

**DANSKE BIDRAG TIL DEN
FAGFÆLLEBEDØMTE
EVALUERINGSFORSKNING**

**MORGENDAGENS VEJLEDENDE ETISKE
PRINCIPPER**

**ER PROGRAMTEORIBASEREDE
EVALUERINGER I STAND TIL AT
VURDERE EFFEKTER AF OFFENTLIGE
INDSATSER?**

DANSKE BIDRAG TIL DEN FAGFÆLLEBEDØMTE EVALUERINGSFORSKNING

INTRODUKTION

Evaluering har som et felt været under rivende udvikling i det seneste årti. Globalt er der blevet etableret nationale evalueringsselskaber i en lang række lande, ligesom der er blevet etableret samarbejdsstrukturer på europæisk og globalt plan.

Flere af disse evalueringsselskaber har vedtaget professionelle standarder eller retningslinjer for evalueringspraksis, der hidrører forhold som anvendelse, gennemførlighed, (metodisk) præcision og etik. Ligeledes har evalueringsselskaberne i Japan og Canada etableret kredencialiserings-programmer, hvor evaluatore kan få deres professionelle kompetencer anerkendt. I andre lande, herunder Danmark, er der oprettet masteruddannelser i evaluering, som også er tegn på professionalisering af evaluatore.

Hidtil har disse forsøg på en professionalisering af evaluering som en særegen disciplin ikke fået et markant fodfæste. En profession karakteriseres ved at have sin egen vidensbase, at have kontrol med viden og ikke mindst en form for monopol på produktion af ny viden.

Selvom evaluering netop kan kendetegnes ved at være tværdisciplinær og betydelige dele af teorier, metoder og standarder kommer fra andre videnskabelige discipliner, så er der tegn på debatter omkring evalueringsteori, metode og vurdering koncentreret i en række evalueringstidsskrifter (Christie & Nesbitt Fleischer, 2010; Heberger, Christie & Alkin, 2010). Tidsskrifter som således er dagsordenssættende for den fremadrettede debat inden for evaluering.

I det seneste nummer af *American Journal of Evaluation* (vol. 32, no. 4) analyserer en række førende tænkere inden for evaluering en række af de tendenser, som forventeligt også vil præge evalueringfeltet i

fremtiden.

Én af tendenserne er, ikke overraskende, en stigende globalisering (Hwalek, 2011) og dermed tilgængelighed og udveksling af evalueringssviden gennem fx internettet (Kistler, 2011). Hvor denne udveksling i stigende grad benytter sig af webinar, blogs, sociale medier mv. som kanaler i den globale landsby, så er det stadig det skrevne ord, som er central for udvekslingen af vor praksis. Og, hævder vi, især den viden som kommunikeret gennem fagfællebedømte kanaler.

Både "professionaliseringen" og globaliseringen af evaluering kan også siges at være udtrykt gennem det stigende antal dedikerede internationale evalueringstidsskrifter, der er blevet etableret i det sidste årti. Det er yderligere blevet påpeget, at den internationale dialog i forskning foregår typisk igennem publikationer; især videnskabelige, fagfællebedømte tidsskrifter er i stigende grad blevet centrale (Canagarajah, 2002: 32-37). Samtidig er engelsk blevet det dominerende sprog for udveksling inden for den internationale forskning (Flowerdew, 1999a, 1999b; Short, Boniche, Kim, & Li, 2001).

I tråd med disse internationale tendenser kan det herhjemme konstateres, at basismidler til danske universiteter siden 2009 tildeles efter universiteternes publicering af forskning i fagfællebedømte forskningspublikationer. Dette dækker såvel nationale som internationale publikationer. Ifølge aftalen mellem stat og universitet omfattes både artikler i tidsskrifter og bøger i denne opgørelse¹. Det vil med andre ord sige, at der er skabt et markant incitament til at publicere igennem fagfællebedømte kanaler.

Af mindst to årsager er det således interessant at undersøge, hvorledes danske evaluatore og evaluate

ringsforskere har bidraget til den internationale evalueringforskning. For det første, som et udtryk for en "professionalisering" der muligvis er undervejs inden for evaluering herhjemme og for det andet for at kortlægge den vidensbase, der udgør det danske bidrag til den internationale evalueringforskning og den dialog, der pågår her.

I denne artikel har vi derfor kortlagt danske evaluators bidrag til fagfællebedømte evalueringstidsskrifter i det seneste årti.

AFGRÆNSNING OG METODE

Vi har foretaget en række afgrænsninger og metodiske valg:

Hvad

Vi har afgrænset kortlægningen til fagfællebedømte evalueringstidsskrifter. Fagfællebedømte bøger er således ikke med i opgørelsen². Ligeledes er fagfællebedømte konferencepapirer fra fx American Evaluation Association ej heller inkluderet.

Vi har i opgørelsen udelukkende medtaget peer-reviewed artikler, ikke boganmeldelser o.lign., som ikke er genstand for anonymiseret bedømmelse.

Hvem

I kortlægningen medtager vi alle forfattere, der er tilknyttet en dansk institution. Disse tæller som *danske bidrag*.

Hvor

Det er ikke entydigt at afgrænse, hvad der konstituerer et evalueringstidsskrift, da det ikke er klart, hvad der definerer og derved afgrænser evaluering. Vi har taget udgangspunkt i (i) fagfællebedømte, (ii) internationale tidsskrifter, (iii) at evaluering er et hovedformål. Vi har herefter sammenlignet med Díaz-Puente, Cazorla & Dorrego (2007) og Christie & Nesbitt Fleischer (2010), der foretager tilsvarende afgrænsninger. Følgende tidsskrifter er medtaget i kortlægningen:

American Journal of Evaluation
Canadian Journal of Program Evaluation
Education Evaluation and Policy Analysis Evaluation
Evaluation and Program Planning

Evaluation and the Health Professions
Evaluation Review
Evidence & Policy (2005-)
Journal of Development Effectiveness (2009-)
Journal of Multidisciplinary Evaluation (2004-)
Measurement and Evaluation in Counseling and Development (2009-)
New Directions for Evaluation
Practical Assessment, Research & Evaluation

Vi har kortlagt alle publikationer i perioden 2000-2011 (oktober).

Hvordan

Der er foretaget en manuel søgning på følgende måde:

Elektroniske søgefunktioner i de forskellige tidsskrifter er anvendt.

Vi har manuelt gennemgået hver enkelt tidsskrifts issue inden for den definerede tidsperiode.

Alle titler, forfattere, og institutionelt tilhørsforhold er registreret i en database.

Metodens begrænsninger er følgende:

- (i) Evaluering og evalueringstidsskrifter kan ikke entydigt afgrænses. Der kan således være andre tidsskrifter, som, man kan argumentere for, burde inkluderes.
- (ii) Vi har ikke kortlagt artikler med evalueringssindhold, der er publiceret i andre internationale fagfællebedømte forskningstidsskrifter inden for særlige samfundsvidenskabelige discipliner eller domæner.

RESULTATER

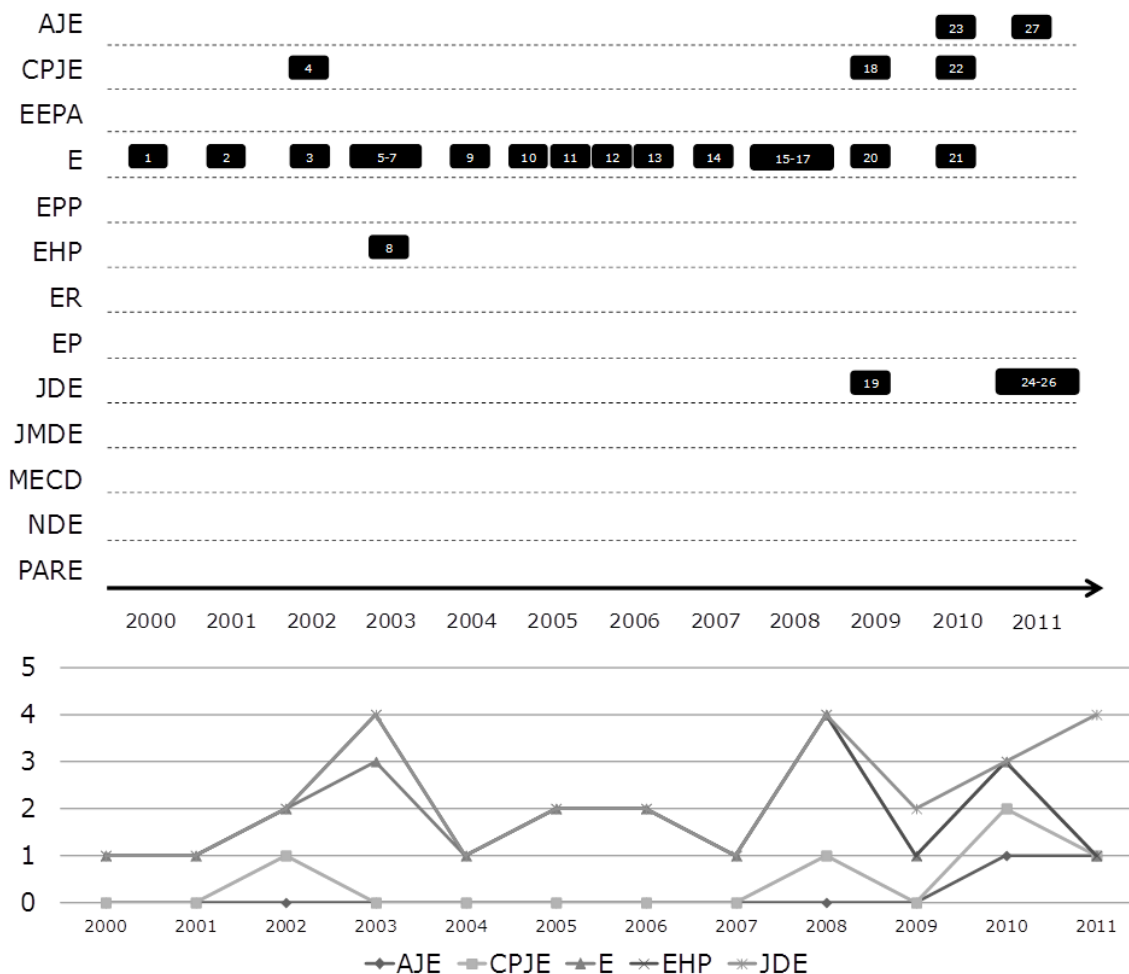
Hvor meget er udgivet?

