



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Evaluering af arbejdsmiljøindsatser - Muligheder og begrænsninger

Hasle, Peter; Nielsen, Klaus Tranetoft; Limborg, Hans Jørgen; Bramming, Pia; Seim, Rikke

Publication date:
2012

Document Version
Tidlig version også kaldet pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Hasle, P., Nielsen, K. T., Limborg, H. J., Bramming, P., & Seim, R. (2012). *Evaluering af arbejdsmiljøindsatser - Muligheder og begrænsninger*. Arbejdsmiljørådet.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Evaluering af arbejdsmiljøindsatser

Muligheder og begrænsninger

FORORD

At skabe dokumentation for effekt af arbejdsmiljøindsatser er sin sag. At blive enige om måden at gøre det på er en anden sag. I årenes løb har aktørerne hver især – og i mange tilfælde på hver deres måde – arbejdet for at evaluere deres indsatser på en måde, der kan vise, at indsatsen har haft effekt i arbejdsmiljøet.

Kravene til dokumentation og legitimering af arbejdsmiljøindsatserne er stadig stigende og viser sig bl.a. i 2020-arbejdsmiljøstrategien, hvor der stilles krav til, at effekten af de igangsatte indsatser bliver synlig.

Målene i 2020-arbejdsmiljøstrategien er samfundsmæssige mål og så ambitiøse, at det kræver et system, som trækker i samme retning og anvender hinandens kompetencer og erfaringer – med respekt for hinandens roller. Der er et stadig behov for at blive klogere på, hvad der virker, og hvad der ikke virker ude i virkeligheden.

Som et led i at skabe klarhed og synlighed på feltet "Evaluering af arbejdsmiljøindsatser" har rådet taget initiativ til denne publikation. En publikation, som kan bidrage til en fælles forståelse og en ramme for evaluering og måling af virkninger på arbejdsmiljøområdet.

Det næste skridt på vejen er en fælles enighed om, at arbejdsmiljøindsatser fremover planlægges, implementeres og evalueres. Arbejdsmiljørådet vil arbejde for, at denne enighed i aktørsystemet opnås, og rådet håber på, at publikationen og kommende drøftelser kan være med til at skabe større transparens i aktørsystemet om evaluering.

Publikationen og dens anbefalinger repræsenterer en tilgang til planlægning og evaluering, der kan bidrage til fremtidig læring af arbejdsmiljøindsatserne og legitimering af arbejdsmiljøarbejdet i partsystemet. Med læringen som omdrejningspunkt for den videre indsats kan det sikres, at de ambitiøse mål i 2020-arbejdsmiljøstrategien opnås.

God læse- og arbejdslyst

Lisbeth Lollike

Formand for Arbejdsmiljørådet

INDHOLDSFORTEGNELSE

Forord	0
Indholdsfortegnelse	1
1. Resumé	2
2. Indledning	5
3. Arbejdsmiljø og evaluering – før og nu	8
3.1 Evaluering af arbejdsmiljøindsatsen i historisk perspektiv	9
3.2 Det særlige ved at evaluere arbejdsmiljøindsatser	12
4. Udvikling af evalueringsteori	14
4.1 Den klassiske effektevaluering	14
4.2 Procesevaluering	15
4.3 Monitorering og evaluering	16
4.4 Evidensbaseret evaluering	17
Effektevaluering	17
Virkningsevaluering	18
4.5 Pragmatisk evaluering	19
4.6 Anvendeligheden af evalueringsparadigmerne	19
Styrker og svagheder ved evalueringsparadigmerne	20
5. Hvordan virker arbejdsmiljøindsatser?	22
5.1 Fra indsats til praksis	23
6. Arbejdsmiljøindsatser fra et afsenderperspektiv	26
6.1 Programteori	26
6.2 Virkemidler	30
6.3 Tre typer af virkemidler	31
6.4 Synergier mellem indsatser	32
6.5 Fra programteori og indsatser til projekter	35
7. Arbejdsmiljøindsatser fra et modtagerperspektiv	37
7.1 Mekanismer	37
7.2 Kontekst	38
7.3 Effekt	40
8. Fra evaluering til handling	44
8.1 Evidens	45
8.2 Validitet	46
8.3 Legitimitet og læring	48
9. Evalueringsscenerier	50
9.1 Før indsatsen	50
9.2 Evaluering af den begrænsede indsats	51
9.3 Evaluering af den omfattende indsats	53
9.4 Tværgående evaluering	54
9.5 Evaluering af et program	55
9.6 Afsluttende bemærkninger	57
10. Perspektivering	58
Litteratur	60

1. RESUMÉ

Evalueringer undersøger som udgangspunkt om noget virker - altså om det fører til de ønskede resultater - ofte betegnet som effekten. Evidens anvendes til at betegne viden om hvor sandsynlig effekten er. Det er et begreb som bruges og misbruges til mange forskellige formål. I sin grundbetydning betyder det blot at vi kan underbygge vores antagelser med oplysninger. Inden for forskningen har evidensbegrebet dog en noget mere ophøjet betydning - som viden der er videnskabeligt underbygget. I den forbindelse bruges evidensbegrebet særligt inden for det medicinske område. Herfra stammer dogmet om at evidens alene kan bygge på forsøg hvor man sammenligner en tilfældigt udvalgt gruppes reaktioner på en intervention med reaktionerne hos en kontrolgruppe der ikke har været udsat for interventionen. Evidens i denne forståelse anvendes ofte om vidensgrundlaget for sammenhængen mellem velafgrænsede påvirkninger i arbejdsmiljøet og konsekvenser for helbredet. Når der gennemføres arbejdsmiljøindsatser er sammenhængene imidlertid mindre velafgrænsede og mere komplekse. Kontekst har fx stor betydning; forebyggelse af ulykker fra en rundsav afhænger ikke kun af om afskærmningen er på plads, men også af arbejdspladsens sikkerhedskultur. I den medicinske tradition øger det evidensen hvis man kan afkoble netop de variable der ikke indgår i en direkte årsagssammenhæng. Men i en arbejdsmiljømæssig sammenhæng viser det sig ofte at være netop sådanne forhold, der er interessante og afgørende for om en indsats virker selvom de er besværlige at sætte på begreb og kvantificere.

Man bliver derfor nødt til at se på effekt i en bredere sammenhæng. I denne publikation betragter vi *virkning* som et begreb for de forandringer en arbejdsmiljøindsats kan have. Dermed reserveres effektbegrebet til de sammenhænge hvor en direkte kausalitet kan dokumenteres. Det er for arbejdsmiljøindsatser ofte vanskeligt. En virkning kan yderligere opdeles i *output*, *outcome* og *impact*. Output er den umiddelbare 'leverance' af indsatsens kerneaktiviteter. Den afgørende virkning er her om indsatsen har formået at bringe disse 'leverancer' frem til den rette målgruppe. Outcome er de virkninger indsatsen har overfor målgruppen på lidt længere sigt, mens impact er den mere langsigtede virkning fx i form af holdbare forandringer og ændringer i opfattelser, prioriteringer og kultur.

Evalueringsforskningen (som den fx diskuteres af Dahler-Larsen 2010;Krogstrup 2011) har peget på at kvaliteten af en evaluering i høj grad er afhængig af de antagelser (i form af en programteori) der lægges til grund for udførelsen af evalueringen. En programteori forstås i denne sammenhæng som det samlede sæt af antagelser om hvordan en indsats vil virke. I denne publikation er det vores 'programteori' at evnen til at designe funktionelle evalueringer på arbejdsmiljøområdet afhænger af at alle aktører samlet set bliver bedre til at udvikle, beskrive og arbejde med refleksive programteorier som led i de evalueringer, der skal gennemføres. En sådan reflektiv programteori kræver at arbejdsmiljøaktørerne forholder sig til mulighederne for at bestemte indsatser vil omforme en given praksis i virksomhederne i den tilsigtede retning. Vi introducerer i den forbindelse begrebet mekanismer der dækker over de kræfter der får omsat en indsats til konkrete ændringer af arbejdsmiljøet. Og vi argumenterer for at mekanismebegrebet gør det muligt at sammenligne forløbet af arbejdsmiljøindsatser på tværs af virksomheder og brancher.

Sammenhængen mellem programteorien bag en arbejdsmiljøindsats og mulighederne for at evaluere dens efterfølgende virkninger vil ofte afhænge af mulighederne for at rumme kompleksiteten af den samlede arbejdsmiljøindsats. Mange arbejdsmiljøproblemer kan betegnes som 'vilde' problemer som er præget af stor kompleksitet, og hvor årsagssammenhængene mellem indsats og virkning bliver vanskelige at klargøre. Det er en udfordring for de gængse forestillinger om troværdighed (reliabilitet) og gyldighed (validitet). Den interne validitet handler om den anvendte metode (teoretisk og praktisk)

faktisk er i stand til at påvise de sammenhænge som man vil undersøge. Mens den eksterne validitet handler om mulighederne for at generalisere resultatet til en bredere sammenhæng. I samfundsmæssige indsats er det stort set aldrig muligt at gentage den samme indsats, og det er derfor særligt vigtigt at kunne generalisere resultaterne for at kunne lære af erfaringerne ved tilrettelæggelsen af den næste indsats. Samtidig er der krav om en omhyggelig metodebeskrivelse der giver andre forskere, evaluatore og beslutningstagere mulighed for at vurdere om en undersøgelse metodemæssigt er gennemført på en troværdig måde.

Ved evalueringen af hvordan en arbejdsmiljøindsats virker, bliver det relevant at se på hvilke virkemidler indsatsen benytter sig af. Vi anvender en overordnet systematisering af virkemidler som man kan finde hos Evert Vedung (1998). Vedung skelner i sin typologi over 'policy instruments' mellem kategorierne regulering (pisk), incitament (gulerod) og viden (prædiken). Begreberne er et udtryk for den måde virkemidlerne forventes at påvirke modtagerne til at gennemføre de ønskede handlinger. Man kan derfor opfatte kategorierne som en overordnet inddeling af de tre grundlæggende mekanismer en samfundsmæssig indsats kan betjene sig af.

I denne publikation præsenteres flere forskellige tilgange til evaluering. Det drejer sig om procesevaluering, monitorering, effektevaluering og virkningsevaluering. Særligt virkningsevalueringerne lever op til mulighederne for at forstå de mekanismer som fører til forandringer, og dermed for at lære af evalueringerne når den næste indsats skal sættes i værk. Mange evalueringer ender imidlertid med at være såkaldte pragmatiske evalueringer der plukker lidt hist og her blandt de forskellige evalueringstilgange. Det kan der være gode begrundelser for. Vælges en 'pragmatisk' evaluering frem for en af de rendyrkede, traditionelle evalueringstilgange, er der dog en oplagt risiko for at man sætter sig mellem to stole. Den information som kommer ud af en sammensat evaluering, kan på den måde blive så usikker og metodisk tvivlsom at den reelt ikke bidrager til et forbedret beslutningsgrundlag.

Det konkrete evalueringdesign er typisk begrundet i de tilgængelige ressourcer og den politiske virkelighed evalueringen skal foretages i. Ofte er det antagelsen at jo mere evalueringerne lever op til de idealer man finder inden for den naturvidenskabelige forskning, des bedre. Det er imidlertid sådan at uvildige, objektive evalueringer er et ideal snarere end en realitet. Det skyldes at evalueringer normalt sættes i gang af de samme beslutningstagere som er ansvarlige for den indsats, som skal evalueres. Dette sammenfald kan give relativt gode muligheder for at anvende evalueringresultaterne og blive klogere på hvad man gjorde godt og skidt, fordi det er den primære bruger som bestiller evalueringen. Men der er også en bagside. Beslutningstagerne vil næppe have interesse i evalueringresultater som påviser at det var en forkert beslutning, og at pengene er spildt. Samtidig er evaluator typisk en konsulentvirksomhed som lever af at lave evalueringer, og den vil normalt ikke have interesse i at lægge sig ud med bestilleren af evalueringen, samtidig med at evaluator selvfølgelig er afhængig af at bevare sit faglige renommé. Evalueringprocessen bliver altså i større eller mindre omfang påvirket af de involverede aktørers forståelser og interesser. Denne problemstilling med interesse-bias kan man forsøge at imødegå ved at inddrage paneler af uafhængige forskere/eksperter i evalueringprocessen. Selvom alle personer og forskningsinstitutioner aldrig er helt interessefrie og selvom denne type tiltag både fordyrer og komplicerer evalueringen, vil det særligt ved mere omfattende evalueringer være relevant.

