluation*, Spring 1998, 55-71.
- Hirdman, Y. (1994). Social engineering and the woman question.Sweden in the thirties. *Studies in political economy*, S. 73-94.
- House, E. R. (1993). *Professional Evaluation. Social impact and Political Consequences*. Newbury Park: Sage.
- House, E. R., Haug, C., & Norris, N. (1996). Producing evaluations in a large bureaucracy. *Evaluation*, 2, April, 135-150.
- House, E. R. & Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, Calif.: Sage Publications.
- House, E. R. (2005). Qualitative Evaluation and Changing Social Policy. In N.K.Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 1069-1081). London: Sage Publications Ltd.
- Krogstrup, H. K. (2001). Brugerinddragelse i evaluering.top-down og bottom-up perspektiver. In P.Dahler-Larsen & H. K. Krogstrup (Eds.), *Tendenser i evaluering* (Odense: Odense Universitetsforlag.
- Kushner, S. (2005). Qualitative Control: A Review of the Framework for Assessing Qualitative Evaluation. *Evaluation*, 11, 111-122.
- Lewin, K. (1945). The Research Center for Group Dynamics at Massachusetts Institute of Technology. *Sociometry*, 8, 126-136.
- Lindblom, C. E. (1959). The Science of "Muddling Through". *Public Administration Review*, XIX, 79-88.
- Lindblom, C. E. & Cohen, D. K. (1979). *Usable knowledge. social science and social problem solving*. new haven, conn..

- Lomborg, B. (1998). Metodologisk vurdering af borger/brugerundersøgelser. *Nordisk administrativt tidsskrift*, 93-120.
- Love, A. J. (1991). *Internal evaluation. building organizations from within*. London: Sage.
- Maxwell, N. L. & Lopus, J. S. (1994). The Lake Wobegon Effect in Student Self-Reported Data. *The American Economic Review Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association*, 84, 201-205.
- Mjøset, L. (2003). Sju teser om positivismekritikkens utfordringer ved århundreskiftet. *Sosiologisk tidsskrift*, 11, 23-44.
- Mouritsen, J. (1997). *Tællelighedens regime. Synlighed, ansvarlighed og økonomistyring gennem mål og rammer i statslige institutioner*. København: Jurist- og Økonomiforbundets Forlag.
- Mulkay, M., Pinch, T., & Ashmore, M. (1987). Colonizing the Mind: Dilemmas in the Application of Social Science. *Social Studies of Science*, 17, 231-256.
- Nepper Larsen, S. (2001). Evalueringsfeber. *Dansk Sociologi*.
- Nonaka, I. & Takeuchi, H. (1995). *The knowledge-creating company. How japanese companies create the dynamics of innovation*. New York: Oxford University Press.
- Olsen, H. (1997). Tal taler ikke uden ord. *Politica*, 29, 295-310.
- Olsen, H. (2002). Kvaler eller kvalitet? *Nordisk Psykologi*, 54, 145-172.
- Patton, M. Q. (1997). Toward distinguishing empowerment evaluation and placing it in a larger context. *Evaluation Practice*, 18, 147-163.
- Peters, T. & Gore, A. (1993). *Creating a government that works better & costs less : the report of the National Performance Review*. Washington D.C.: Plume.
- Polanyi, M. (1983). *The Tacit Dimension*. Gloucester, Mass.: Peter Smith / Doubleday.
- Pollitt, C. (1996). Justification by works or by faith?.Evaluating the New Public Management. *Evaluation*, 1, 133-154.
- Pollitt, C. & Summa, H. (1997). Trajectories of Reform: Public Management Change in Four Countries. *Public Money and*

Management, 17, 7-18.