Kortlægningen viser, at danske evaluators i perioden 2000-2011 har bidraget til 27 peer-reviewed artikler i de ovenfor nævnte tidsskrifter. Dette er ensbetydende med ca. to udgivelser per år (se bilag).

Et overblik fremgår af figuren nedenfor:

Figur 1 – oversigt over danske forfatteres publikationer i peer-reviewed evalueringstidsskrifter 2000-2011

VALG AF TIDSSKRIFTER



American Journal of Evaluation (AJE), Canadian Journal of Program Evaluation (CPJE), Education Evaluation and Policy Analysis (EEPA), Evaluation (E), Evaluation and Program Planning (EPP), Evaluation and the Health Professions (EHP), Evaluation Review (ER), Evidence & Policy (EP), Journal of Development Effectiveness (JDE), Journal of Multidisciplinary Evaluation (JMDE), Measurement and Evaluation in Counseling and Development (MECD), New Directions for Evaluation (NDE), Practical Assessment, Research & Evaluation (PARE)

VALG AF TIDSSKRIFTER

Samtidig viser kortlægningen, at danske forfattere kun har bidraget til 5 af de 13 undersøgte tidsskrifter. Langt størstedelen af artiklerne er publiceret i tidsskriftet *Evaluation. The International Journal of Theory, Research and Practice*. Hele 17 af 27 (63 %) artikler er således publiceret her. *Evaluation* er et af de relativt få tidsskrifter, der er publiceret i Europa med støtte fra European Evaluation Society.

Siden sin etablering i 2009 har *Journal of Development Effectiveness* haft et markant dansk islæt med hele fire bidrag.

Kun seks artikler er blevet publiceret i nordamerikanske tidsskrifter siden 2000, heraf de 3 inden for de seneste to år. Díaz-Puente et al (2007) har vist, at der er en markant skævhed i de nordamerikanske tidsskrifters publikationsmønster, hvor ikke-nordamerikanske bidrag udgør en mindre andel.

Dette mønster bekræftes af danske evaluators publikationsmønster³.

HVEM HAR PUBLICERET?

I alt har 32 forskellige evaluatører bidraget til de 27 artikler.

Ingen dansk evaluator har bidraget med en markant andel af det samlede antal publikationer. Den forfatter med flest bidrag er Olaf Rieper med fire publikationer. Alle er publiceret i tidsskriftet *Evaluation*.

Kun syv evaluatører har bidraget til mere end én artikel. Tilsvarende har 25 evaluatører kun bidraget til en enkelt artikel.

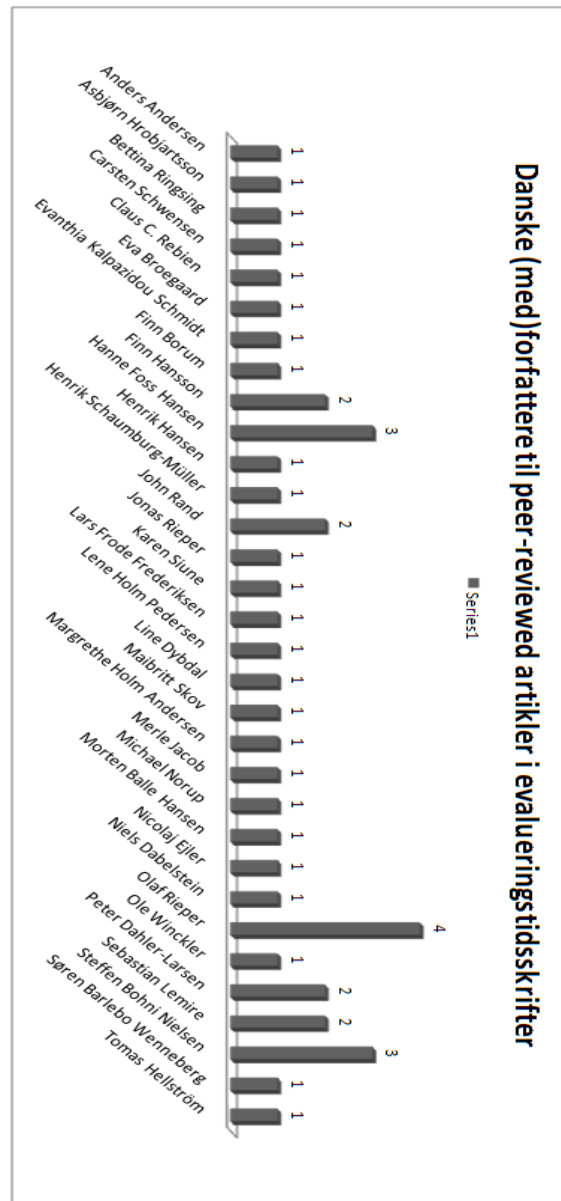
Steffen Bohni Nielsen er den evaluator, der har haft størst diversitet med publikationer i tre forskellige tidsskrifter.

Figur 2 (modsat side)- Danske (med)forfattere til peer-reviewed artikler i evalueringstidsskrifter 2000-2011

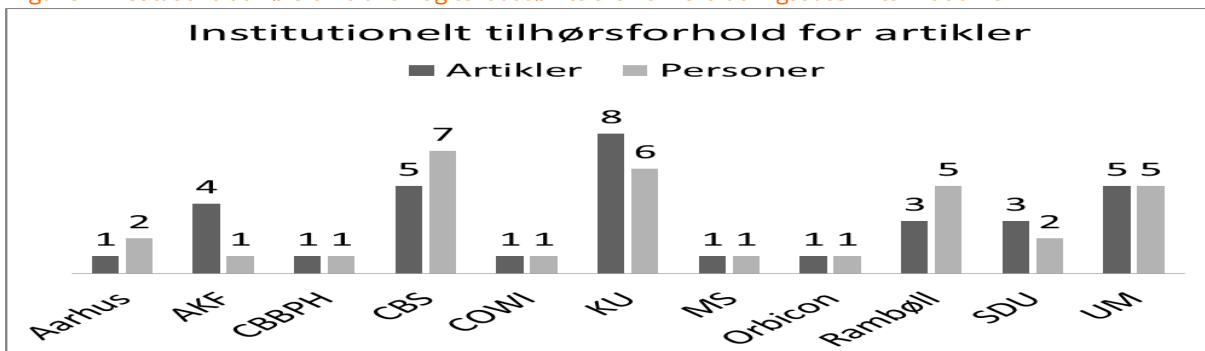
HVOR KOMMER EVALUATORERNE FRA?

Universiteterne har ikke overraskende bidraget med flest publikationer. 17 artikler har bidrag fra universitetsforskere. Forfattere fra hhv. offentlige institutioner, rådgivningsvirksomheder begge bidraget til fem artikler hver, mens sektorforskningsinstitutioner har bidraget til 4. Én artikel er skrevet af evaluator med tilhørsforhold til en NGO.

I forhold til institutioner er det Københavns Universitet, der har bidraget til flest artikler (dog fra flere forskellige institutter), mens det er Copenhagen Business School, hvorfra flest forskellige evaluatører har bidraget. Det er markant, at alle bidrag fra offentlige institutioner kommer fra Udenrigsministeriet (Danida).



Figur 3 - Institutionelt tilhørsforhold for fagfællebedømte artikler i evalueringstidsskrifter 2000-2011



STØRRE PRODUKTIVITET I EVALUERINGS- TIDSSKRIFTER

Når man ser på antallet af artikler, så er det øget betydeligt i anden halvdel af perioden (2000-2005: 11 artikler, 2006-2011: 16 artikler). Især har rådgivningsvirksomheder øget deres bidrag fra én til fire publikationer, hvilket forklarer størstedelen af variationen.

Når man ser på betydningen af de tidsskrifter, som danske evaluatore har publiceret i, så er der to kilder, der kan informere herom. Den første er Forsknings- og Innovationsstyrelsens (FI) autoritetsliste, hvor et udvalg inden for hver videnskabelig disciplin angiver, om et tidsskrift er niveau 1 eller niveau 2. Niveau 2-tidsskrifter anses for mere prestigøse og sværere at få optaget artikler i.

Det anden er Thomson Reuters årlige Journal Citation Report, hvori ca. 9500 fagfælle-bedømte tidsskrifter inden for natur- og samfundsvidenskaberne er inkluderet. Et tidsskrifts impact factor beregner det gennemsnitlige antal referencer til artikler i et givent tidsskrift. Impact faktoren er en proxy-indikator for et tidsskrifts vigtighed. Ikke alle benævnte tidsskrifter indgår i Thomson Reuters rapport, jf. tabel 1.