Udover at valget af evalueringdesign er begrundet med det praktisk mulige, er det også meget afhængigt af hvilken type arbejdsmiljøindsats der er tale om. Den sidste del af publikationen handler netop om hvordan man kan sammensætte (pragmatiske) evalueringindsatser som tilpasses forskellige typer indsats. Vi arbejder med fire forskellige evalueringsscenarier som hver for sig på forskellig måde skal tilpasses til de tilgængelige ressourcer.

Evaluering af den begrænsede indsats

Evalueringsdesignet bør her være simpelt og benytte meget få ressourcer der alligevel bidrager med relevant viden. Ofte vil selvevaluering være den mest relevante evalueringsform.

Evaluering af den omfattende indsats

Evalueringsdesignet må her bygge på klare programteorier som kan danne grundlag for stærkt prioriterede evalueringer. Der bør udformes en dynamisk, strategisk planlægningsmodel med følgende trin: Situationsanalyse, strategiske mål, kritiske succesfaktorer, målepunkter, handlingsplan, evaluering, læring og opfølgning.

Tværgående evaluering

I denne evalueringsform er der mulighed for at evaluere sammenlignelige indsatser på forskellige områder. Evalueringsdesignet må her forholde sig til at indsatserne har forskellige målgrupper, og måske dermed også har varierende programteorier, men omvendt er fordelene at erfaringerne kan puljes og bruges på tværs.

Evaluering af et program

Et program består typisk af flere projekter/indsatser. Evalueringen må derfor opbygges således at evaluator gennem evalueringen af de mange enkelte projekter bliver i stand til at kunne gennemføre tværgående analyser af projekterne. Derved bliver det muligt at nå frem til generaliserede konklusioner, om hvad der virker og ikke virker i projekterne.

Hovedformålet med at evaluere arbejdsmiljøindsatser er at producere læring der tillader arbejdsmiljøsystemet konstant at forbedre effektiviteten af indsatserne. Der er behov for at udvikle en reflektiv evalueringkultur. Udviklingen af en sådan kultur kan støde ind i snubletråde. En af disse snubletråde er muligheden af at evalueringer bliver til legitimeringsforanstaltninger frem for at give læringsmuligheder. En anden snubletråd er at det stadige og tiltagende krav om evalueringer kan komme til at påvirke hvad der kommer til at være indholdet i indsatserne. Dette skal ses i sammenhæng med det store fokus på evidens. Hvis vi kun iværksætter indsatser som vi på forhånd har hug og stikfast viden om virker, bliver det måske så som så med udviklingen og læringen. Derfor er stadige og løbende diskussioner af både det faglige og praktiske indhold i evalueringer centrale. De er nødvendige for at sikre at vi til stadighed lærer og derigennem får at påvirke arbejdsmiljøet i en positiv retning.

Som vi har argumenteret for, er det afgørende at man vælger den evalueringsform der giver bedst mening i forhold til den indsats man ønsker at evaluere. Virkningsevaluering er den evalueringsform (tilpasset de konkrete omstændigheder) der i mange tilfælde er at foretrække fordi den bedst tilgodeser de læringsformål der i særlig grad knytter sig til evaluering af arbejdsmiljøindsatser.

2. INDLEDNING

Denne publikation bygger på evalueringsteori. Formålet er at give et væsentligt bidrag til at forstå hvad evaluering er, og hvordan og hvornår den bedst kan anvendes til evaluering af forskellige typer projekter inden for arbejdsmiljøområdet. Arbejdsmiljøområdet er præget af en lang og stærk forskningstradition der har været med til at udvikle en bred forståelse af hvad arbejdsmiljø er, og hvordan det hænger sammen med helbred og trivsel. Denne forståelse er bygget op gennem en lang række forskningsbidrag der tilsammen har skabt viden om arbejdsmiljø og effekten af arbejdsmiljøinterventioner. Der er nu behov for at anvende den eksisterende viden om arbejdsmiljø til at få en bredere forståelse af hvordan samfundsmæssige arbejdsmiljøindsatser fungerer. Det kan bl.a. ske i forbindelse med den evalueringspraksis der er under opbygning i disse år. I denne udvikling er det vigtigt at tage hensyn til de særlige problemstillinger og dilemmaer som knytter sig til evaluering og som bl. a. er med til at adskille forskning og evaluering.

Evaluering kan overordnet defineres som en systematisk bedømmelse af udfald, præstationer og organisering af samfundsmæssige aktiviteter. Det er en bedømmelse som er rettet mod fremtidige handlingssituationer, og dermed har evalueringer et anvendelsessigte (Dahler-Larsen 2010; Vedung 1997).

Det konkrete anvendelsessigte, der er forbundet med evalueringer, adskiller evaluering fra forskning fordi forskning ikke nødvendigvis har et direkte anvendelsesformål. På trods af denne skillelinje mellem evaluering og forskning er der betydelige fællestræk: Fx benytter evaluering metoder som også anvendes specielt indenfor samfundsvidenskaben og i stigende grad også af kvalitetskrav som kan genfindes i forskningsverdenen. Evaluering er, som den samfundsvidenskabelige forskning, også præget af store diskussioner og uenigheder både om de overordnede tilgange (paradigmer) for evaluering, og hvilke konkrete metoder der er bedst egnede. Derudover har felter inden for forskning også kastet sig over evaluering som et særskilt forskningsobjekt.

Evaluering er en målrettet aktivitet der interesserer sig for viden om konkrete indsatser som grundlag for at træffe beslutninger om nye indsatser. Selvom forskningen leverer væsentlig viden som indsatser kan bygge på, kan den sjældent på kort sigt anvendes direkte som grundlag for beslutningstagning. Forskningen vil typisk bidrage med bredere svar med flere nuanceringer, og ofte lang tid efter beslutningstagerne har været nødt til at tage stilling til den næste indsats. Evaluering bliver derfor vigtig for at sikre det bedst mulige grundlag for aktuelle beslutninger.

Der er imidlertid en række særlige problemstillinger knyttet til evaluering – nogle generelle og andre særligt knyttet til arbejdsmiljøområdet. Et af hovedproblemerne handler om at finde frem til sammenhænge mellem indsats og effekt. Vi har måske nogen viden om at der er en vis sammenhæng mellem de forskellige led i kæden. Vi mangler blot sikre metoder til at skille direkte effekter fra alle mulige andre faktorer som også har betydning, og dermed fungerer som en slags baggrundsstøj. Så selvom vi ikke har absolut viden om virkningerne hen gennem kæden, betyder det ikke at der ikke er nogen virkning. Et væsentligt element der ofte vil 'mudre' billedet, er de generelle udviklingsstrømme i samfundet. For eksempel har den økonomiske krise og afmatning i byggesektoren betydet stigende arbejdsløshed i denne branche. Effekten af samtidige arbejdsmiljøindsatser der målrettes mod nedbringelse af ulykkesfrekvensen i branchen, vil derfor være svær at måle; for hvad skyldes faldet i beskæftigelsen og mindsket travlhed, og hvad kan tilskrives den specifikke arbejdsmiljøindsats?

Den stigende kompleksitet der karakteriserer de problemer som arbejdsmiljøreguleringen skal forholde sig til, kan lidt forenklet eksemplificeres som forskellen mellem at regulere høreskadende støj og stress. Støjen er umiddelbart til at "få øje på"; den kan relativt let måles, hvad der forårsager støjen er

ofte enkelt at fastslå, at pege på løsninger er ligetil (men ikke altid), og virkningen på sundheden har vi et simpelt mål for. Hvis vi skal evaluere indsatsen mod støj i arbejdsmiljøet, er der altså en række værktøjer, som vi kan bruge. Stress-epidemien – som nogen kalder det – er tilsyneladende omfattende, men dog ikke i en sådan grad at der ikke hersker stor uenighed om dens realitet og omfang. Der er ingen sikre biologiske markører for stress. Hvad der forårsager stress, er ofte helt uklart og åbent for diskussion. Der er kun i begrænset omfang udviklet løsninger som bare med rimelig sikkerhed kan reducere medarbejdernes stress – for slet ikke at tale om at sikre at arbejdsforholdene ikke er stressende generelt set – og den langsigtede virkning af stress på befolkningens sundhed er i bedste fald uklar. Omvendt er det stadig nødvendigt at anvende den tilgængelige viden med al sin usikkerhed til at forsøge at forebygge stress, men denne usikkerhed gør det vanskeligt at svare på, hvordan man evaluerer arbejdsmiljøindsatser overfor stress. Dette er selvfølgelig på ingen måde enestående for arbejdsmiljøarbejdet. Samfundet er præget af en stigende kompleksitet hvilket er en udfordring som generelt præger forskning, evaluering og politik.

En given arbejdsmiljøindsats fungerer normalt i sammenhæng med mange andre indsatser, og det er den fælles påvirkning fra disse indsatser som forsøger på at ændre adfærden hos virksomhedens ledelse og medarbejdere. Der er tilsynsførende fra Arbejdstilsynet, der er informationskampagner fra BAR og Videncenter for Arbejdsmiljø, der er arbejdsmiljørådgivere, der er arbejdsmiljørepræsentanter med viden fra arbejdsmiljøkurser, og der er ledere der har arbejdsmiljøperformance inkluderet i deres bonusvilkår. Det bliver derfor vanskeligt at adskille et enkeltstående BAR-projekts virkning fra synergien med de mange andre indsatser som i princippet søger at opnå en tilsvarende adfærd i virksomhederne.

Et andet klassisk problem når det gælder koblingen mellem arbejdsmiljø og sundhed, er latenstiden. Fra udsættelsen for kræftfremkaldende stoffer eller støj til diagnoser som kræft eller varigt høretab går der ofte flere årtier. Andre virkningsmekanismer har væsentligt kortere tid fra påvirkning til effekt, men selv i koblingen mellem arbejdsmiljøarbejde og arbejdsmiljø kan der sagtens være en vis forsinkelse; det er jo netop i det lange seje træk i ledelsen og arbejdsmiljøorganisationen at grundlaget lægges for en organisation der kan og vil gøre noget ved arbejdsmiljøproblemerne.

Det måske vanskeligste problem er arbejdsmiljøets øgede kompleksitet. Arbejdsmiljøreguleringen har i stigende omfang fået kompleksitet som et grundvilkår. Der er langt fra børnearbejde til arbejdsrelateret stress. I Arbejdsmiljølovens levetid har arbejdsmiljøproblemstillingerne udviklet sig ganske betydeligt. I den første periode var der en væsentlig udvikling omkring stoffer og materialer, hvor opmærksomheden på kræft, reproduktionsskader, hjerneskader og hudirritation steg betydeligt. Siden er indeklima, muskelskeletbesvær og psykisk arbejdsmiljø blevet tildelt en stadig større opmærksomhed. Samtidigt har det samfundsmæssige mål med arbejdsmiljøindsatsen også udviklet sig fra at have fokus på at begrænse nedslidning og sygdom til at reducere sygefraværet.

Endelig udgør omkostninger en barriere for at gennemføre så grundige evalueringer at man kan opnå en sikker viden om reguleringens virkning - at etablere viden er ikke omkostningsfrit. Det vil typisk være sådan at jo mere præcis viden man ønsker sig, jo længere tid og jo større omkostninger vil der være forbundet med at opnå denne viden. Hvis den ønskede viden yderligere skal have en høj grad af validitet, vil det typisk kræve en indsnævring af undersøgelsesfeltet, idet 'støjen' skal holdes ude. Derved bliver det også sværere at anvende denne mere sikre viden i andre sammenhænge. Der skal ganske naturligt være et rimeligt forhold mellem budgettet for selve indsatsen og omkostninger til at vurdere den. Og i forlængelse af denne pointe er evalueringer og det evaluerede ikke hermetisk lukkede i forhold til hinanden. Målinger påvirker det målte, ligesom det ikke er tilfældigt hvad man udvælger at evaluere.