- Pollitt, C. & Bouckaert, G. (2000). *Public management reform.a comparative analysis*. Oxford: Oxford University Press.
- Popper, K. (1995). *The Open society and its Enemies*. London: Routledge.
- Porter, T. M. (1992). Quantification and the Accounting Ideal in Science. *Social Studies of Science, 22*, 633-652.
- Porter, T. M. (1995). *Trust in Numbers. The pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Power, M. (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Quine, W. V. (1996). Epistemology naturalized. In P.K.Moser (Ed.), *Empirical knowledge - readings in contemporary epistemology* (Landham: Rowman & Littlefield.
- Quine, W. V. O. (1964). *From a logical point of view*. Cambridge: Harvard University Press.
- Rescher, N. (1995). *Public concerns. Philosophical studies of Social Issues*. Lanham: Rowman & Littlefield.
- Rossi, P. H., Freeman, H. E., & Sandefur, G. D. (1993). *Workbook for evaluation.a systematic approach*. London: Sage.
- Ryan, K. & Johnson, T. (2000). Democratizing Evaluation: Meanings and Methods from Practice. *New Directions for Evaluation, 85*, 39-50.
- Schwandt, T. A. (1997). Evaluation as Practical Hermeneutics. *Evaluation, 5*, 69-83.
- Schwandt, T. A. (2001). A Postscript on Thinking about Dialogue. *Evaluation, 7*, 264-276.
- Schwandt, T. A. (2003). 'Back to the Rough Ground!' Beyond Theory to Practice in Evaluation. *Evaluation, 9*, 353-364.
- Scriven, M. (1997). Empowerment evaluation examined. *Evaluation Practice, 18*, 165-175.
- Shadish, W. R., Chacon-Moscoso, S., & Sanchez-Meca, J. (2005). Evidence-based Decision Making: Enhancing Systematic Reviews of Program Evaluation Results in Europe. *Evaluation, 11*, 95-109.

- Simon, H. A. (1965). *Administrative behavior. a study of decision-making processes in administrative organization*. New York: MacMillan.
- Simon, H. A. (1982). *Economic Analysis and Public Policy*. Cambridge, Mass.: The MIT Press Ltd.
- Simon, Herbert (1992). *Economics, bounded rationality and the cognitive revolution*. Aldershot: Edward Elgar.
- Spencer, L., Ritchie, J., Lewis, J., & Dillon, L. (2003). *Quality in Qualitative Evaluation: a Framework for Assessing Research Evidence*. London: Governments Chief Social Researcher's Office.
- Stehr, N. (1994). *Knowledge Society*. London: Sage.
- Toulmin, Stephen and Gustavsen, Björn (1996). *Beyond theory. Changing organizations through participation*. Amsterdam: John Benjamins.
- Torres, R. T., Preskill, H. S., & Piontek, M. E. (1997). Communicating and reporting: Practices and concerns of internal and external evaluators. *Evaluation Practice*, 18, 105-125.
- Tsoukas, H. & Vladimirou, E. (2001). What is organizational knowledge? *Journal of Management Studies*, 38, 973-993.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick: Transaction Publishers.
- Wargo, M. J. (1995). The impact of federal government reinvention on federal evaluation activity. *Evaluation Practice*, 16, 227-237.
- Weiss, C. H. (1979). The many meanings of Research utilization. *Public Administration Review*, 426-431.
- Weiss, C. H. & Bucuvalas, M. (1980). *Social Science Research and Decision-making*. New York.
- Weiss, C. H. (1997). Theory-Based Evaluation: Past, Present, and Future. *New Directions for Evaluation*, 41-56.
- Weiss, C. H. (1999). The interface between evaluation and public policy. *Evaluation*, 5, 468-486.
- Whitley, R. (1984). *Intellectual and social organization of the sciences*. Oxford: Clarendon Press.
- Whitley, R. (1995). Academic Knowledge and Work Jurisdiction in

Management. *Organization Studies*, 16, 81-105.

Wright, C. (1968). Evaluation Research. In D.Sills (Ed.), *International Encyclopedia of the Social Sciences*, vol. 5. *Elec to Freq* (pp. 198). New York: MacMillan & Free Press.