Tabel 1 – Peer-reviewed evalueringstidsskrifters niveau, impact factor og ranking

Tidsskrift navn	Akronym	Niveau ⁴	Område	Impact factor ⁵	Ranking ⁶
American Journal of Evaluation	AJE	1	Statskundskab	1.157	17 out of 83 in Social Sciences, Interdisciplinary
Canadian Journal of Program Evaluation	CPJE	1	Psykologi	N/A	N/A
Education Evaluation and Policy Analysis	EEPA	2	Pædagogik, uddannelse, didaktik	1.919	14 out of 177 in Education & Educational Research
Evaluation	E	1	Statskundskab	N/A	N/A
Evaluation and Program Planning	EPP	2	Sociologi, socialt arbejde	0.711	N/A
Evaluation and the Health Professions	EHP	1	Folkesundhed	1.208	49 out of 71 in Health Care Sciences & Services
Evaluation Review	ER	1	Psykologi	0.840	29 out of 83 in Social Sciences, Interdisciplinary
Evidence & Policy (2005-)	EP	1	Statskundskab	N/A	N/A
Journal of Development Effectiveness (2009-)	JDE	N/A	Samfundøkonomi	N/A	N/A
Journal of Multidisciplinary Evaluation (2004-)	JMDE	N/A	N/A	0.850 ⁷	N/A
Measurement and Evaluation in Counseling and Development (2009-2011)	MECD	1	Psykologi	0.903	45 out of 67 in Psychology, Applied and 31 out of 50 in Psychology, Educational
New Directions for Evaluation	NED	1	Pædagogik, uddannelse, didaktik	N/A	N/A
Practical Assessment, Research & Evaluation	PARE	1	Pædagogik, uddannelse, didaktik	N/A	N/A

Som det fremgår af tabel 1, er der en lidt kuriøs diskrepans mellem FIs niveaudeling og Thomson Reuters impact factor. Fx er et multidisciplinært evalueringstidsskrift som Evaluation & Program Planning rangeret som niveau 2 i én disciplin, mens American Journal of Evaluation er rangeret som niveau 1 i en anden disciplin, men med en markant højere impact factor.

Derudover kan det konstateres, at Evaluation, der er den primære kanal for danske evaluators publikationer, end ikke er medtaget i Thomson Reuters rapport. Desuden kan det konstateres, at tidsskrifter med en klar domæneprofil (Evaluation and the Health Professions og Education Evaluation and Policy Analysis har en højere impact factor end de tværdisciplinære tidsskrifter).

EN BETYDELIG VARIATION I EMNER

Når man ser på artiklernes indhold, så kan det konstateres, at størstedelen af artiklerne har et empirisk indhold. Dog er det ikke nødvendigvis det primære formål at afrapportere fund fra konkrete evalueringer. Det empiriske indhold er nærmere en case for at udvikle og/eller diskutere en teoretisk eller metodologisk problematik. Dette ligger fint i tråd med de fleste evalueringstidsskrifters retningslinjer.

Artiklerne har en betydelig sektorspredning. Syv artikler er skrevet med udgangspunkt i en bistandskontekst. Fem artikler handler om aspekter af evaluering af forskning og forskningsinstitutioner.

Enkelte artikler har ikke en konkret sektorspecificering – mens de øvrige artikler er spredt ud over en række sektorer.

Tematisk er anvendelse af og i evaluering - en meget bred kategori – det primære indhold i syv af de 27 artikler. Aspekter og problematikker vedrørende effektvurdering er ligeledes godt repræsenteret, særligt Journal of Development Effectiveness, hvilket giver mening qua tidsskriftets formål.

Endelig er der også flere bidrag om teoribaseret evaluering.

Tabel 2 – Overordnet emne i peer-reviewed artikler 2000-2011

Emne	Antal artikler
Evaluering og advocacy	1
Resultatbaseret styring	1
Evidensbaseret politik og praksis	1
Generel evalueringsteori	2
Evalueringens rolle i governance	3
Evalueringens kapacitetsopbygning	3
Teoribaseret evaluering	4
Effekt-evaluering	5
Anvendelse af og i evaluering	7
Total	27

HVORDAN KLARER VI OS I FORHOLD TIL ANDRE?

Hvis vi sammenligner antallet af artikler af danske forfattere i forhold til landene omkring os kan vi konstatere, at vi ligger på niveau med Sverige og foran Tyskland, Norge og Finland.

Tabel 3 – Sammenligning med peer-reviewed publikationer i landene omkring os 2000-2011 (se næste side)

Forfatteres tilhørsforhold til land og institution, fagfællebedømte artikler i evalueringstidsskrifter i perioden 2000-2011						
Land	Universitet	Offentlig Institution	Rådgivningsvirksomhed	Anden forskningsinstitution	NGO	*Total
DE	15	0	3	2	0	20
DK	17	5	5	4	1	32
FI	8	5	4	4	0	21
NO	7	3	2	3	0	15
SE	26	5	0	2	0	33

*Antallet overstiger antallet af faktiske artikler, fordi artikler har flere forfattere. Samtidig kan artikler være registreret i forskellige kategorier fx er første forfatter fra et universitet og anden forfatter er fra en NGO. Artiklen tæller således i begge kategorier.

BLIVER DET LÆST OG ANVENDT?

Som vist ovenfor er størsteparten af de danske bidrag blevet publiceret i tidsskriftet, *Evaluation*. Vi har derfor indhentet data fra Sage, der udgiver tidsskriftet, om hvilke 100 artikler er de mest læste gennem tiderne⁸. Dette kan blive et indblik i det omfang danske bidrag til evalueringsslitteraturen rent faktisk bliver læst.

Ifølge Sage er fem af de 17 danske bidrag at finde på denne liste. Det er især værdt at bemærke, at to danske bidrag placerer sig i top 10 over mest læste artikler.

"yngste" artikel i top 20. Ingen andre nordiske artikler er at finde i top ti.

Hvis vi ser Sage liste over mest 50 citerede artikler fra *Evaluation*, i dette og andre tidsskrifter udgivet af Sage, er det imidlertid andre artikler som dukker op. Her er Peter Dahler-Larsens artikel om teori-baseret evaluering (2001) placeret som nr. 15, mens Kim Forss og Claus Rebiens artikel om processuel brug af evaluering (2002) er placeret på en 23. plads⁹.

Rang	Titel	Forfattere	Antal downloads
1	Choosing Evaluation Models: A Discussion on Evaluation Design	Hanne Foss Hansen (2004)	4768
10	Improving Performance?: Exploring the Complementarities between Evaluation and Performance Management	Steffen Bohni Nielsen & Nicolaj Ejler (2008)	2958
40	Knowledge without Goals? Evaluation of Knowledge Management Programmes	Tomas Hellstrom (2003)	1598
62	Organizational Use of Evaluations: Governance and Control in Research Evaluation	Finn Hansson (2005)	1392
97	Learning about Advocacy: A Case-Study of Challenges, Everyday Practices and Tensions	Bettina Ringsing (2008)	1055

Hanne Foss Hansens artikel (2004) om valg af evalueringsmodeller er således med afstand den mest læste artikel i *Evaluation*. Steffen Bohni Nielsen og Nicolaj Ejlers (2008) artikel er ligeledes i top ti og er den

(Kilde: *Evaluation*. Sage Publication, december 2011.)

KONKLUSION OG PERSPEKTIVER

I en analyse af diskussioner inden for evalueringsforskningen vedrørende nytte, metode og vurdering fandt Heberger, Christie og Alkin at;

“...evaluation theorists draw upon the intellectual work of a diverse set of fields, [but] evaluation authors often publish their work in evaluation-related journals such as American Journal of Evaluation, Evaluation Review, and Evaluation and Program Planning. These journals are included in interdisciplinary and multidisciplinary subject categories, offering evidence for evaluation as its own discipline as well as further supporting the transdisciplinary nature of evaluation.” (Heberger, Christie & Alkin, 2010: 39)

Det kan konstateres, at:

- * Danske evaluators bidrag til den internationale evalueringsviden gennem fagfællebedømte evalueringstidsskrifter ligger på niveau med Sverige og foran Norge, Finland og Tyskland.
- * Evaluators fra andet end forskningsinstitutioner har bidraget til ca. 1/3 af artiklerne.
- * Evaluators fra universiteterne har bidraget til ca. 2/3 af artiklerne.
- * Antallet af publikationer er øget og diversificeret i anden halvdel af perioden.
- * Der er betydelig emne- og sektormæssig spredning. Tidsskriftet *Evaluation* er markant mest anvendt som kanal for danske evaluators.
- * At to danske artikler er repræsenteret blandt de ti mest læste artikler i *Evaluation* og fem artikler er placeret i top 100 mest læste.

Samtidig peger et blot kursorisk kig på publikationslisterne hos vores fremmeste evalueringsforskere, at især bidrag til internationale forskningsantologier er en anden markant kanal, som anvendes.

Samlet set viser kortlægningen, at der er en tilsyneladende stigende produktion af internationale, fagfællebedømte artikler, og at forfatterne af disse i højere

grad end tidligere afspejler den diversitet, der er blandt de, der beskæftiger sig med evaluering i Danmark.

At en del af vidensproduktionen foregår uden for universiteterne, kan indikere en stigende professionel bevidsthed og internationalt udsyn blandt danske evaluators. Spørgsmålet er, om dette også vil manifestere sig i de institutionelle rammer og professionelle standarder, vi stiller til os selv som evaluators.