I denne publikation diskuteres en række teorier og begreber. Formålet er at få indsigt i de særlige problematikker evaluering af arbejdsmiljøindsatser kan afstedkomme. Det er vores vurdering at der er behov for en præcisering af de begreber der ofte anvendes i flæng, når vi snakker om arbejdsmiljøindsatser, kampagner, interventioner, virkemidler og lignende. Det er en nødvendighed for at forstå hvordan en given form for en indsats kan virke, og efterfølgende for at kunne evaluere om den faktisk har virket efter hensigten. Når det lykkes, kan evalueringer netop give værdifuld læring, som kan anvendes til at forbedre efterfølgende arbejdsmiljøindsatser.

I det næste kapitel vil vi beskæftige os med hvordan samfundets arbejdsmiljøindsats har udviklet sig, og hvordan man allerede på et tidligt tidspunkt har søgt at evaluere den. I kapitel 4 præsenterer vi udviklingen inden for evalueringsteori. Dernæst diskuterer vi i kapitel 5 hvordan arbejdsmiljøindsatser virker. I kapitel 6 og 7 arbejder vi med begrebet programteori som er et nødvendigt grundlag for at designe og evaluere effektive indsatser. I kapitel 8 kobler vi de forskellige evalueringsparadigmer til overvejelser om hvordan man kan omsætte disse mere teoretiske overvejelser om evaluering til konkrete handlinger. Kapitel 9 kommer ind på mulighederne for at tilrettelægge konkrete evalueringer af arbejdsmiljøindsatser bl.a. i forhold til fire forskellige typer indsatser. I dette kapitel diskuterer rapporten også hvordan man kan evaluere hele programindsatser hvor mange forskellige indsatser (der alle skal evalueres hver for sig) virker samlet. Endelig indeholder kapitel 10 en diskussion af perspektiver for den videre udvikling af evaluering af arbejdsmiljøindsatser.

3. ARBEJDSMILJØ OG EVALUERING – FØR OG NU

I handlingsplanen "Nye veje til et bedre arbejdsmiljø – Regeringens strategi for arbejdsmiljøindsatsen frem til 2020" (Beskæftigelsesministeriet 2011) kommer tankegangen om evaluering som byggesten for fremtidige beslutninger tydeligt til udtryk. . Handlingsplanen har en række direkte referencer til overvågning og evaluering som grundlæggende politikredskaber på arbejdsmiljøområdet.¹

Det fremgår af handlingsplanen at den er vidensbaseret for så vidt som den *"... hviler på et stærkt fagligt fundament. De seneste år er der indhentet både viden og erfaringer med arbejdsmiljøindsatsen. Arbejdstilsynet, Det Nationale Forskningscenter for Arbejdsmiljø og Arbejdsskadesstyrelsen har lavet grundig analyse og vurdering af de centrale udfordringer ..."* (s. 1).

Handlingsplanen henviser da også under overskriften *"Et solidt fagligt grundlag"* helt direkte til rapporten *"Fremtidens arbejdsmiljø 2020 – Fagligt grundlag for prioritering af arbejdsmiljøindsatsen"* (Arbejdstilsynet 2010b), som er Arbejdstilsynets rapportering af det indhentede vidensgrundlag.

Handlingsplanen er prioriteret og indeholder, som det hedder, *"Tre ambitiøse og konkrete mål"* for:

1. Antallet af arbejdsulykker
2. Antallet af beskæftigede, som er psykisk overbelastede
3. Antallet af beskæftigede, der har muskelskeletoverbelastninger.

Handlingsplanen er formuleret som 19 initiativer hvoraf flere peger på evalueringer. Således har gruppen med initiativ 18 og 19 overskriften *"Resultaterne skal dokumenteres"*. Initiativ 18 er *"Måling af fremdriften i arbejdsmiljøet"*, og her konstateres det i teksten at det i dag ikke er muligt at følge med i, om man kommer nærmere målsætningerne undervejs. Derfor skal der etableres målinger, der kan muliggøre dette. Initiativ 19 omhandler *"Effektmålinger af de konkrete indsatser"*. Her hedder det: *"Ved at evaluere effekten af de forskellige aktiviteter får man et reelt grundlag for at sikre, at aktiviteterne bidrager bedst muligt til at opfylde målsætningerne."* Og videre: *"Regeringen vil fra 2011 stille krav om, at Arbejdstilsynet og andre arbejdsmiljøaktører, der modtager midler fra det offentlige, skal dokumentere effekten af væsentlige aktiviteter."* (Beskæftigelsesministeriet 2011, s. 14)

Men det er ikke kun de sidste to initiativer som trækker på evalueringstankegangen. Initiativ 14, *"Bedre vejledning og information om arbejdsmiljø"* handler om gennem en ekstern undersøgelse at *"... afklare, hvordan virksomhederne får større gavn af den virksomhedsrettede vejledning og information, ..."*. Initiativ 10 *"Undersøgelse af erfaringerne med virksomhedernes arbejdsmiljøorganisation"* stipulerer en undersøgelse af erfaringerne med omlægningen på området inden 2015. Endeligt er der initiativ *"Risikobaseret tilsyn – Fokus på virksomheder med arbejdsmiljøproblemer"*, hvor kernen er at tilsynet med virksomhederne skal afpasses sandsynligheden for at virksomhederne har problemer med arbejdsmiljøet.

Risikobaseret tilsyn er nærmest en modebølge indenfor en række reguleringsfelter (Black & Baldwin 2010) og bygger i høj grad på en evalueringstankegang. Risikobaseret regulering er en fællesbetegnelse som Black & Baldwin beskriver på følgende måde: *"Den vigtigste komponent ... er evaluering af risi-*

¹ Handlingsplanen udmønter en aftale mellem den daværende regering (dvs. Venstre og Konservative), Socialdemokraterne, Dansk Folkeparti og det Radikale Venstre fra marts 2011.

koen for overtrædelse og en vurdering af den betydning som overtrædelse vil have for reguleringsmyndighedens muligheder for at opnå sin målsætning. I den ideelle form tilbyder risikobaseret regulering en evidensbaseret metode til at målrette brugen af ressourcer og til at prioritere opmærksomhed på den største risiko i overensstemmelse med en gennemskuelig, systematisk og begrundet overordnet ramme" (s. 181, vor oversættelse). Kernen i risikobaseret regulering adskiller sig således ikke fra centrale dele af evalueringstankegangen, og begrebet risikobaseret tilsyn læner sig på den måde også op ad en evalueringstænkning.

Der er således stærke træk i hele den handlingsplan for arbejdsmiljøindsatsen der gælder for de kommende år, som knytter an til aktuelle forestillinger om evaluering og evidensbaseret politik. Klarest udtalt i initiativ 18, men også i en række af de andre initiativer. Der refereres til evaluering relativt overordnet og med lidt forskellige ord som både handler om at dokumentere effekter, og om at finde ud af hvordan virksomhederne får størst gavn af indsatser.

Det er netop formålet med denne publikation at afklare hvordan evaluering kan anvendes i arbejdsmiljøindsatsen på den bedste måde. For at komme dette formål nærmere vil vi først dykke ned i, hvordan evaluering af arbejdsmiljøindsatsen i lidt bredere forstand har udviklet sig historisk set.

3.1 EVALUERING AF ARBEJDSMILJØINDSATSEN I HISTORISK PERSPEKTIV

Forventninger om at regulering og politiklægning på arbejdsmiljøområdet skal ske på et gennemsigtigt, systematisk og forsvarligt grundlag, er jo ikke ny, omend forventningerne uden tvivl er taget til i styrke som vi skal se senere. Denne udvikling skal naturligvis ses i lyset af at både arbejdsmiljøet og vores forståelse heraf også udvikler sig. I det følgende vil vi foretage nogle centrale nedslag i udviklingen af evalueringsbølgen på området.

I en vis forstand ligger det i enhver myndigheds institutionelle natur at opsamle erfaringer og udviklingstræk på kvantitativ vis med henblik at korrigere og begrunde sin egen indsats. Af Gyldendals Store Danske Encyklopædi fremgår det at: "*Ordet statistik kommer via ty. fra nylat. statisticus 'hvad der vedrører statsager'*" og "*Statistik, videnskab, som i sin oprindelige betydning omfatter en metodisk, empirisk funderet beskrivelse af samfundsforhold vha. indsamling og systematisering af talmaterialer.*" Danmarks Statistik blev etableret i 1850, og stat og statistik er sammenbundet fra første færd i etableringen af den grundlovsbaserede forvaltning i Danmark. I Arbejdstilsynets første årsberetning "Beretning om Arbejdstilsynets Virksomhed indtil Udgangen af Aaret 1874" kan vi finde nedenstående tabel over børnearbejdere på fabrikker.

Efter de fremkomne Oplysninger var Bestanden af Fabrikerne den 1ste Decbr. 1873 følgende:

Bedrifternes Art.	Antal Bedrifter inddragne.	B o r n				Unge Mennesker 14—18 Aar.	
		under 10 Aar.		fra 10—14 Aar.		Mandsjon.	Kvindesjon.
		Mandsjon.	Kvindesjon.	Mandsjon.	Kvindesjon.		
Bernstøberier og Ma- skinfabriker	67	"	"	6	"	389	"
Mekaniske Fabriker etc	42	"	"	3	"	113	"
Andre Metalvarefabriker	13	6	1	14	"	23	2
Papirfabriker	5	"	"	2	3	5	7
Forskjellige Papir- og Pappvarefabriker	9	5	7	23	16	12	25
Tapet-, Kulflegardin- og Børdugfabriker	10	12	1	60	2	19	1
Pibeslange-, Børstefa- briker o. desl.	3	"	1	5	39	"	13
Trævare- og Skorfæ- briker	6	3	"	3	"	9	"
Bogtrykkerier og Sten- trykkerier	49	"	"	49	5	156	26
Udvarefabriker	64	8	3	123	78	113	54
Bomulds- og Linned- varefabriker	19	12	6	98	56	76	93
Forskjellige femiste Fa- briker	27	1	3	59	17	41	21
Lædnstilsfabriker	7	12	"	93	48	27	28
Cement-, Sten- og Stukkaturfabriker	3	"	"	7	2	3	2
Porellæn-, Fajance- og Lervarefabriker	18	28	"	74	10	88	6
Glasværker	5	10	"	66	3	44	"
Tobak- og Cigarfabriker	116	282	26	906	173	387	88
Dampmøller	8	"	"	"	"	27	"
Chokolade- og Sukker- varefabriker	17	"	"	12	3	16	23
Sifoniefabriker	6	5	2	48	19	21	1
	494	384	50	1651	474	1569	380
			434		2125		1949
				2559			

Figur 1: Uddrag af kopi fra Arbejdstilsynets først årsberetning fundet på nettet²

Vi kan henføre principperne i vores gældende arbejdsmiljølovgivning til Arbejds miljøloven af 1975 – som siden kun er blevet ændret gennem mindre omfattende lovændringer. Da arbejdsmiljøloven blev udarbejdet, var det bl.a. på grundlag af fire rapporter fra Arbejds miljøgruppen af 1972 (1973; 1974; 1975; 1976) som langt hen af vejen kan opfattes som en systematisk evaluering af arbejdsmiljøtilstanden i begyndelsen af 1970'erne.

² (<http://www.emu.dk/gym/fag/hi/inspiration/tema/bornarb/28tabel1874.pdf>)

Med handlingsprogrammet "Rent arbejdsmiljø år 2005" fra 1996 begyndte Arbejdstilsynet at formulere målbare prioriteringer for sig selv og det samlede arbejdsmiljøsystem. I 1996 var de syv visioner dog ikke mere målsatte end at de udpegede problemer bare skulle reduceres mest muligt/helt undgå, som det hed.

Det var ikke kun arbejdsmiljøsystemet selv der stod som initiativtager til det forøgede fokus på evaluering. Rigsrevisionen spillede en væsentlig rolle i form af en undersøgelse af Arbejdstilsynets styring af tilsynsfunktionen i 2000 (Rigsrevisionen 2000). Nogle af de centrale og illustrative punkter i sammenfatningen var at:

"uanset at Handlingsprogram 2005 har skullet virke siden 1996, udviser antallet af arbejdsulykker og -lidelser, opdelt på de 7 visionsområder og opgjort på baggrund af de til Arbejdstilsynet anmeldte arbejdsskader, ikke en klar stigende eller faldende tendens. (...)"

"Arbejdstilsynets resultatkontrakt ikke indeholder mål for og resultatkrav til kvalitet og produktivitet samt effekt af Arbejdstilsynets opgavevaretagelse,"

"Arbejdstilsynet i undersøgelsesperioden ikke har foretaget produktivetsmålinger ud over opgørelse af enhedsomkostningerne for udvalgte aktiviteter,"

"Arbejdstilsynet ikke har et samlet ledelsesinformationssystem, der systematisk opgør resultater, kvaliteten heraf, produktivitet og effekt".

Der var således et betydeligt pres på en forøget evalueringsvirksomhed, og det var måske en af årsagerne til at der i 2002 blev strammet op på 2005-handlingsplanen med konkrete reduktionsmål for nogle af visionerne (alvorlige ulykker – 15 %, tunge løft – 15 %, ensidigt, gentaget arbejde – 10 %, psykisk arbejdsmiljø – 5 %). En vigtig detalje i "Rent arbejdsmiljø år 2005" var også ambitionen om "at udbygge statistik og dokumentation både i Arbejdstilsynet og Arbejdsskadestyrelsen så alle arbejdsmiljøaktører – herunder virksomheder – kan bruge det statistiske materiale." Der har siden været en væsentligt forøget monitorering af arbejdsmiljøet, fx har man udarbejdet de såkaldte overvågningsrapporter siden 1999 og gennemført de såkaldte VOV-undersøgelser (VOV for VirksomhedsOverVågning) siden 2001.

Handlingsprogrammet "Rent arbejdsmiljø 2005" blev evalueret i rapporten "10 år med handlingsprogrammet Rent Arbejdsmiljø år 2005" (Arbejdstilsynet 2007a). Resultatet var lidt nedslående for så vidt som målet ikke blev nået for tre af målsætningerne, mens der for højrepetitivt arbejde blev fundet et fald på 18 %, og dette mål var således blevet opfyldt.

I 2005 kom "Fremtidens arbejdsmiljø 2010" (Arbejdstilsynet 2005) som også havde konkrete reduktionsmål på sine fire prioriterede indsatsområder: Ulykker – 20 %, psykisk arbejdsmiljø – 10 %, muskelskeletbesvær – 10 %, støj – 15 % for høreskadende støj, 10 % for øvrig støj. En særlig detalje er at reduktionsmålene denne gang var formuleret som effekt på sygefraværet – ikke som reduktion i belastningen. Man kan sige, at evalueringsmetoderne udvikles og evalueringsindsatsen intensiveres, samtidigt med at målsætningerne for selve indsatsen fokuseres, så de er mere præcist formulerede i forhold til det man egentligt ønsker at opnå. I denne periode var der netop et stærkt fokus på arbejdsudbuddet.

I forbindelse med arbejdsmiljøreformen 2004 iværksatte Arbejdstilsynet en større evaluering af reformen. Evalueringen er afrapporteret i en række arbejdsrapporter, fire delrapporter og en slutrapport (Arbejdstilsynet 2010a). En lang række aktører eller institutioner er også blevet evalueret i de senere år fx:

- "Evaluering af Videncenter for Arbejdsmiljø – Centerets virksomhed 2005 til primo 2008" (Advice A/S 2008)
- "Stikprøvevis effektevaluering af branchearbejdsmiljørådernes projekter" (Capacent & Arbejdstilsynet 2007)
- "Evaluering af Arbejdsmiljøforskningsfonden" (Arbejdstilsynet 2011)

Uanset den mere gennemgribende evalueringsindsats er det dog vanskeligt at påvise resultater på det overordnede plan. Den umiddelbare evaluering af målopfyldelsen i forhold til 2010-målene (Arbejdstilsynet 2012) der netop er udkommet, viser at målene kun er nået på ét område (høreskadende støj), mens de andre områder udviser modsatrettede tendenser.

Mere generelt peger ovenstående gennemgang på at evalueringer på arbejdsmiljøområdet ikke er nogen ny foreteelse, men har spillet en væsentlig rolle for såvel den faglige som den politiske udvikling på området i hvert fald siden årtusindeskiftet, men med lange rødder tilbage i historien.

Man kan finde parallelle træk i udviklingen i mange andre landes arbejdsmiljøstrategier bl.a. i Sverige, Norge og England. I England har den britiske pendant til det danske Arbejdstilsyn - Health & Safety Executive (HSE) – således indført årlige såkaldte *business plans* (forretningsplaner) hvori overordnede kvantitative mål omsættes til konkrete aktiviteter ved hjælp af et målhierarki. På den måde anvendes målinger og evidens til at omsætte overordnede mål til konkret operationaliserbart niveau der kan evalueres. I forretningsplanen for 2010/11 ses følgende eksempel (for perioden 2000-10) på målhierarkiet: "*Forbedre arbejdsmiljøet*" skal opnås ved at "*Skabe sunde og mere sikre arbejdspladser*", og det skal ske gennem "*At skabe effektive interventioner*" baseret på "*forbedret datafangst*", at "*forbedre og bruge evidensbase*" samt at "*udvikle passende præstationsmål*". (Health and Safety Executive 2012). De nævnte operationaliserede mål i hierarkiet er blot nogle få af os udvalgte eksempler blandt en lang række som illustrerer hvordan evalueringslogikken også flourerer i Storbritannien.

Den udvikling vi har set i Danmark, når det gælder den strategiske tilgang til reguleringen af arbejdsmiljøet med øgede krav om evaluering og evidens, kan således ikke ses som et isoleret dansk fænomen. Det er snarere en international tendens som i virkeligheden ikke kun er fremherskende på arbejdsmiljøområdet, men også indenfor en lang række andre samfundsmæssige reguleringsfelter.

3.2 DET SÆRLIGE VED AT EVALUERE ARBEJDSMILJØINDSATSER

Som det foregående afsnit viser, har det historisk set optaget forskere og politikere at kunne dokumentere betydningen af arbejdsmiljøindsatserne og at kunne prioritere forskellige reguleringsindsatser på et solidt fagligt grundlag. Vi kan også observere en øget tendens til at efterspørge stadig mere dokumentation. Men til trods for en lang forskningstradition på feltet og en væsentlig forøget mængde information gennem de senere år, er der jo ikke meget der tyder på at vi over en bred kam er ved at få styr på hvordan vi kan sikre et sikkert og sundt arbejdsmiljø.

En forklaring på dette forhold skal søges i at der er en lang og kompleks virkningskæde mellem den indsats, man gør i arbejdsmiljøsystemet, og det man ønsker at opnå i den sidste ende. Denne kæde er i en simpel udgave skildret i nedenstående figur.



Figur 2. Virkningskæden på arbejdsmiljøområdet

Der er imidlertid ikke klare årsags-virknings-sammenhænge mellem de enkelte led i kæden. I praksis er det derfor svært – eller måske umuligt – at opnå en "alt andet lige"-situation hvor en relevant kvantificering af effekten længere nede af virkningskæden kan lade sig gøre. Der er flere forhold som udfordrer tankegangen om direkte årsag-virkningssammenhænge.

I den oprindelige arbejdsmiljøregulering før den nuværende arbejdsmiljølov bestod kæden kun af tre led: Regulering – arbejdsmiljø – helbred. Her stillede myndighederne klare krav, fx om afskærmning på rundsave, som efterfølgende blev kontrolleret. I et sådant tilfælde kunne der relativt enkelt konstateres en sammenhæng mellem antal afskårne fingre og tilstedeværelsen af afskærmning i træindustrien. Sådanne enkle sammenhænge genfindes imidlertid kun i en mindre del af arbejdsmiljøet hvilket netop var en væsentlig årsag til, at en arbejdsmiljølov med et universelt arbejdsgiveransvar og reflek-siv evaluering blev indført (Hasle et al. 2000). Det blev således nødvendigt at indføre arbejdsmiljøarbejdet som en mellemvariabel. Opmærksomheden er i stigende grad blevet rettet mod netop arbejdsmiljøarbejdet med diskussioner af sikkerhedsorganisationen i 1980'erne (Møller, Jensen, & Broberg 1988; Stranddorf, Møller, & Jensen 1992) og i det seneste årti systematisk arbejdsmiljøledelse (Frick & Wren 2000; Hasle & Zwetsloot 2011). Dette ekstra led i kæden er på den ene side nødvendigt for at kunne påvirke det stadig mere komplekse arbejdsmiljø, og på den anden side er arbejdsmiljøarbejdet i virksomhederne en mangesidig størrelse som vanskeligt lader sig indfange i simple modeller.

Hvis pilene mellem de forskellige led i virkningskæden var stærke kausalitetsmekanismer, ville sammenknytningen mellem kædens yderste led også være relativt robust. Men ingen af modellens pile er stærke kausalitetsmekanismer: Vi ved fx ikke præcist hvordan forskellige reguleringsstile – typisk benævnt kontrollerende henholdsvis understøttende (Nielsen et al. 2010) – påvirker arbejdsmiljøet i små og mellemstore virksomheder. Vi ved ikke præcist om et systematisk ledelsesarbejde som vi fx ser det i certificerede virksomheder, giver et godt arbejdsmiljø. (Hasle & Zwetsloot 2011; Robson et al. 2007). Vi ved ikke helt præcist hvilke forhold i arbejdsmiljøet, der udløser stressrelaterede sygdomme. (Briner & Reynolds 1999; Semmer 2006). Vi kender en del til koblingen mellem sundhed og sygefravær, men vi ved også at den er mere kompliceret end bare et spørgsmål om at manglende sundhed giver sygefravær. At etablere viden om virkninger mellem de forskellige kasser blot to og to er altså langt fra et banalt problem. Som en konsekvens er vores viden om den samlede virkningskæde naturligvis også meget usikker.

Vores interesse er altså ikke blot at undersøge om forskellige arbejdsmiljøtiltag får den forventede virkning eller ej. Vi ønsker med denne publikation at blive klogere på hvordan man i evaluering kan belyse *sammenhænge* mellem de valgte virkemidler (som fx systematisk ledelsesarbejde eller en ændret organisering af arbejdsmiljøorganisationen) og deres effekt i praksis. Det er derfor nødvendigt at få en forståelse af hvordan virkemidler i arbejdsmiljøindsatsen kan virke. Derved bliver det muligt at tilrettelægge evalueringer som netop kan belyse virkningen af en arbejdsmiljøindsats. Inden vi ser nærmere på virkemidler og mulighederne for at evaluere deres virkning, beskriver vi i det næste kapitel forskellige tilgange til evaluering, og vi diskutere på det grundlag styrker og svagheder ved de forskellige tilgange.

4. UDVIKLING AF EVALUERINGSSTEORI

Det er et komplekst forehavende at evaluere arbejdsmiljøindsatser. Arbejdsmiljøfeltet har sine særlige forhold der komplicerer evalueringsopgaven; men evalueringer som sådan har også aspekter der i sig selv gør det at evaluere til en vanskelig opgave. I det følgende undersøger vi hvad evalueringsteorien og den tilknyttede værktøjskasse kan tilbyde i forhold til arbejdsmiljøfeltet. Vi redegør her for udviklingen af evalueringsteorien med særligt fokus på fire væsentlige evalueringsparadigmer som succesivt har domineret reguleringsfeltet, og vi slutter af med en diskussion af mulighederne for at anvende de fire paradigmer til evaluering af arbejdsmiljøindsatser.