ARTIKLER MED DANSKE (MED) FORFATTERE

- Finn Borum**, CBS, **Hanne Foss Hansen** (2000), University of Copenhagen. The Local Construction and Enactment of Standards for Research Evaluation. The Case of the Copenhagen Business School. *Evaluation*, vol. 6, no. 3, pp. 281-299.
- Peter Dahler-Larsen** (2001). University of Southern Denmark. From Programme Theory to Constructivism On Tragic, Magic and Competing Programmes. *Evaluation*, vol. 7, no. 3, pp. 331-349.
- Kim Forss. Andante-Tools for thinking, Sweden.
- Claus C. Rebien**, COWI, Denmark (2002). Process Use of Evaluations Types of Use that Precede Lessons Learned and Feedback. *Evaluation*, vol. 8, no. 1, pp. 29-45.
- Anders Andersen** (2002). The Danish ministry of foreign Affairs, Copenhagen. Evaluating organizational capacity development. *Canadian Journal of Program Evaluation*, vol. 17, no. 2, pp. 121-150.
- Lars Frode Frederiksen, Finn Hansson, Søren Barlebo Wenneberg** (2003), CBS, Denmark. The Agora and the Role of Research Evaluation, *Evaluation*. vol. 9, no. 2, pp. 149-172.
- Niels Dabelstein** (2003). Danida. Evaluation Capacity Development: Lessons Learned. *Evaluation*, vol. 9, no. 3, pp. 365-369.
- Tomas Hellström** og **Merle Jacob** (2003). CBS, Denmark. Knowledge without Goals? Evaluation of Knowledge Management Programmes. *Evaluation*, vol. 9, no. 1, pp. 55-72.

- Asbjørn Hróbjartsson** (2003). The Nordic Cochrane Centre, University of Copenhagen. **Michael Norup**, University of Copenhagen. The Use of Placebo Interventions in Medical Practice – A National Questionnaire Survey of Danish Clinicians. *Evaluation and the Health Professions*, vol. 26, no. 2, pp. 153-165.
- Olaf Rieper** (2004). AKF, Institute of Local Government Studies, Denmark. Exploring Popper's Relevance for the Evaluation Community. *Evaluation*, vol. 10, no. 1, pp. 92-100.
- Hanne Foss Hansen**. University of Copenhagen (2004). Choosing Evaluation Models A Discussion on Evaluation Design. *Evaluation*, vol. 11, no. 4, pp. 447-462.
- Henrik Schaumburg-Müller** (2004). CBS, Denmark. Use of Aid Evaluation from an Organizational Perspective. *Evaluation*, vol. 11, no. 2, pp. 207-222.
- Peter Dahler-Larsen** (2005). University of Southern Denmark at Odense, Denmark. When Evaluation Meets the 'Rough Ground' in Communities. *Evaluation*, vol. 12, no. 4, pp. 496-505.
- Finn Hansson** (2005). CBS, Denmark. Organizational Use of Evaluations Governance and Control in Research Evaluation. *Evaluation*, vol. 12, no. 2, pp. 159-178.
- Olaf Rieper**, AKF. **Jonas Rieper**, Centre for Bridge Building in Health Care, Denmark. Laila Lounso, University, Tromsø, Norway (2007). Evaluative Feedback as a Contribution to Learning between Groups of Professionals. *Evaluation*, vol. 13, no. 3, pp. 306-322.
- Lene Holm Pedersen**. University of Copenhagen. **Olaf Rieper**, AKF (2008). *Is Realist Evaluation a Realistic Approach for Complex Reforms?* *Evaluation*, vol. 14, no. 3, pp. 271-293.
- Bettina Ringsing**. Danish Association for International Co-operation (2008), Denmark. Learning about Advocacy. A Case-Study of Challenges, Everyday Practices and Tensions. *Evaluation*, vol. 14, no. 4, pp. 413-336.
- Steffen Bohni Nielsen** og **Nicolaj Ejler**. Ramboll Management Consulting (2008). Improving Performance? Exploring the Complementarities between Evaluation and Performance Management. *Evaluation*, vol. 14, no. 2, pp. 171-192.
- Evanthia Kalpazidou Schmidt** og **Karen Siune** (2008). University of Aarhus, Denmark. Evaluating inter-, multi-, and transdisciplinary research in the European Research Area. *Canadian Journal of Program Evaluation*, vol. 23, no. 1, pp. 179-200.
- John Rand**. Department of Economics, University of Copenhagen, Denmark (2009). Impact of an aquaculture extension project in Bangladesh. *Journal of Development Effectiveness*, vol. 1, no. 2, pp. 130-146.
- Hanne Foss Hansen**. University of Copenhagen. **Olaf Rieper**, AKF, Danish Institute of Governmental Research, Denmark (2009). The Evidence Movement. The Development and Consequences of Methodologies in Review Practices. *Evaluation*, vol. 15, no. 2, pp. 141-163.
- Margrethe Holm Andersen**. Ministry of Foreign Affairs, Denmark (2010). Involving Developing Countries in the Evaluation of Development Cooperation: The Case of a Joint Evaluation of Development Cooperation between Ghana and Denmark. *Evaluation*, vol. 16, no. 3, pp. 295-308.
- Line Dybdal**, **Steffen Bohni Nielsen** og **Sebastian Lemire**. Ramboll Management Consulting (2010). Contribution Analysis Applied: Reflections on Scope and Methodology. *Canadian Journal of Program Evaluation*, vol. 25, no. 2, pp. 29-57.
- Morten Balle Hansen**. University of Southern Denmark. Evert Vedung. Uppsala University, Uppsala and Gävle, Sweden (2010). Theory-based Stakeholder Evaluation. *American Journal of Evaluation*, vol. 31, no. 3, pp. 295-313.
- Henrik Hansen**. Institute of Food and Resource Economics, University of Copenhagen, Denmark. **Ole Winckler Andersen**. Ministry of Foreign Affairs, Copenhagen, Denmark (2011). Impact evaluation of infrastructure interventions. *Journal of Development Effectiveness*, vol. 3, no. 1, pp. 1-8.

Eva Broegaard. Danidas Evaluation Department, Ministry of Foreign Affairs of Denmark, Copenhagen, DK, Denmark. **Carsten Schwensen.** Orbicon, DK, Ballerup, Denmark (2011). Experience from a phased mixed-methods approach to impact evaluation of Danida support to rural transport infrastructure in Nicaragua. *Journal of Development Effectiveness*, vol. 3, no. 1, pp. 9-27.

John Rand. Institute of Food and Resource Economics, University of Copenhagen, Frederiksberg, DK, Denmark (2011). Evaluating the employment-generating impact of rural roads in Nicaragua. *Journal of Development Effectiveness*, vol. 3, no. 1, pp. 28-43.

Steffen Bohni Nielsen, Sebastian Lemire og Maibritt Skov. Ramboll Management Consulting (2011). Measuring Evaluation Capacity - Results and implication of a Danish Study. *American Journal of Evaluation*, vol. 32, no. 3, pp. 323-344.

REFERENCER

Canagarajah, A. S. (2002). *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press.

Christie, C. & [Nesbitt Fleischer](#), D. (2010). Insight Into Evaluation Practice: A Content Analysis of Designs and Methods Used in Evaluation Studies Published in North American Evaluation-Focused Journals. *American Journal of Evaluation*, vol. 31 (3), pp. 326-346.

Díaz-Puente, J.M. Cazorla, A. and Dorrego, A. (2007). Crossing National, Continental, and Linguistic Boundaries : Toward a Worldwide Evaluation Research Community in Journals of Evaluation. *American Journal of Evaluation*, vol. 28 (4), pp. 399-415.

Flowerdew, J. (1999a). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, vol. 8 (3), pp. 243-264.

Flowerdew, J. (1999b). Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, vol. 8 (2), pp. 123-145.

Heberger A.E., Christie, C.A. & Alkin, M.A. (2010). A Bibliometric Analysis of the Academic Influences of and on Evaluation Theorists' Published Works. *American Journal of Evaluation*, vol. 31 (1), pp. 24-44.

Hwalek, M. (2011). The Starts are Aligned for Disruptive Change in Evaluation Practice. *American Journal of Evaluation*, vol. 32 (4), pp. 567-72.

Kistler, S. (2011). Technology, Social Networking and the Evaluation Community, *American Journal of Evaluation*, vol. 32 (4), pp. 567-72.

Short, R., Boniche, A., Kim, Y., & Li, P. L. (2001). Cultural globalization, global English, and geography journals. *The Professional Geographer*, 53 (1), pp. 1-11.

Thomson Reuters. 2011. *2010 Journal Citation Reports*.

NOTER

1: Ministeriet for Forskning, Innovation og Videregående Uddannelser. Aftale om basismidler efter resultat. 30.6.2009. <http://fivu.dk/lovstof/politiske-aftaler/basismidler-efter-resultat>

2: Det kan konstateres, at danske evaluører har bidraget til en lang række bøger omhandlende evaluering, især i INTEVAL-serien udgivet af Transaction Publishers.

3: På seneste AEA-konference (2011) angav Thomas Schwandt, redaktør for American Journal of Evaluation, at i 2011 kommer ca. halvdelen af deres bidrag fra forfattere uden for USA (dvs. inkl. Canada).

4: Forsknings- og Innovationsstyrelsens autoritetsliste for tidsskrifter og forlag, 2010 og 2011.

5: 2010 Journal Citation Reports (Thomson Reuters, 2011).

6: 010 Journal Citation Reports (Thomson Reuters, 2011).

7: Uofficiel impact factor ifølge tidsskriftets hjemmeside.

8: Evaluation har eksisteret siden 1995. Vi antager, i tråd med Sage, at antallet af downloads er en indikator for hvor meget artiklen er læst.