Udviklingen af evalueringsteori kan sammenfattes i fire bølger som hver indeholder forståelser af hvad evaluering skal bruges til, og hvordan den kan gennemføres (Krogstrup 2011):

- Den klassiske effektevaluering
- Procesevaluering
- Monitorering og evaluering
- Evidensbaseret evaluering

Som altid med denne type opdelinger er der ikke altid skarpe skillelinjer. Der er overlap, og nye tendenser bygger typisk videre på de foregående, samtidig med at disse lever i bedste velgående. Samlet set giver de fire bølger dog et udmærket billede af hvordan udviklingen er forløbet op til i dag. Vi tager afsæt i vores præsentation af de fire tilgange primært med afsæt i Krogstrup (2011) (se også www.bar-eva.dk).

4.1 DEN KLASSISKE EFFEKTEVALUERING

Evaluering både som begreb og konkret aktivitet opstod i 1960erne - først i USA med en hurtig spredning til Europa. På det tidspunkt herskede en forestilling om en konstant udvikling mod et bedre og rigere samfund hvor det handlede om at finde frem til den 'rigtige' metode som kunne sikre den fortsatte positive udvikling. I USA blev det formuleret som "the experimental society" hvor systematiske forsøg kunne føre til afprøvning af forskellige politikker, og man derved kunne eksperimentere sig frem til den rigtige måde at gøre tingene på. Der blev derfor udviklet et evalueringsdesign som handlede om at måle situationen før en bestemt intervention, gennemføre interventionen og efterfølgende gennemføre en udfaldsmåling som skulle påvise effekten af interventionen. Dette design forfines ved at anvende en kontrolgruppe hvor der ikke gennemføres intervention, og effektmålingen styrkes ved at sammenligne før- og eftermålinger for interventionsgruppen og for kontrolgruppen. Denne type evaluering bygger på kvantitative dataindsamlings- og analysemetoder. I den klassiske effektevaluering er objektivitet et centralt tema, og det indebærer at evaluator ikke må påvirke interventionen.

Det er imidlertid også et begreb som allerede på et tidligt tidspunkt blev kritiseret. Blandt de væsentligste kritikpunkter kan fremhæves (Krogstrup 2011, se også Pawson 2006):

- Samfundsmæssige programmer er tæt knyttet sammen med politik og politiske ideologier. Af denne grund kan det være vanskeligt, at gennemføre evalueringen som et vellykket eksperiment. Man kan spore en tendens til at de programmer der har politisk bevågenhed, gennemføres af politiske årsager, og evalueringens resultater kommer i anden række.
- Det er metodisk vanskeligt at adskille betydningen af de samfundsmæssige interventioner og udviklinger som i øvrigt sker i samfundet. Der vil altid ske en påvirkning fra andre samfunds-

mæssige udviklinger som kommer til at påvirke evalueringseresultatet, og det bliver dermed ikke klart hvad der faktisk har forårsaget de identificerede forandringer.

- Dette kritikpunkt blev forsøgt imødegået ved at etablere kontrolgrupper, men de er vanskelige at etablere i de komplekse sammenhænge som samfundsmæssige interventionen afprøves i. Der er i praksis sjældent kontrolgrupper som kun adskiller sig i forhold til den gennemførte intervention.

Denne kritik førte efterhånden til nye evalueringsskemaer som vi præsenterer og diskuterer i det følgende, men den klassiske effekt-evaluering kom også til at danne afsæt for mere avancerede effekt-evalueringer som også anvendes indenfor arbejdsmiljøområdet i både evaluering- og forskningsmæssig sammenhæng, og effektforståelsen ligger i høj grad til grund for anvendelsen af evidensbegrebet. Vi vender tilbage til de nyere udviklinger af tilgangen i afsnit 4.4 om "Evidensbaseret evaluering".

4.2 PROCES-EVALUERING

Proces-evaluering er en evalueringstilgang der udviklede sig som en reaktion på den klassiske effekt-evaluering, og hvor man så at sige hoppede over i den modsatte grøft. Fra midten af 1970'erne begyndte man at fokusere på de faktiske programmer i samfundet som politikerne under alle omstændigheder etablerer, og formålet med evaluering blev derfor set som forsøg på at bidrage til at gøre programmer så effektive som muligt. Udgangspunktet var at det ikke blev anset som muligt at udskille egentlige sluteffekter af et program på grund af de mange faktorer som influerer programmet. Ud fra den forståelse bliver det væsentligste at give feedback til de interessenter som beskæftiger sig med programmet, således at de får muligheder for at forbedre implementeringen. Denne form for evaluering kaldes i evalueringsteorien også responsiv evaluering eller formativ evaluering. I almindelig sprogbrug anvendes imidlertid begrebet proces-evaluering, og det virker derfor mest meningsfuldt at fastholde anvendelsen af dette udtryk. Dog er det vigtigt at være opmærksom på at den praktiske anvendelse af begrebet i nogle tilfælde også rækker ind over virkningsevaluering som introduceres i afsnit 4.4. Det gælder når proces-evalueringen ikke anvendes til at give feedback undervejs, men til at vurdere procesforløbet i en afsluttende evaluering. Her anvender vi proces-evaluering som udtryk for en evalueringsskema hvis primære formål er at give feedback til deltagerne og beslutningstagere undervejs i en indsats.

Proces-evaluering betjener sig først og fremmest af kvalitative metoder i form af interview og forskellige former for observation. Kvantitative metoder anvendes dog også. Fx kan det være relevant at anvende spørgeskemaer til at finde ud af om deltagerne i et program har hørt om en indsats og er blevet inddraget. Proces-evaluering med kvalitative metoder indebærer at evaluator ikke kan holde sig adskilt og dermed forholde sig objektivt til programmet. Tværtimod bliver evaluator netop inddraget i programmet, for derigennem at kunne forstå hvad der foregår, og for efterfølgende at kunne give feedback til deltagerne således at de kan forbedre deres indsats.

Der er forskellige holdninger til hvorledes evaluators inddragelse skal håndteres. I den mest oprindelige version skal evaluator godt nok være til stede i feltet for derved at observere hvad der foregår, men uden i øvrigt at blande sig, og opgaven består i det tilfælde i at videregive observationerne til interessenterne som derefter selv fortolker og beslutter sig til, hvilke konsekvenser disse måtte have. I senere retninger inddrages evaluator i forskelligt omfang og sammenhænge mere aktivt i feedback mekanismerne. Der kan således være tale om at facilitere at interessenterne selv i høj grad står for evalueringen - fx ved at organisere dialogaktiviteter eller ved at træde ind i forskellige andre varianter af faciliterende eller coachende roller. Inden for forskningen er denne rolle bl.a. udtrykt i aktions-

forskning (Gustavsen 1992;Nielsen & Svensson 2006) som lægger op til at felt og forsker i fællesskab ændrer virkeligheden. Derigennem opnår de både en ny og mere ønskværdig virkelighed, og en forståelse af hvordan denne virkelighed kan påvirkes.

Også procesevaluering er blevet udsat for kritik på flere punkter:

- Set fra den klassiske effektevaluering er den responsive evaluering uvidenskabelig, idet det bliver tilfældigt hvad der fokuseres på, og fravalget af effektmåling betyder at forsøgene på at optimere ikke har et grundlag af viden om effekt at basere sig på.
- Ud fra tilgangens egne forudsætninger kan det desuden være problematisk at evalueringen i for høj grad styres af evaluators og/eller interessenters egne præferencer, og evalueringen kan dermed bl.a. få en konserverende effekt fordi der ikke stilles spørgsmålstejn ved programmet som sådan og optimeringen sker indenfor de givne rammer.

Tilgangens styrke ligger primært i det læringspotentiale den kan skabe i det evaluerede felt gennem feedback-mekanismerne, mens mulighederne for generalisering er mere begrænsede.

4.3 MONITORERING OG EVALUERING

I takt med fremvæksten af New Public Management (NPM) i 1980'erne vokser også den næste evalueringebølge frem. NPM henter inspiration fra den private sektor og lægger vægt på markedsgørelse og ansvarliggørelse af de relevante ledere. Dette implementeres ofte gennem målstyring. Det bliver i den sammenhæng væsentligt at finde ud af om offentlige institutioner opfylder deres mål, om borgerne får de ydelser de har krav på, og om leverandører leverer det de har lovet. Det bliver derfor nødvendigt at udvikle målbare parametre som kan belyse, om målopfyldelsen finder sted. Typisk skrives målene ind i resultatkontrakter for de enkelte institutioner, og deres ledere, og institutionernes budgetter (og lederlønninger) bliver afhængige af graden af målopfyldelse. I de seneste års version er man i stigende grad – igen inspireret af den private sektor – begyndt at anvende *Key Performance Indicators* (KPI'er) som omdrejningspunkt for opfølgningen på målene. Løbende overvågning af de opstillede parametre og mål betegnes monitorering, og monitorering er blevet et integreret element i styringen af den offentlige sektor. På arbejdsmiljøområdet får vi som en konsekvens af handlingsprogrammet "Rent arbejdsmiljø år 2005" (Arbejdstilsynet 2007b) en markant forøget indsats, når det gælder "statistik og dokumentation" fra omkring 1999 og frem. Det udmøntes bl.a. i årlige overvågningsrapporter.

Monitoreringen støder dog ind i nogle af de samme barrierer som den klassiske effektevaluering, og som også er en del af årsagen til, at det er vanskeligt at evaluere arbejdsmiljøindsatser. Der er langt fra de konkrete aktiviteter til de slutmål som offentlige institutioner skal levere: fx skoleelever som fortsætter i en videregående uddannelse eller reduktion i antallet af førtidspensionister. Derfor kommer monitoreringen til at fokusere på procesvariable som fx andelen af langtidssygemeldte som har været til samtaler eller deltaget i arbejdsprøvning, mens selve den ønskede effekt ikke bliver målt. En anden version er at lave brugertilfredshedsmålinger som kan være ganske vanskelige at knytte til de egentlige mål for institutionen. Fx viser patienttilfredshedsmålinger typisk høj tilfredshed³, men spørgsmålet er, om denne tilfredshed udtrykker behandlingstilfredshed eller patienternes oplevelse af at personalet, de møder, optræder sympatisk. Det bliver simpelthen for dyrt og for langsigtet i forhold til både det daglige arbejde og de årlige budgettildelinger hvis målstyringen skulle forholde sig til de langsig-

³ <http://patientoplevelser.dk/index.asp?id=499>

tede effektmål. Det samme gælder typisk for monitorering af arbejdsmiljøindsatser. Her kan man sjældent koble en konkret indsats sammen med et udfald i effekt. Monitorering anvendes derfor først og fremmest i den generelle overvågning fx i forhold til 2020-arbejdsmiljøstrategien hvor de overordnede tendenser kan følges, men uden at årsagerne til tendenserne kan identificeres ud fra selve overvågningen.

Der kan derfor stilles spørgsmålstejn ved om monitorering i snæver forstand er et evalueringsparadigme (Krogstrup 2011), men som med de øvrige evalueringsbølger er den udtryk for et ønske om at lave systematisk vidensindsamling som gennem en bedømmelse, kan anvendes som fremtidigt beslutningsgrundlag. Desuden giver den mulighed for at følge med i langsigtede udviklingstendenser hvilket de øvrige evalueringsparadigmer i mindre grad er velegnede til. Monitorering er samtidig en tilgang til evaluering som i høj grad danner grundlag for nutidens dominerende evalueringsbølge.