9: <http://evi.sagepub.com/reports/most-cited>

MORGENDAGENS VEJLEDENDE ETISKE PRINCIPPER FOR EVALUERING

Etik i evaluering har været på dagsordenen i over 25 år. Alligevel står det mere end nogen sinde før klart, at vi står på tærsklen til en ny tidsalder, når vi taler om etik. Min argumentation går på, at der er behov for et fokusskifte fra ´etik, som noget, der er knyttet til evaluator baseret på individtænkning´ til ´etik, som er knyttet til selve evalueringen, dvs. som en særlig praksis i et fællesskab´. Det vil jeg uddybe nærmere i denne artikel.

ERKENDELSESTEORI GØR EN FORSKEL

Evaluering er ikke bare evaluering. Afhængig af hvilke teorier og erkendelsesteorier, der organiserer evalueringen, vil formålet og de anvendte metoder være vidt forskellige. Den valgte erkendelsesteori bag evalueringen har afgørende betydning for både medarbejdere og lederes tanke- og handlerum, såvel som evaluators. Evaluering GØR noget ved den måde, hvorpå ledere og medarbejdere forstår sig selv, hinanden og deres arbejdsliv, og har væsentlige effekter¹ af såvel positiv som negativ karakter. Både for det enkelte organisationsmedlem og for livet i organisationer.

Det er derfor ikke ligegyldigt, hvilke teorier og grundlæggende antagelser, vi gør brug af, når vi evaluerer. Der ligger en ofte ubemærket, men meget væsentlig magt i praktiseringen af evaluering, som det er vigtigt vi holder os for øje som evaluator og rekvirent af evalueringer. Næmlig magten til at definere virkeligheden. At praktisere evaluering medfører således i min optik en væsentlig etisk fordring, - nemlig at gøre sig overvejelser om den definitionsmagt, vi bringer i spil, når vi evaluerer. De etiske udfordringer heri knytter an til selve mandatet for evaluering, som til evalueringens instrumentet og – processen.

ETIK ER MERE OG ANDET END ET INDIVIDUELT ANSVAR BASERET PÅ MAVEFORNEMLER

Allerede i 1994 vedtog American Evaluation Association "The Guiding Principles for Evaluator²" som et sæt standarder for enhver evaluator. Disse vejledende

principper er normative og anvisende for, hvad evaluator bør medtænke og vise hensyn til i sit evalueringsarbejde. De har siden hen affødt en massiv udvikling af evalueringsstandarder og diskussioner omkring etik i evaluering rundt om i verden.

Michael Morris³ (2011) efterlyser case-baserede etiske analyser og mere empirisk forskning af etiske dimensioner i evaluering, især som de opleves af de forskellige interessenter, der er i kløerne på en evaluator.

Endvidere udforsker han forståelser af etik i evaluering og organiserer sin undersøgelse i overensstemmelse med AEA´s "The Guiding Principles for Evaluators"; hvor 1. er Systematisk undersøgelse, 2. Kompetencer, 3. Integritet/Ærlighed, 4. Respekt for mennesker og 5. er Ansvar for generel og offentlig velfærd. Perspektivet er her, at enhver evaluator støder på etiske udfordringer og dilemmaer i alle faser i en evaluering. Målet er her, at evaluator gennem udvikling af egne faglige kompetencer og personlige egenskaber, skal blive bedre til at håndtere de moralske dilemmaer, som hendes/hans evalueringsarbejde støder ind i.

Jeg vil imidlertid introducere en tænkning, hvor fokus flyttes fra evaluators personlige egenskaber og umiddelbare mavefornemmelser som middel til at håndtere moralske udfordringer til et fokus på de rammer og muligheder, som *evalueringsspraksissen skaber*. I denne optik giver det ingen mening at tale om vejledende

etiske principper for evaluator, men i stedet om vejledende etiske principper for evaluering, som en særlig praksis, hvor alle parter deler et fælles etisk ansvar baseret på solidarisk refleksivitet og forpligtelsen til dialog⁴.

Også Helen Simons (2006) er på linje med Michael Morris' opfattelse af etik i evaluering, der er bundet op på evaluators personlige egenskaber, valg og vurderinger, men hun lægger mere vægt på betydningen af kompleksitet og kontekst for de etiske overvejelser, som evaluator skal gøre sig. Hun skriver (min oversættelse): "De fleste evaluering- og forskningsartikler, hvor etik er nævnt, begynder med at huske os på, at der ikke findes nogen kontekstfrie abstrakte etiske principper for evaluering, som kan anvendes til at guide handlinger og beslutninger af etisk karakter i evaluering. Det, vi møder i praktisk evalueringssarbejde, er moralske dilemmaer, hvor vi skal foretage komplekse vurderinger, vælge mellem alternative veje til forskellige handlinger, tage myriader af faktorer i betragtning, - sociale, personlige, politiske, kulturelle - sådan som de forekommer i den givne kontekst". Simons har her en vigtig pointe, som jeg gerne vil arbejde videre med ved at vise, hvordan et skift i erkendelsesteori kan få stor betydning for de etiske overvejelser og handlemuligheder.

Virtanen og Laitinen (2004) taler for at løfte etikspørgsmålet op på et højere plan og de skriver (min oversættelse): "Vi sigter mod at placere denne evige udvikling af etiske standarder og retningslinjer i en bredere diskussion af etik i evaluering. Vores pointe er her, at evaluator bør have en stærkere fornemmelse for de moralske og politiske meninger og implikationer ved evalueringsspraksis. Evaluator bør kunne genkende begrænsningerne ved de nuværende konceptualiseringer af standarder og retningslinjer. I denne optik har den hidtidige debat om etiske standarder for evaluering primært været fokuseret på tekniske aspekter.

Sådanne typer af standarder kan være nødvendige, men giver ikke et tilstrækkeligt etisk fundament for ethvert evalueringssarbejde. Alle bør derfor optræde på måder, som sikrer at enhver evaluator viser et

ensartet højt niveau af ansvarlighed." Dertil kunne man spørge, hvordan man kan vurdere, hvad der er et højt og lavt niveau af ansvarlighed? Og hvem er det lige, der vurderer det?

Virtanen og Laitinen introducerer her en tænkning, hvor evaluators etiske forpligtelser rækker ud over de strengt videnskabelige og metodiske (tekniske aspekter), men fastholder samtidig, at det er evaluators evner og fornemmelser for evalueringssarbejdets moralske og politiske meninger og implikationer, der er midlet til håndtering af etiske udfordringer. Fokus er stadig på evaluator som enkeltindivid. Og vi er ikke kommet meget længere eller blevet meget klogere på, hvordan etiske udfordringer håndteres i praksis i evalueringssarbejdet. Der er altså ikke megen hjælp at hente, når vi i sidste ende kun har vores mavefornemmelse at støtte os til, og det angiver på ingen måder et højt ensartet niveau af ansvarlighed.

Jeg har her blot givet tre eksempler på evalueringssfaglige bidrag fra nogle inden for evalueringssfeltet, der er optaget af moral og etik. Diskussionen om etiske udfordringer i evaluering flytter sig ikke væsentligt, hvilket jeg mener, hænger sammen med, at drøftelserne udspiller sig inden for de snævre rammer af realismen som erkendelsesteori og individtænkningen. Men de tre eksempler, jeg her har fremdraget sender et klart budskab. Evaluering har fortsærende store udfordringer med hensyn til etik.

ETIK- OG MORALFORSTÅELSER I EVALUERING

Jeg vil derfor vende perspektivet 180 grader og komme med mit bud på, hvordan vi kan forstå, imødekomme og håndtere etiske overvejelser i evaluering. Mit håb er, at denne artikel belyser en ny tids opfattelse af etik. Samtidig er det min ambition at give et bud på morgendagens vejledende etiske principper for evaluering. Heri ligger en ambition om at gentænke evalueringssbegrebet, evaluatorrollen og evalueringers praktiske anvendelser i organisationer og institutioner. Men lad os lige for en kort stund dvæle ved begreberne etik og moral (Haslebo & Haslebo, 2007).

Etik kan forstås som filosofiske refleksioner over mo-

ralen og handler om overvejelser over, hvad der er en moralsk forsvarlig handling. Der er tale om en kritisk refleksion, der ikke nødvendigvis godtager, hvad dagligdagens moralske overvejelser har ført frem til. Moral, i betydningen "vores skyldighed overfor andre mennesker" siger noget om rigtige og forkerte handlinger og gode og dårlige handlinger og her skelnes mellem en første ordens moral og en anden ordens moral.

Ud fra en første ordens moral skal formuleringen af universelle etiske retningslinjer styre menneskers handlinger og således bidrage til overholdelse af de eksplicite og implicite normer. Normerne kan være "altomfattende" og således sige noget *hvad*, man skal gøre og *på hvilken måde* det skal gøres, *hvornår* det skal gøres og at det skal gøres autentisk og med troværdighed. Blot en lille afvigelse kan vække uro og mistro. Første ordens moral fokuserer på individet, og det anses for vigtigt, at ansvar placeres på enkeltpersoner og først, når man er stillet til regnskab for sine fejl (og mangler/ugerninger) vil man forbedre sig. Fastholdelse af egen moral sker gennem moralsk fordømmelse af personer uden for eget (fag) fællesskab eller af personer, der bryder normerne indenfor eget fællesskab. Sammenstød mellem forskellige førsteordens moralopfattelser kan nemt føre til beskadigelse af relationer og måske ophør af kommunikation.

Anden ordens moral udspringer af en ikke-fundamentalistisk etik, som stræber mod at bevare selve muligheden for at samskabe, hvad der er godt for de implicerede parter. De etiske overvejelser gøres fra et helikopterperspektiv, og ingen handlinger anses for at være onde eller dårlige i sig selv. Handlinger får deres betydningsskabende mening i relationer og kontekster. Det individuelle ansvar erstattes af et relationelt ansvar, hvilket vil sige et fælles ansvar for at bevare muligheden for at koordinere handlinger. Den etiske udfordring er at samarbejde og koordinere handlinger, som genopretter muligheden for at skabe en fælles forståelse med udgangspunkt i en omsorg for relationer, frem for for det enkelte individ.