4.4 EVIDENSBASERET EVALUERING

Efter årtusindskiftet er evidensbaseret evaluering i særlig grad sat i højsædet. Et krav om at offentlige midler skal anvendes hvor der er evidens for effekten, er på mange måder en naturlig forlængelse af New Public Management og monitoreringsbølgen. Det gælder ikke mindst indenfor arbejdsmiljøet hvor den tætte tilknytning til det biomedicinske forskningsparadigme skaber en særlig grobund for kravet om evidens for arbejdsmiljøindsatser. Der er imidlertid generelt set ikke enighed om hvordan evidens skal forstås, og der udspiller sig ofte heftige diskussioner om hvornår der er tilstrækkelig evidens for at iværksætte særlige tiltag. Det er også en velkendt diskussion fra arbejdsmiljøområdet hvor der ofte er diskussioner af evidensen for om noget er skadeligt for helbredet, og om bestemte forebyggende foranstaltninger har en effekt eller ej. I nogle tilfælde afløser en teknisk diskussion af sådanne spørgsmål den politiske uenighed som faktisk ligger bag. Vi vender senere tilbage til en bredere diskussion af evidensbegrebet, mens vi her fokuserer på betydning for evalueringsparadigmerne. Der er særligt to grundlæggende forskellige positioner som dominerer debatten om evidensbaseret evaluering. Den ene bygger på den klassiske effektevaluering hvor det randomiserede kontrollerede forsøg (randomized control trials – RCT) er blevet udpeget som den gyldne standard. Den anden position betegner vi virkningsevaluering, og den bygger i høj grad på realistisk evaluering (Pawson 2006) som fremhæver viden om mekanismer og kontekst som afgørende for evidens.

EFFEKTEVALUERING

RCT-tilgangen henter i høj grad inspirationen fra USA, og den bygger på den medicinske forskning hvor effekten af medicin undersøges gennem sammenligning af den aktive ingrediens og kalkpiller (placebo). Der trækkes lod mellem deltagerne, og forskerne ved ikke hvem der får hvilken type piller. Ideen med denne metode er at sikre fuldstændig sammenlignelighed, således at den eneste forskel er medicinen (interventionen). Metoden er indenfor medicinen fuldstændig gennemprøvet og standardiseret. Med inspiration herfra overføres metoden i dette evalueringsparadigme til andre anvendelsesområder, som både omfatter individer og organisationer. Fx anvendes tilgangen nu også indenfor en række socialpolitiske felter. Her kan der fx trækkes lod mellem belastede unge om deltagelse eller ikke-deltagelse i afprøvning af forskellige støtteformer. Med hensyn til organisationer kan man fx trække lod mellem skoler eller klasser om afprøvning af bestemte undervisningsmaterialer. På arbejdsmiljøområdet er metoden først og fremmest anvendt til interventioner som retter sig mod enkeltindivider. Et typisk eksempel kan være tilbud om træning som afgøres ved lodtrækning, og hvor man eventuelt efterfølgende giver kontrolgruppen det samme tilbud. Man kan også tænke sig randomisering på organisationsniveau, hvor man fx trækker lod mellem teams eller mindre enheder om deltagelse/ikke deltagelse. Metoden kan dog ikke anvendes i den ideelle form, fordi de deltagere som gennem lod-

trækningen ikke får interventionen, er bekendt hermed og dermed på forskellig måde kan blive påvirket af situationen. Netop fordi det kan være svært at randomisere, anerkendes også metoder med anvendelse af "ikke-randomiserede" kontrolgrupper samt såkaldte "naturlige eksperimenter" og før-og-efter-målinger som relevante - om end metodemæssigt svagere muligheder. I denne tilgang anvendes i høj grad systematiske review af litteraturen hvor de anvendte forskningsmetoders kvalitet bedømmes, og der er udarbejdet internationale modeller for hvorledes det skal gennemføres i praksis. Den medicinske model betegnes Cochrane⁴-modellen og den samfundsvidenskabelige Campbell⁵-modellen, og begge modeller er veletablerede med solide institutionelle understøtninger.

VIRKNINGSEVALUERING

I de seneste år har specielt Dahler-Larsen og Krogstrup (2004) i Danmark udviklet begrebet virkningsevaluering. Det svarer på mange måder til det internationale begreb realistisk evaluering som først og fremmest er beskrevet af Pawson og kolleger (Pawson 2006; Pawson & Tilley 1997; Pawson, Wong, & Owen 2011). De to evalueringstilgange er ikke fuldstændig overlappende, men det er vores vurdering at anvendelsen af udtrykket virkningsevaluering i høj grad vil dække over den samme forståelse som ligger bag realistisk evaluering, og vi har derfor valgt at fastholde virkningsevaluering som det udtryk der anvendes på dansk. I det følgende diskuterer vi virkningsevaluering med udgangspunkt i den engelsksprogede litteratur om realistisk evaluering.

Udgangspunktet for denne forståelse af evaluering er at der indenfor samfundsmæssige interventioner yderst sjældent findes sikre effekter som er kontekstafhængige. Det betyder, ifølge den realistiske evaluering, at selv den bedst gennemførte effektevaluering med anvendelse af RCT-lignende metoder ikke kan udtale sig om effekten udenfor den kontekst interventionen er gennemført i. Det væsentlige består derfor i at finde frem til de mekanismer som i samspil med konteksten, kan føre til et bestemt resultat. Dermed bliver det muligt at finde ud af hvad der virker, og hvordan det virker i en konkret kontekst, og gennem forståelsen af mekanismerne kan denne viden overføres til andre sammenhænge. Virkningsevaluering benytter sig af både kvalitative og kvantitative metoder.

For realistisk evaluering er systematiske review også et kerneelement (Pawson 2006). Argumentet herfor er at man gennem sammenligning med flere tilsvarende indsatser kan udtale sig mere generelt om den konkrete indsats' virkning. Det giver en bedre mulighed for at få en forståelse af de generaliserbare mekanismer idet en enkelt evaluering sjældent vil være tilstrækkeligt omfattende til at kunne tegne et generelt billede af de relevante mekanismer. Sammenligning af flere tilsvarende indsatser vil imidlertid kunne give mere pålidelig viden om indsatsens virkning. Mens effektevalueringen har en mangeårig tradition bag sig, er den realistiske evaluering endnu relativt ung, og erfaringerne med at anvende metoden er under opbygning. Inden for arbejdsmiljøområdet er der endnu kun spredte eksempler på denne type evaluering (Se fx Olsen, Legg, & Hasle 2012; Pedersen, Nielsen, & Kines 2012). En af vanskelighederne som virkningsevaluering skal overvinde – ikke mindst i arbejdsmiljøsammenhæng – er at få en mere præcis beskrivelse af hvad en mekanisme er, og hvordan den fungerer. Et af midlerne hertil er en systematisk anvendelse af programteori, således som vi beskriver i kapitel 5. Holdbarheden af denne programteori kan efterfølgende undersøges i evalueringen, idet man får et grundlag for at undersøge mekanismer, da programteorien på forhånd har opstillet hypoteser for de mekanismer, som skal få interventionen til at virke.

⁴ <http://www.cochrane.org>

⁵ <http://www.campbellcollaboration.org>

4.5 PRAGMATISK EVALUERING

I den praktiske verden foretages kun få evalueringer som klart følger det ene eller det andet paradigmes metodekrav. Når det sker, drejer det sig typisk om egentlig forskning som stiller større krav til systematisk metodeanvendelse. De fleste evalueringer plukker fra hylderne og indeholder både kvantitative og kvalitative elementer. Kvantitative effektmål findes måske i den løbende monitorering af en institutions aktiviteter, eller måske gennem spørgsmål til brugere som spørges til deres holdning. Parallelt hermed gennemføres kvalitative interview af hvad der foregår i praksis, og som dermed måske kan sige noget om mekanismer som er i spil. Disse data anvendes ofte til midtvejsevalueringer som forsøger at give feedback for derved at kunne styrke implementeringen af indsatsen. Årsagen er udpræget pragmatisk: Det er begrænset hvor mange ressourcer der kan anvendes til evaluering, og der er typisk mange interesser som evalueringen skal tilgodese. Indvendingerne mod den pragmatiske fremgangsmåde kan være at den sætter sig mellem to stole og spreder de begrænsede ressourcer for meget ud. Det kan være problematisk ud fra et klassisk synspunkt at effektmålene bliver for upræcise og tilfældige, fordi de ikke er fremkommet gennem anvendelsen af tilstrækkeligt systematiske metoder, som bl.a. sikrer effektivt mod risikoen for *bias*; altså skævvredne resultater. Som følge heraf ved man derfor ikke præcist, om effekten skyldes interventionen eller noget andet. Og tilsvarende vil begrænsede interview af de involverede parter måske give et billede af deres oplevelser, men de vil ikke kunne svare på hvilke mekanismer som faktisk har drevet processen igennem. Vi vender i slutningen af kapitel 8 og i kapitel 9 tilbage til en diskussion af hvordan pragmatisk evaluering kan tilrettelægges, således at den får mulighed for at opnå en tilstrækkelig høj kvalitet til at kunne anvendes i praksis.

4.6 ANVENDELIGHEDEN AF EVALUERINGS-PARADIGMERNE

Spørgsmålet er hvordan de forskellige paradigmer kan anvendes i praksis. I overvejslen heraf tager vi udgangspunkt i en forståelse af de fleste arbejdsmiljøindsatser som såkaldte 'vilde' problemer, og vi kobler denne forståelse sammen med mulighederne for at evaluere arbejdsmiljøindsatser ved hjælp af de forskellige evalueringssparadigmer.

Evaluering som et både praktisk og teoretisk felt har en historie på et halvt hundrede år, og evaluering har udviklet sig fra at være enkeltstående, relativt sjældne aktiviteter til at være et gennemgribende element i nutidens samfund. Evaluering har nu fået et sådan omfang at vi nu er på vej ind i et 'evalueringssamfund' (Dahler-Larsen 2003). Nu igangsætter beslutningstagere – specielt i den offentlige sektor – sjældent aktiviteter uden de på den ene eller anden måde bliver fulgt op af en evaluering, og sjældent (i hvert fald i princippet) uden at aktiviteterne kan begrundes med evidens fra tidligere evalueringer, der viser at de med en vis sandsynlighed kan levere de resultater som forventes. Denne udvikling præger også arbejdsmiljøområdet hvor der også stilles stadig større krav til evaluering. Det gælder ikke mindst, som tidligere beskrevet, 2020-arbejdsmiljøstrategien hvor en række punkter indeholder krav om evaluering.

Arbejdsmiljøområdet er dog – som så mange andre områder – i den situation at der sjældent er simple kausale sammenhænge mellem en konkret indsats og den ønskede effekt på sikkerhed og sundhed. Årsags-virkningskæden er lang med mange muligheder for brud undervejs, og det gør det vanskeligt at være sikker på sammenhængen mellem indsats og effekt. Et simpelt eksempel er afskærmning af rundsave. Som konkret løsning er det ganske lige til: En skærm forhindrer at fingrene får kontakt med savklingen. Men spørgsmålet om implementeringen af indsatsen kan forstås som et vildt problem. Kravet om afskærmning er en af de ældste arbejdsmiljøregler, men branchen kæmper stadig med at få

den overholdt, fordi der i praksis er en række forhold omkring produktivitet, viden, kultur og professionsidentitet, som betyder at der alligevel mangler skærme på mange rundsave.

I vurderingen af de enkelte evalueringsparadigmer kan der således være behov for at undersøge en række forskellige forhold. Med udgangspunkt i rundsavseksemplet kan der være behov for at vurdere forhold i direkte tilknytning til afskærmning som løsning på et sikkerhedsproblem. Det drejer sig om løsningen overhovedet kan løse problemet (fjerner afskærmningen den dominerende risiko), i hvilken udstrækning løsningen implementeres, og om ulykkestallet faktisk falder som følge af implementeringen. Men derudover kan det eksempelvis være relevant at spørge hvordan det kan lade sig gøre at få implementeret afskærmning, om omkostningerne ved forskellige indsatser for at få implementeret løsningen, og om implementeringen er bæredygtig på langt sigt. En evaluering kan næppe svare på alle disse spørgsmål, og de forskellige evalueringsparadigmer vil egne sig til at belyse nogle af disse spørgsmål og i mindre grad eller slet ikke andre.

I det følgende diskuterer vi kort de væsentligste styrker og svagheder ved paradigmerne i forhold til at evaluere arbejdsmiljøindsatser.

STYRKER OG SVAGHEDER VED EVALUERINGS PARADIGMERNE

Den klassiske effektevaluering i sin rene form spiller i dag ikke nogen stor rolle, og der er derfor fire paradigmer som er væsentlige i praksis: procesevaluering, monitorering, effektevaluering og virkningsevaluering hvor de to sidste er forskellige tilgange til evidensbaseret evaluering.

Procesevaluering vil være i stand til at beskrive og følge det flow en indsats udvikler sig gennem, og den vil have mulighed for at bidrage til at udvikle processerne mellem de forskellige interessenter og mellem aftager- og modtagerniveau. En svaghed vil være at procesevaluering måske fokuserer på en smal kontekst, og den dermed ikke kan følge en særlig indsats helt fra politisk program til effekt, og den vil kunne komme til at fokusere på alle mulige andre tiltag der kan have påvirket virksomhedskonteksten. Samtidig vil denne tilgang have svært ved at give oplysninger om effekten af en indsats.

Monitorering bruges typisk til at følge mere langsigtede udviklingstræk, og svagheden er at der ofte er langt fra de konkrete aktiviteter til de slutmål man ønsker at kunne udtale sig om. Målingerne kan derfor ofte ende med at blive statiske og ikke opleves som relevante for de konkrete indsatser. Samtidig kan de dog have stor politisk gennemslagskraft, fordi de kan belyse langsigtede udviklinger der på den ene eller anden måde er behov for at gribe ind overfor.

Den evidensbaserede evaluering indeholder som nævnt to meget forskellige paradigmer – effektevaluering og virkningsevaluering. En klassisk *effektevaluering* vil måle situationen før og efter en bestemt intervention for at finde frem til hvad der kom ud af interventionen – effekten. Styrken ved denne evalueringsform er at der fremkommer tal for effekt, og de vil ofte have stor politisk gennemslagskraft. Men effektevalueringen har vanskeligt ved at tage højde for de mere eller mindre uforudsigelige kontekster som interventionen møder i praksis. Det kan derfor ikke eller kun i begrænset omfang vurderes hvad der er årsagen til at en given effekt opstår. Moderne projektdesign og avancerede statistiske analyser kan til en vis grad kompensere herfor, og det kan derigennem blive muligt at udelukke visse konkurrerende faktorer. Men dels vil der sjældent være mulighed for at anvende denne slags metoder i almindelige evalueringer, og dels har også disse metoder vanskeligt ved at forstå konteksten hvilket er nødvendigt, hvis man ønsker at generalisere en bestemt type indsats over til en anden kontekst.

Formålet med *virkningsevaluering* er også at opnå evidens som kan anvendes som grundlag for beslutninger. Her er hovedsigtet at finde frem til de mekanismer som i samspil med konteksten, kan føre til et bestemt resultat. Formålet er at finde ud af hvad der virker, for hvem og under hvilke betingelser

(kontekst). Besvarelsen af dette spørgsmål giver en forståelse af mekanismerne som kan anvendes til at overføre den opnåede viden til andre sammenhænge. Problemet er imidlertid at vores viden om mekanismerne i arbejdsmiljøssammenhænge er sparsom, og der forestår derfor et betydeligt udviklingsarbejde for at kunne udnytte mulighederne i dette evalueringsparadigme fuldt ud.

Ovenstående korte diskussion af styrker og svagheder ved de fire dominerende paradigmer peger på at ingen af dem kan forkastes som forkerte eller ubrugelige, og som følge heraf kan der heller ikke udpeges et enkelt paradigme som det rigtige. I Figur 3 sammenfatter vi styrker og svagheder ved de fire paradigmer.

Paradigme	Styrker	Svagheder
Procesevaluering (responsiv, formativ)	<ul style="list-style-type: none"> • mulighed for at forbedre indsatsen undervejs 	<ul style="list-style-type: none"> • begrænset viden om resultatet af indsatsen
Monitorering	<ul style="list-style-type: none"> • mulighed for at følge en langsigtet udvikling og dermed reagere på den 	<ul style="list-style-type: none"> • begrænset viden om årsager til udviklingen
Effektevaluering	<ul style="list-style-type: none"> • viden om resultatet af en indsats 	<ul style="list-style-type: none"> • begrænset viden om årsagerne til det fremkomne resultat
Virkningsevaluering (realistisk)	<ul style="list-style-type: none"> • viden om de mekanismer som i forskellige kontekster fører til resultater 	<ul style="list-style-type: none"> • vanskeligt at generalisere mekanismer og deres relation til konteksten

Figur 3: Styrker og svagheder ved de fire evalueringsparadigmer

Et væsentligt problem i evalueringssammenhæng er at det er omkostningsfuldt at gennemføre gode og gyldige evalueringer, og at beslutningstagerne ofte vil ønske at få resultater som bedst opnås gennem evalueringer der indeholder elementer af flere eller alle paradigmerne. Ofte ender evalueringer derfor i det femte paradigme – den pragmatiske evaluering. Det kan der være gode begrundelser for, men der er også en oplagt risiko for at de enkelte elementer i evalueringen bliver for overfladiske. Den information som kommer ud af en for sammensat evaluering med ønsker om lidt fra alle hylder, kan være så usikker at den reelt ikke bidrager til at forbedret beslutningsgrundlag. Denne problemstilling diskuterer vi yderligere i de næste kapitler.

5. HVORDAN VIRKER ARBEJDSMILJØINDSATSER?

Der er altså adskillige mulige tilgange til at evaluere arbejdsmiljøindsatser, og det bliver dermed nødvendigt med et grundlag for at vælge en relevant tilgang og designe en konkret evaluering. Dette grundlag handler om en forståelse af hvordan arbejdsmiljøindsatser virker. Vi vil derfor præsentere et begrebsapparat som kan anvendes til at beskrive hvordan arbejdsmiljøindsatser virker. En sådan beskrivelse giver mulighed for at forstå sammenhængen mellem indsatsen og indsatsens effekt, og den er dermed afgørende for at kunne tilrettelægge evalueringer på en sådan måde at de giver ny viden som kan anvendes til at forbedre kvaliteten af de næste indsatser. Dette begrebsapparat tager udgangspunkt i Pawsons (1997; 2006) teori for realistisk evaluering (virkningsevaluering) som vi præsenterede i forrige kapitel. Et af de centrale spørgsmål, vi rejser i det følgende, er hvorfor arbejdsmiljøindsatser ofte har meget forskellig virkning. Hvorfor har nogen indsatser en stor effekt på én gruppe virksomheders arbejdsmiljøindsats, mens andre synes uberørte? Vi indleder med en konstrueret historie som kan belyse dette spørgsmål. Det er baseret på erfaringer fra projektet: "Ergonomi på mindre mejerier". Der er der tale om et netværk af mindre mejerier, som har gennemført et prisvindende projekt hvor de eksterne aktører i stort omfang har planlagt og gennemført indsatsen på en forbilledlig måde. Det indeholder spændende nytænkning bl.a. gennem forankring i et netværk af virksomheder der har givet hinanden håndslag på at samarbejde. Selv sådanne projekter giver dog ingen garanti for at den ønskede effekt på arbejdsmiljøet opnås, når resultaterne vurderes på den enkelte virksomhed. Denne situation illustreres med følgende historie fra mejeriet "Solskin". Historien er som nævnt opdigtet, men bygger dog på vanskeligheder med implementeringen som faktisk viste sig undervejs i projektet.

"Hvorfor står maskinen i hjørnet?"

Mejeriet "Solskin" der er en del af "Arbejdsmiljønetværk for mindre mejerier", havde gennem en måned afprøvet det nye hjælpemiddel der tidligere var blevet udviklet på et andet mejeri. Hjælpemidlet var en vakuumløfter som skulle anvendes ved pakning af opskårne oste i plastkasser. Da Gerda overtog fra natholdet, skubbede hun det lidt klodsede og girafagtige instrument over i hjørnet. "Hvorfor flytter du den?" spurgte Inge, "den er da en rigtig god aflastning". Gerda tog kittel og handsker på, før hun svarede. "Ikke for mig, jeg synes det tager alt for lang tid, og så står den i vejen. Jeg tror ikke jeg kommer til at bruge den".

Lidt senere kom Finn, som er arbejdsmiljørepræsentant, forbi og bemærkede hjælpemidlet ovre i hjørnet. Han henvendte sig til Gerda: "Den er her jo, for at vi skal prøve den. Så jeg synes det ville være godt, hvis du arbejdede med den. Den er jo taget ud af bonussen så det betyder ikke noget at det tager lidt længere tid i starten." Han begyndte selv at skubbe rundt med redskabet. "Jeg kan da godt se at den er lidt uhandy, men det er jo også en prototype. Det bliver jo bedre når vi en gang får bygget den sammen med pakkelinjen, ligesom vi har løftere ovre ved de store oste. Dem bruger I da altid." "Hvornår bliver det?" spurgte Inge, "det ser jeg frem til." Finn drejede lidt på et af håndtagene før han svarede. "Det bliver nok ikke lige med samme. I dag vil alle supermarkederne have deres egne størrelser af papkasser. Vi er allerede oppe på 14 forskellige slags, og lige nu dur den jo kun til de blå plastkasser. Så enten skal vi have noget, der er mere fleksibelt, eller også må vi ha' branchen til at forhandle med supermarkederne." Han sparkede blidt til gummihjulet og kørte redskabet over i hjørnet igen.

Historien er opdigtet, men inspireret af projektet "Ergonomi på mindre mejerier".

Som historien illustrerer, er et godt projekt og en god teknisk løsning ikke altid nok. En række andre forhold er afgørende for om en arbejdsmiljøindsats i praksis ender med at forebygge belastninger i arbejdet. Det er derfor væsentligt at blive bedre til på forhånd at overskue hvordan en konkret indsats omsættes til forbedringer af arbejdsmiljøet. I projektet 'Ergonomi på mindre mejerier' var udviklingen af prototypen tænkt som et forpligtende samarbejde mellem de mindre mejerier. Udviklingsdelen synes også at være relativt succesfuld. I eksemplet viser problemet sig da prototypen skulle indsættes i den konkrete arbejdskontekst. Projektejerne har formodentlig ikke forestillet sig de praktiske omstændigheder i hvilke prototypen skulle fungere, og den måde medarbejderne ville modtage maskinen på. De har måske været mere fokuserede på selve udviklingen af prototypen. Eksemplet viser altså hvordan de implicite og eksplicite antagelser der ligger bag et konkret projekt, får betydning for effekten på arbejdsmiljøet af en indsats.

Disse antagelser kan man også kalde en *programteori*. Begrebet programteori dækker over antagelser om hvordan en indsats vil virke. Antagelser som i større eller mindre omfang, kan vise sig at holde stik når arbejdsmiljøindsatsen får sit eget – ofte uforudsigelige – liv i mødet med arbejdspladsernes praksis. Mejerihistorien viser netop at det ikke i alle tilfælde går som forventet. Men netop læringen fra sådanne erfaringer fra praksis kan anvendes til at forbedre indsatsen. Som Finn i historien peger på, kan det være en god idé at fortsætte udviklingen af hjælpemidlet med forsøg med mere fleksible løsninger hvilket dog kræver et samarbejde med supermarkederne.