Min erfaring er, at evalueringspraktikere i stadig sti-

gende grad forventer, at en evaluering bidrager til organisationsfællesskabet og understøtter kerneydelserne, er meningsfulde og relationelt hensigtsmæssige og dermed tager afsæt i en anden ordens moralopfattelse. Min påstand er, at de etiske udfordringer i evalueringsarbejdet bliver langt nemmere at håndtere, når overvejelserne tager afsæt i følgende to ting, der ville springe rammerne for den snævre tænkning:

Etiske overvejelser ses ikke kun som evaluators individuelle ansvar, men tager afsæt i den relationelle etik og altså en andenordens moralopfattelse.

Evalueringen rammesættes eksplicit af den socialkonstruktionistiske erkendelsesteori.

Med dette mener jeg, at socialkonstruktionismen bliver det organiserende princip for evaluering⁵, hvilket betyder at etikbegrebet forstås som relationel og kontekstuel. Etik i Evaluering ud fra den forståelse skærper opmærksomheden for, hvordan en evalueringsproces kan forbedre vilkårene for at samskabe mening og koordinere handlinger på måder, der er nyttige og respektfulde både for flest muligt/alle implicerede parter og for organisationens kerneopgaver.

VEJLEDENDE ETISKE PRINCIPPER FOR EVALUERING PEGER I RETNING AF SAMSKABELSE OG FÆLLES ANSVAR

Et skift til socialkonstruktionistisk erkendelsesteori betyder et skift fra fokus på individ til et fokus på relationer og kontekst. Det betyder, at AEA's vejledende principper for evaluator bør udvikles til vejledende etiske principper for evaluering. En måde at gøre det på er at nytænke og kraftigt omforme de 5 principper.

Det kunne gøres sådan:

Systematisk undersøgelse:

I stedet for at tænke "systematisk undersøgelse", bliver det vigtigt at anvende **systemisk tænkning!** Troen på at en systematik skulle højne fagligheden hører til inden for realismen. Ud fra et socialkonstruktionistisk perspektiv giver det ingen mening at sige, at det er systematikken i sig selv, der skulle højne fagligheden, for hvad er det, der skal systematiseres? Det socialkonstruktionistiske perspektiv åbner op for muli-

tivers, hvilket betyder, at hver aktør oplever verden fra sit bestemte perspektiv. Derfor bliver det dels vanskeligt at foretage dækkende og systematiske beskrivelser, da billedet ofte vil pege i alle retninger. Og dels vil bestræbelsen efter den systematiske beskrivelse af et komplekst og broget multivers kræve mængder af ressourcer, som vi kun i sjældne tilfælde råder over.

I stedet er vi optagede af, hvordan handlinger og meninger fletter sig ind i hinanden i et system, der er flygtigt og dynamisk. Viden samskabes i relationen og kommunikationen og varierer afhængig af position og er knyttet til kontekster og aktørers handlemuligheder. Når vi tænker systemisk, giver vi plads til konkret handleanvisende viden fremfor generel, abstrakt og objektiv viden. At arbejde systemisk bygger også på en systematik, men ikke i betydningen: en dækkende stabil beskrivelse. Systematikken vil snarere gå på at samskabe forskellig viden bygget på erfaringer set fra forskellige perspektiver.

Kompetencer.

I stedet for et at tænke i besiddelse af individuelle kompetencer, bliver **mobilisering af kompetencer hos alle interessenter** vigtigt. Tænkningen om, at evaluators kompetencer og personlige egenskaber og evner til at begå sig i et komplekst felt er noget, man som evaluator enten besidder eller ikke besidder, udskiftes med en tænkning om, at kompetencer mobiliseres gennem samarbejde og dialog og her spiller øvrige interessenters kompetencer en lige så stor rolle som evaluators. I stedet for at se kompetencer som iboende et individ, synliggøres kompetencer i relationer. Dette skift i tænkning handler om, hvordan evalueringen kan bidrage til at mobilisere kompetencer hos alle interessenter og dermed udvikle og berige den praksis, der lader sig evaluere og den praksis, foretager evalueringen. Med kompetencer følger medansvar og medinddragelse.

Integritet/Ærlighed:

Det snævre fokus på integritet og ærlighed som et individuelt og personligt kendetegn giver ingen færdigheder i **mestring af anerkendende** dialog. Idéen om, at Individet ses i forhold til sig selv udskiftes med et fokus på individet i relation til andre. Troen på, at hvis

individet er i overensstemmelse med sig selv, så vil det automatisk udrette noget godt, udskiftes med en tanke om, at individer er hinandens forudsætninger for at fremstå kompetente og moralsk ansvarlige. Dette sker gennem kommunikation og handlinger. Ærlighed bliver hermed ikke et klart pejlemærke, idet det forudsætter, at der er en entydig sandhed, vi kan være ærlige overfor. Når vi erkender, at der altid er flere versioner af sandheden, bliver princippet om ærlighed reduceret til at være tro mod sig selv og sit eget perspektiv. I en mangetydig og modsætningsfyldt verden er den slags ærlighed og integritet et fattigt redskab. Vi har i stedet brug for at kunne mestre en anerkendende dialog, der bygger bro over konflikter og modsatrettede forventninger og inddrager forskellige perspektiver og oplevelser.

Respekt for mennesker.

I stedet for respekt for mennesker er **omsorg for relationer** mellem mennesker vigtig. Princippet om respekt for mennesker betyder i praksis respekt for det enkelte individ, men når individer er forskellige og ser forskelligt på ting, hvordan kan respekt for den enes perspektiv, så undgå at være despekt for den andens? Når der drages omsorg for relationer åbner det op for en hensyntagen til det organisatoriske fællesskab fremfor det enkelte individ, samt at alle perspektiver får en stemme og ikke mindst, at evalueringen udføres på en sådan måde, at relationerne støttes og ikke beskadiges i processen med at samskabe brugbar viden.

Ansvar for generel og offentlig velfærd.

Princippet om ansvar for generel og offentlig velfærd bliver til et princip om **nytteværdi for de relevante fællesskaber**. Evaluering udretter jo ikke noget i sig selv, men skal gøre nogle aktører i stand til at udrette noget bedre, end de gjorde før evalueringen. Resultater i en evaluering er død viden, med mindre den kan anvendes af nogen. I tanken om nytteværdi ligger også et hensyn til ressourceanvendelsen i forhold til evalueringens udbytte. Hvis ikke evalueringen skaber væsentlig og brugbar viden og giver værdi til organisationen, så er den ikke ressourceforbruget værd, og vi var bedre stillet med at bruge ressourcerne andetsteds.

Det ville give nogle helt andre handlemuligheder og – anvisninger, hvis vi i evaluering således arbejder ud fra følgende vejledende etiske principper: 1. Systemisk tænkning. 2. Mobilisering af kompetencer hos alle interessenter. 3. Mestring af anerkendende dialog. 4. Omsorg for relationer og 5. Nyttelværdi for de relevante fællesskaber.

RELATIONEL ETIK BASERER SIG PÅ DIALOG OG SOLIDARISK REFLEKSIVITET

Et skift i erkendelsesteori giver således radikalt nye input til etik-diskussionerne samt til drøftelserne om, hvorvidt vi i Dansk EvalueringsSelskab skal udarbejde vejledende etiske principper for evaluering og i givet fald, hvordan de så skal se ud. Spørgsmålet, som vi må stille i den anledning er, om formålet med de vejledende principper er at samle og holde sammen på en profession, sådan som tankerne var med AEA´s guiding principles?

Et andet spørgsmål, vi må stille os er, om det er hjælpsomt og nyttigt at tro, at et sæt vejledende etiske principper skal kunne rumme den diversitet og mangfoldighed af evalueringsmetoder og tilgange, som findes i feltet? Som Dansk EvalueringsSelskab og evalueringsfeltet ser ud i Danmark, har vi - som jeg ser det mere brug for vejledende etiske principper, der kan bruges i en dialog mellem evaluator, rekvirent og brugere af evaluering.

Udfordringen med AEA´s guiding principles er, at de på den ene side skal være så tilpas abstrakte, at de kan anvendes i forbindelse med enhver evaluering og samtidig anvisende og konkrete nok til, at de indholdsmæssigt også gør en forskel fremfor slet ingen at give.

En løsning på dette skisma kunne være at udvikle mindst to sæt vejledende etiske principper for evaluering; et på et socialkonstruktionistisk grundlag og et andet på realismens grundlag, og så lade rekvirenter, opdragsgivere og evaluører i fællesskab vælge mellem de specifikke tilgange med dertilhørende forskellige etiske principper alt efter evaluanden, den valgte metode, den organisatoriske kontekst samt anvendel-

sessigtet.

I mange organisationer og i samfundet i det hele taget er der en stigende viden om erkendelsesteorier og forskellige tilgange til metoder. Derfor vil rekvirenter af evalueringer i stigende grad have forventninger om at evaluører kan gøre rede for deres valg af erkendelsesteori, metode, proces og de vejledende etiske principper, der er relevante i forhold til disse valg. At arbejde med mindst 2 sæt vejledende etiske principper ville være en kærkommen udfordring for evalueringsfeltet. Ikke mindst vil valgmuligheden mellem de to sæt være en ansats til, at evaluator, rekvirent og øvrige involverede får drøftet mulige etiske udfordringer og foretaget et valg mellem de to eller måske udvikle en tredje model. Ikke mindst vil det socialkonstruktionistiske sæt vejledende etiske principper åbne mange nye perspektiver og sætte en relationel etik i spil, som i mine øjne, er det nødvendige næste skridt i retningen mod en femte generationsevaluering.

LITTERATURLISTE:

AEA (1994): www.AEA.com

Gergen, Kenneth J.(2005): Virkeligheder og relationer. Dansk Psykologisk Forlag.

Gergen, Kenneth J.(2010): En invitation til social Konstruktion. Forlaget Mindspace.

Haslebo, Gitte og Maja Loua Haslebo (2007): "Etik i organisationer. Fra gode hensigter til bedre handlemuligheder". Dansk Psykologisk forlag.

Larsen, Peter Dahler og Hanne Kathrine Krogstrup (red.)(2001): Tendenser i evaluering. Syddansk Universitetsforlag.

Lund, Gro Emmertsen (2010): "Evaluering i et socialkonstruktionistisk perspektiv". Relationel Praxis. Forlaget Mindspace.

Lund, Gro Emmertsen (2011): Femte Generation Evaluering. I: Dansk Evaluerings Selskabs Tidsskrift, se linket: <http://www.danskevalueringsselskab.dk/tidsskrift.htm>

Michael Morris (2011): "The Good, the bad and the evaluator: 25 years of AJE Ethics". American Journal of evaluation.

Simons, Helen (2006) Ethics in evaluation. In, Shaw, Ian F., Greene, Jennifer C. and Mark, Melvin M. (eds.) The SAGE Handbook of Evaluation. London, GB, Sage, 213-232.

Virtanen, Petri og Ilpo Laitinen (2004): Beyond Evaluation Standards? European Journal of spatial Development. <http://www.nordregio.se/EJSD/> -ISSN 1650-9544-Refereed Articles oct. no 13

NOTER

1: Effekter forstået her som konstitutive virkninger af evalueringen. (Peter Dahler-Larsen, 2001)

2: The five Guiding Principles for Evaluators (Systematic Inquiry, Competence, Integrity/Honesty, Respect for People, and Responsibilities for General and Public Welfare). AEA (1994)

3: Michael Morris (2011): "The Good, the Bad, and the Evaluator: 25 Years of AJE Ethics". American Journal of Evaluation.

4: Haslebo & Haslebo (2007)

5: Læs mere herom i Lund 2010: Evaluering i et socialkonstruktionistisk perspektiv. Relationel Praksis, Forlaget Mindspace.

ER PROGRAMTEORIBASEREDE EVALUERINGER I STAND TIL AT VURDERE EFFEKTER AF OFFENTLIGE INDSATSER¹?

Artiklen diskuterer en række kritikker, der har været fremført imod programteoribaserede evalueringers evne til at evaluere effekter. Herudfra analyseres 8 aktuelle eksempler på denne type af evalueringer. Konklusionen er, at programteoribaserede kan anvendes til effektevaluering, men også at der indenfor aktuell evalueringspraksis i Danmark er brug for at arbejde mere systematisk med en række punkter, f.eks. med at lede efter beviser, der kun er sandsynlige, givet programteorien er sand eller afvise trusler mod mekanismerne i programteorien.

INDLEDNING

Effekter af politiske initiativer har gennem en årrække fået en fremtrædende rolle i den danske velfærdsstat. Således blev der fra politisk side afsat knap 100 millioner kr. i 2004 til, at kaste lys over hvilke effekter forskellige velfærdsforanstaltninger havde (Egelund og Lausten 2010: 4). Der er således et øget krav om, at sociale initiativer skal kunne dokumentere en effekt, før de anses som legitime.

Når man er interesseret i at evaluere effekten af en handling, er man interesseret i forskellen mellem, hvad der sker efter en handling har fundet sted, og hvad der ville være sket, hvis handlingen ikke havde været igangsat (Shadish, Cook & Campbell 2002: 5). Gyldigheden af kausalsammenhængen er derfor i fokus. Traditionelt har der været en opfattelse af, at eksperimenter er den eneste valide måde hvorpå, effekten af en indsats kan klarlægges (Hede og Andersen 2007: 31-32; Rieper og Hansen 2007: 115-119). Årsagen er, at eksperimenter giver nogle unikke muligheder for at kontrollere en række klassiske udfordringer i forhold til at fastslå kausalitet (Shadish, Cook & Campbell 2002: 5-7, 53-54, 63, 247-53). Eksperi-

menter bruges dog kun sjældent i Danmark, da de er dyre at gennemføre, og da de af en række andre grunde anses for problematiske (Dybdahl 2010: 12). Mange indsatser på velfærdsområder implementeres således under forhold, hvor det hverken er praktisk eller etisk muligt at lave et randomiseret eksperiment.

UDFORDRING FOR PROGRAMTEORETISK EFFEKTEVALUERINGER

Måske fordi eksperimenter er sjældent anvendt, er evalueringsmodeller, der tager udgangspunkt i en eller anden form for programteori, såsom virkningsevaluering, logiske modeller, indsatsteori, realistic evaluation osv., meget populære. Alle disse evalueringsmodeller benytter sig af programteori til at klarlægge *hvordan* og *hvorvidt* en indsats fungerer.

Det er velkendt at programteoribaserede effektevalueringer består af to dele: En teoretisk model, der udfolder kausalkæden mellem indsatsen og outcome, samt en evalueringsdel der guides af den teoretiske model. (Se bl.a. Dahler-Larsen 2009).

Evalueringemetoden har dog mødt en del kritik, idet

den anklages for ikke at forholde sig til de klassiske kriterier for kausalitet (Cook 2000: 29-31). Kritikken går på, at programteoribaserede effektevalueringer:

- *Ikke kan klarlægge, om der er samvariation mellem indsatsen og outcome.
- *Ikke kan udelukke tredjevariable, da de ikke estimerer den kontrafaktiske situation.
- *Ikke har kontrol over tidsrækkefølgen.
- *Ikke afgør, om rivaliserende programteorier er mere plausible, end den der er opstillet.

Kriteriet om samvariation mellem interventionen og outcome kan i programteoribaserede effektevalueringer afgøres ved at se efter en ændring i outcome efter indsatsen er gennemført².

Mere alvorligt forholder det sig med de øvrige kritikpunkter. Håndteres de ikke, kan man i princippet ikke afgøre, om indsatsen virker. Men betyder det så, at vi ikke kan bruge programteoribaserede effektevalueringer til at udtale os om effekten af en indsats? Min påstand er, at problemet kan løses med inspiration fra Process Tracing og Contribution Analysis.

HÅNDBLIVNING AF KAUSALITETSPROBLEMET I PROGRAMTEORIBASEREDE EFFEKTEVALUERINGER

Ved Process Tracing udelukkes tredjevariable ved at finde stærke beviser for mekanismerne i kausalkæden. Tanken stammer fra bayesiansk logik, og fokuserer på, at ikke alle beviser for en mekanisme i programteorien, er lige stærke. Meget groft kan logikken forklares ved, at observationer, der kun er sandsynlige at finde, hvis en kausalmekanisme er sand, tillægges stor værdi, mens observationer, der også er sandsynlige at finde, hvis kausalmekanismen ikke er sand, tillægges lille værdi (Beach & Pedersen 2010: 217-236). Logikken kan eksemplificeres med en analogi til politiets opklaringsarbejde. Et stærkt bevis er således en overvågningsfilm, der viser den mistænkte myrde. Et mindre stærkt bevis er derimod, at den mistænkte ikke kan forklare, hvad han lavede på gerningstidspunktet (Ibid.).

Tredje variable afvises altså ikke direkte, men de gøres usandsynlige set i lyset af de stærke beviser for den opstillede kausalmekanisme. På samme måde medfører sandsynliggørelsen af kausalmekanismen, at konkurrerende kausalteorier bliver mindre sandsynlige.

Ved Contribution Analysis håndteres kausalitetsproblemet ligeledes ved at styrke troen på mekanismerne i kausalteorien (Mayne 1999: 13-17, Dybdal: 2010).

Dette gøres ved at³:

- *finde beviser for mekanismerne i kausalkæden
- *inkorporere trusler mod mekanismerne i programteorien, som testes og afvises.
- *undersøge rivaliserende programteorier.
- *undersøge hvilke kontekstuelle faktorer, der har betydning for kausalmekanismen.

Kritikken angående tidsrækkefølgen håndteres på to forskellige måde af Process Tracing og Contribution Analysis.

Ifølge Process Tracing følger mekanismerne logisk tidsmæssigt efter hinanden (Beach & Pedersen 2010: 222). Problemet skal således løses ved at argumentere for tidsrækkefølgen.

Ved Contribution Analysis anerkendes det, at der kan være problemer (Rogers: 2005:18). Derfor skal det overvejes, hvornår man måler på indikatorerne for mekanismerne.

Ingen af de præsenterede løsninger kan med sikkerhed garantere, at de observerede ændring i outcome stammer fra indsatsen. Dette er dog en af pointerne med tilgangene. Effekter er altid betinget af kontekstuelle forhold, der har betydning for, hvordan en indsats virker. Derfor handler kausalinferens om at sandsynliggøre sammenhæng og tage de nødvendige kontekstuelle forbehold. Derudover giver programteoribaserede effektevalueringer ofte viden, der kan være med til at forbedre en indsats.

HVORDAN HÅNDBTERER PROGRAMTEORIBASEREDE EFFEKTEVALUERINGER FRA DANMARK KAUSALITETS-PROBLEMET?

Jeg har gennemført en analyse af 8 simpelt tilfældige udvalgte evalueringer⁴, for at klarlægge i hvor høj grad de opfylder kriterierne ovenfor. De udvalgte evalueringer bygger alle på tilgange, der kan betegnes som programteoribaserede effektevalueringer. Derudover er kvaliteten af evalueringernes opstillede kausalkæder, indikatorer, samt de metodiske tilgang også vurderet. Analysen vil dog primært behandle evalueringernes håndtering af kausalitetsproblematikken og kun afslutningsvist berøre de andre emner.

Evalueringernes håndtering af kausalitetsproblemet vurderes ud fra følgende parametre: 1. Vises det, at der er samvariation mellem indsatsen og outcome? 2. Argumenteres der for tidsrækkefølgen? 3. Sandsynliggøres programteorien? 4. Inddrages den kontekst evalueringen foregår i?

De fleste evalueringer lever op til kriteriet om *samvariation*. Således klarlægger syv ud af otte evalueringer, hvorvidt der er samvariation mellem indsatsen og outcome. Dette foregår enten via en nulpunktsmålning og/eller via en sammenligningsgruppe.

Mere problematisk forholder det sig med evalueringernes argumentation for *tidsrækkefølgen*. Ingen af de gennemgåede evalueringer forholder sig på nogen måde til tidsrækkefølgen i programteorien. Fraværet af argumentation for tidsrækkefølgen kan dog skyldes, at denne opfattes som logisk. Det er derfor usikkert, om dette udgør et egentligt problem.

Sandsynliggørelse af programteorien er helt centralt for muligheden for at fortage kausalinferens. Som tidligere nævnt kan dette gøres via værktøjerne fra Process Tracing eller Contribution Analysis. Dvs. ved at finde beviser for mekanismerne, som kun er sandsynlige givet teorien er sand, og/eller afvise trusler mod mekanismerne i kausalteorien og undersøge revitaliserende kausalteorier. Seks ud af otte evalueringer, opfylder ikke dette kriterium. Kun to af evalueringerne opfylder kravet i nogen grad. Eksempelvis "Viden der forandrer - Virkningsevaluering af læsevejlederen som

fagligt fyrtårn". Evalueringen undersøger om tilstedeværelsen af en læsevejledning, støtter lærerne i deres undervisning af eleverne. Her findes der frem til, at læsevejledningen kun virker under nogle ganske særlige (kontekstuelle) omstændigheder (EVA 2009:52). Beviserne herfor er så specifikke, at de kun synes sandsynlige, givet mekanismen er sand. Da evalueringen ikke argumenterer for sandsynligheden af alle mekanismerne i kæden, kategoriseres den dog til kun at opfylde kriteriet i nogen grad.

Det sidste kriterium for kausalinferens ved programteoribaserede effektevalueringer, er klarlæggelse af kontekstuelle forhold med betydning for programteorien. Seks af de otte evalueringer forholder sig til kontekstuelle forhold. Evalueringerne opfylder således i stor udstrækning dette kriterium.

Næsten alle evalueringerne håndterer altså enkelte af kravene i forbindelse med kausalinferens. Denne håndtering er dog yderst overfladisk og mangelfuld. Kun enkelte evalueringer håndterer kausalitetsproblemet i en sådan grad, at der skabes troværdighed om de fundne sammenhæng. Problemet bliver forstærket af, at evalueringerne alt for sjældent opstiller konkrete og målbare indikatorer for mekanismerne i programteorien, og kun sparsomt diskuterer de metodiske valg, der er foretaget. På trods af at denne analyse bygger på en meget lille stikprøve, tegnes der et billede af, at programteoribaserede effektevalueringer generelt ikke er af en kvalitet, der tillader, at der foretages kvalificerede politiske beslutninger på baggrund af dem. Hvor efterlader det så programteoribaserede effektevalueringer?

KONKLUSION: FORSLAG TIL FORBEDRING AF PROGRAMTEORIBASEREDE EFFEKTEVALUERINGER

Skal programteoribaserede effektevalueringer i fremtiden kunne bruges til at vurdere effekten af forskellige politiske indsatser, er der behov for at ændre evalueringspraksis. Nedenfor opstilles en række forslag til, hvordan programteoribaserede effektevalueringer kan forbedres.

*Opstil en programteori; argumenter for denne teoretisk og konstruer målbare og velovervejede indikatorer

for mekanismerne i programteorien. Kun velunderbyggede programteorier kan forsvares mod konkurrerende programteori, og uden målbare indikatorer er det umuligt at sandsynliggøre mekanismerne i programteorien.

- * Styrk troen på programteorien og sandsynliggør eksistensen af en sammenhæng mellem indsatsen og outcome ved at: 1) finde beviser for mekanismerne, der kun er sandsynlige at finde, givet programteorien er sand. 2) og/eller afvise trusler mod mekanismerne.
- * Overvej om der eksisterer rivaliserende programteorier og forsøg at afvise disse. Dette vil styrke tilliden til den oprindelige kausalteori.
- * Overvej eksplicit om der er tidsrækkefølgeproblemer. Er dette er tilfældet, skal de forsøgs håndteret.
- * Overvej eksplicit om de anvendte metoder og kilder er hensigtsmæssige til at måle de enkelte mekanismer i programteorien og vær ærlig om svagheder

Slutteligt kan det indvendes, at et for ensidigt fokus på teknisk gode programteoribaseret effektevaluering kan medføre et tab af ejerskab hos de personer, der indgår i evalueringen. Denne indvending er rimelig, men hvis fokus er at evaluere effekten af en indsats, bør evalueringen foretages sådan, at det reelt er muligt udtale sig om effekten. Ønsker man derimod brugerinddragelse og ejerskab, er det også helt legitimt. Til gengæld må man acceptere, at dette kan påvirke muligheden for at vurdere effekterne af en indsats.

LITTERATUR

- Beach, Derek og Pedersen, Rasmus Brun (2010). Process tracing: Metode, design og forskningslogik. I Andersen, Lotte Bøgh, Hansen, Kasper Møller og Klemmensen, Robert: *Metoder i Statskundskab*. København: Hans Reitzels Forlag, ss. 215-237
- Cook, Thomas D. (2000). The False Choice Between Theory-Based Evaluation and Experimentation. *New Directions for Evaluation*, no 87, Fall 2000, pp 27-34
- Dahler-Larsen, Peter (2009). Opskrift på virkningsevaluering. I Dahler-Larsen, Peter og Krogstrup, Hanne Katrine: *Nye veje i evaluering*. Viborg: Akademia, pp. 51 – 79

Dybdal, Line (2010). Contribution analysis - Et alternativ til klassiskeffektdesign? I: *Evalueringsnyt 24 - Tema: Effektmåling - Metoder, Anvendelse og Effekter*. <http://www.danskevalueringsselskab.dk/pdf/Evalueringsnytnr24.pdf>

Egelund, Niels og Laustsen, Helen (2010). Effekttundersøgelser inden for det specialpædagogiske område - Metodiske problemstillinger og overvejelser. I: *Evalueringsnyt 24 - Tema: Effektmåling - Metoder, Anvendelse og Effekter*. <http://www.danskevalueringsselskab.dk/pdf/Evalueringsnytnr24.pdf>

EVA (2009). *En lærerig vej til resultater. Håndbog i evaluering ved hjælp af indsatssteori*. <http://www.eva.dk/projekter/2007/evaluering-af-udviklingsprojekter/projektprodukter/en-laererig-vei-til-resultater>

Hede, Anders (2007). Systematiske reviews og evidens. I Fulgsang, Lars et al. (red.): *Teknikker i samfundsvidenskaberne*. Frederiksberg: Roskilde Universitetsforlag, pp 28 – 53

Mayne, John (1999). Addressing Attribution through Contribution Analysis: Using Performance Measures Sensibly. Discussion <http://dsp-psd.pwgsc.gc.ca/Collection/FA3-31-1999E.pdf>

Rieper, Olaf og Hansen, Hanne Foss (2007). Metode-debatten om evidens. København: AKF Forlaget, s. 4-122

Rogers, Patricia (2005). Program Theory: Not Whether Programs Work but How They Work. I Eliot Stern (red): *Evaluation Research Methods*. Volume III. London: Sage, s. 3 – 30

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin. Pp. 1 - 61, 82 - 93, 96 - 102, 161 - 165, 246 - 290

NOTER

1: Jeg skylder stor tak til Niels Matti Søndergaard for hans kyndige støtte og vejledning i forbindelse med denne artikel.

2: I praksis er dette dog sværere end som så (Mayne 1999: 1-3)

3: En mere uddybende gennemgang af Contribution Analysis findes i Line Dybdals (2010) artikel "Contribution analysis - Et alternativ til klassiskeffekt-design?" Bragt i Evaluerings Nyt 24.

4: Følgende evalueringer indgår i analysen: 1) Bonde, Ane og Schmidt, Rikke Pedersen (2007). "Evaluering af sundhedsfremme på arbejdspladsen med personlige sundhedsprofiler". 2) CAPI (2005). "Ambassadørordningen - under luppen. Evaluering af Projekt "Ambassadørordningen"". 3) Danmarks Evalueringsinstitut (2008). "Viden der forandrer. Virkningsevaluering af læsevejlederen som fagligt fyrtårn". 4) Incitre (2006). "Evaluering af fire metodeudviklingsprojekter støttet af puljen "Udvikling af behandlingstilbud til utilpassede unge"". 5) Københavns kommune (2009). "Evaluering af sikker by - Den korte snor". 6) Persson, Lenie (2008). "Virkningsevalueringsrapport - Lighed i sundhed Sundere liv i Greve Nord". 7) Rambøll (2009). "Rapport. Evaluering af tværfaglige samværs-møder i højkonfliktsager". 8) Region Midt (2009). "Projekt "Samfund og sprog som rehabilitering" - slutevaluering". Center for kvalitets udvikling.