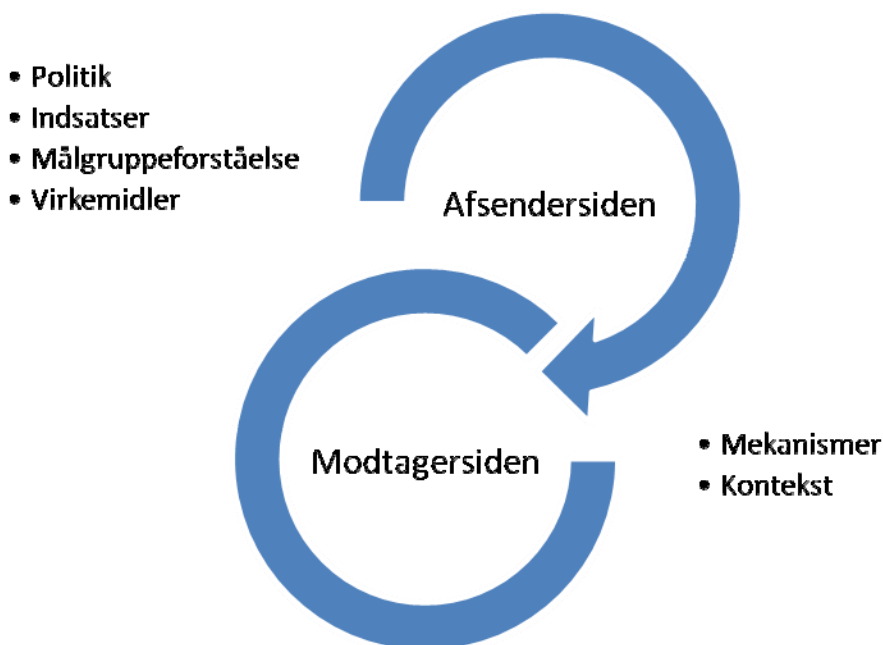
Forskning og gode evalueringer vil formodentlig på længere sigt kunne sige noget mere præcist om hvad der sker i mødet mellem arbejdsmiljøindsatser og praksis i virksomhederne. Dermed bliver der mulighed for at åbne den 'black box' (som det hedder i forskningsslang) som koblingen mellem arbejdsmiljøindsatser og praksis i virksomhederne udgør. Black box betegner det faktum at man ikke har umiddelbar (erkendelsesmæssig) adgang til hvorfor en forandringsproces finder sted, selvom man kan observere at der faktisk sker en forandring. Budskabet er at man i rollen som afsender af arbejdsmiljøindsatser, men også som evaluator af disse indsatser, skal forholde sig til hele forløbet fra afsender til modtager. Man skal således både forholde sig til afsendersidens intentioner og konkrete indsatser til de måder modtagersiden i praksis vil kunne formodes at modtage og omforme indsatserne til praksis. Programteorien omfatter på denne måde det samlede sæt af antagelser om hvordan en indsats vil virke.

5.1 FRA INDSATS TIL PRAKSIS

Vi skelner i det følgende mellem *afsendersiden* hvor politikker og indsatser udvikles og planlægges, og *modtagersiden* hvor indsatserne udfoldes og omsættes til praksis. Vi vil dog ikke konkret komme ind på konkrete politikker, men fokusere på de forskellige arbejdsmiljøaktører som udformer arbejdsmiljøindsatser. Det vil primært sige Arbejdstilsynets og arbejdsmarkedets parter samt de fælles institutioner i form af først og fremmest Arbejdsmiljørådet og BAR'ene. Andre aktører som fx Forebyggelsesfonden, aftalebaserede organer, forskere og forskellige konsulenter indgår også. Det er her de konkrete arbejdsmiljøindsatser udformes, og det er disse aktører som gør sig antagelser om hvad der vil virke, og hvordan det vil virke. Programteorien er arbejdsmiljøaktørernes antagelser om *hvad der vil virke, og hvordan det vil virke*. På afsendersiden betjener man sig af forskellige virkemidler som er de konkrete tiltag der udformer indsatsen, og bringer den fra ide til praksis. Antagelserne om hvad der vil virke, og hvordan det vil virke, findes altid, men de kan være mere eller mindre eksplicite og sammenhængende.

Arbejdsmiljøaktørerne skal altså forholde sig til deres egne antagelser om hvad der vil virke, hvilket i høj grad vil bygge på modtagersidens formodede reaktioner. Modtagersiden omfatter først og fremmest virksomhederne og deres arbejdsmiljøorganisation. Det er dog vigtig at nævne at indsatsen også kan rettes mod andre modtagere end netop virksomhederne. Et eksempel kunne være indsatsen der ønsker at skabe forandringer i måden hvorpå tekniske skoler underviser i arbejdsmiljø på. Her er de tekniske skoler modtagerne, men det ændrer ikke på modellen grundlæggende. På modtagerniveau konfronteres intentionerne bag indsatsen med den praktiske virkelighed. Dermed kommer den specifikke kontekst i virksomhederne og deres omgivelser til at have stor betydning for arbejdsmiljøindsatsens faktiske forløb. Disse sammenhænge er – set fra arbejdsmiljøaktørens perspektiv - grafisk afbilledet i en model i figur 4.

Mens indsatsen tager udgangspunkt i en politik for arbejdsmiljøet og indeholder bestemte virkemidler som er rettet mod bestemte målgrupper, er mekanismer og kontekst de centrale på modtagersiden. Mekanismerne kan forklare hvordan en arbejdsmiljøindsats fører til at den konkrete arbejdsmiljøpraksis udvikler sig på en bestemt måde. Nogle mekanismer kan være meget specifikke og kun gælde i et begrænset antal kontekster, mens andre måske kan genfindes i mange forskellige situationer. Pointen er at gennem identifikation af mekanismer som er generelle, kan man lære af dem på tværs af deres specifikke kontekst, selvom denne kontekst i sidste ende bliver afgørende for det endelige resultat. Når arbejdsmiljøaktørerne udformer en arbejdsmiljøindsats vil viden om sådanne mekanismer (det der sker, når indsatsen møder praksis) kunne forbedre sandsynligheden for, at indsatsen rent faktisk får den ønskede effekt.



Figur 4: Forståelsesmodel for arbejdsmiljøindsatser

Der er sjældent en direkte vej fra den centrale arbejdsmiljøaktør som står bag en indsats, til modtagerne i en virksomhed. Indsatsen vil forløbe i et flow hvor den passerer fra den ene aktør til den næste og videre til yderligere aktører, og når den ankommer til virksomhederne vil der også der være et flow fra den ene aktør til den næste. En ny regel fra Arbejdstilsynet vil fx bevæge sig fra et centralt niveau til de tilsynsførende, samtidig med at BAR'ene vil arbejde med vejledninger for hvordan reglen kan efterkommes, og arbejdsmiljørådgivere vil inddrage den i deres rådgivning. Alle disse aktører vil have deres egen forståelse af hvordan reglen skal fortolkes, og hvilken betydning den skal have i praksis. I

virksomheden vil reglen og dens fortolkning måske blive modtaget af en arbejdsmiljøkoordinator, som skal formidle den til en arbejdsmiljøorganisation som igen skal formidle videre til ledere og medarbejdere for at påvirke deres praksis. Det betyder at indsatsen efterhånden får sit eget liv og i større eller mindre omfang transformeres af aktørernes egen forståelse af hvordan den giver mening.

Denne forståelse – eller nærmere disse forståelser – kan derfor ofte ende med at være ganske forskellig fra de antagelser som lå i programteorien, og resultatet af indsatsen kan dermed blive ganske anderledes end forventet. Nogle indsatses fungerer kun hvis kæden af aktører i et vist omfang gennemfører indsatsen som forventet af de centrale igangsættere af indsatsen, andre indsatses kommer netop til at fungere fordi indsatsen transformeres og tilpasses gennem det flow den gennemløber frem til den endelige målgruppe.

Holdbarheden af de antagelser der ligger bag en indsats, spiller således en væsentlig rolle for hvor stort et gennemslag en konkret indsats vil få. I en evaluering vil et eftersyn af antagelserne derfor også være et væsentligt input til at forstå hvorfor en indsats lykkes, eller hvorfor den kun får en begrænset betydning. Når arbejdsmiljøaktøren udvikler en programteori der dækker både afsender- og modtagerside, kan sandsynligheden øges for at indsatsen i højere grad får den tilsigtede effekt og at evaluering er mulig. Det er i den forbindelse væsentligt at skelne mellem implementeringsfejl og teorifejl (Kristensen 2005).

- Implementeringsfejl: Indsatsen implementeres ikke som forudsat i programteorien (fx hvis den aldrig når ud til målgruppen, eller et mellemlid har omformet indsatsen til noget helt andet end intenderet)
- Teorifejl: Hvis indsatsen gennemføres som forudsat i programteorien, men ikke opnår den forventede effekt. (Fx hvis en meget teksttung pjece udsendes til en stor gruppe mindre virksomheder, men ingen læser den.)

Implementeringsfejl skyldes således en mangelfuld programteori, hvor gennemførelsen af forandringerne ikke er blevet gennemtænkt, fx fordi man ikke har gjort sig klart hvordan man når ud til målgruppen. Teorifejl skyldes en forkert programteori hvor forandringen ikke indebærer de forbedringer som man forestillede sig. I begge tilfælde kan der måske konstateres noget der ligner en effekt - positiv eller negativ - men det må i så fald siges at være en tilfældighed.

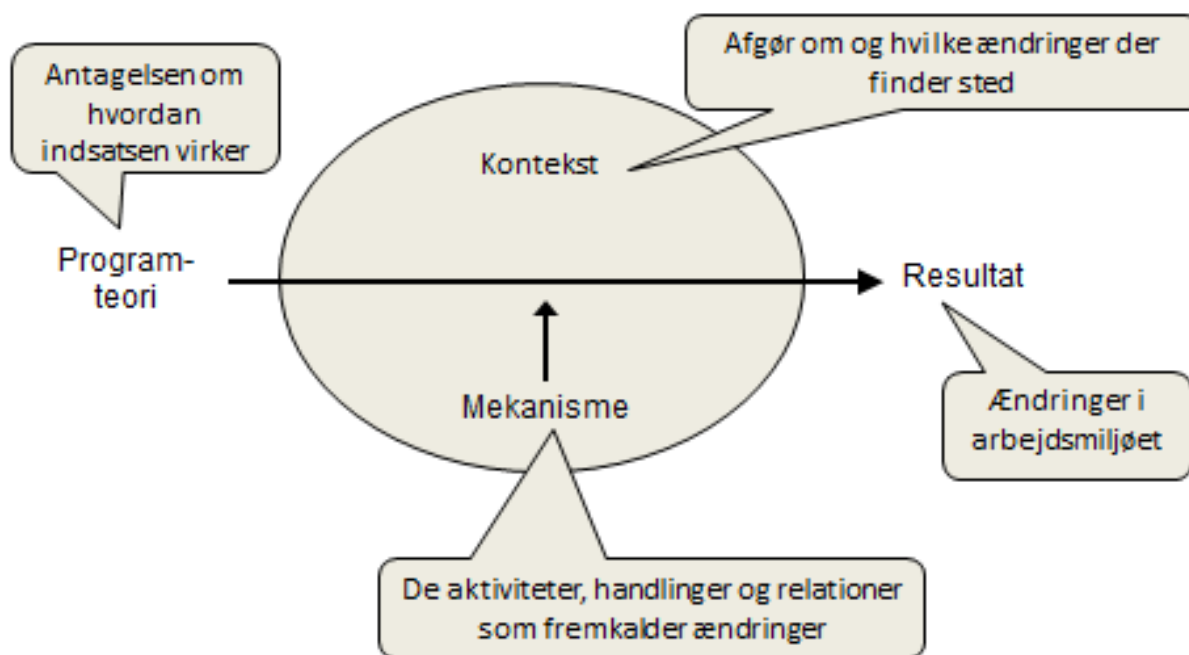
Med til det billede vi har forsøgt at tegne af konkrete arbejdsmiljøindsatser, hører, at det ofte kan være ganske svært at få øje på sammenhængene mellem intention og effekt i praksis. Derfor har vi i dette afsnit introduceret begreberne programteori og mekanisme. Sammen med en indsigt i den kontekst som en indsats udspiller sig i, er disse begreber med til at kunne give os en bedre forståelse af sammenhængen mellem indsats og effekt, der viser hvad der virker og hvordan. For at opnå dette er en skelnen mellem afsender og modtager central. Afsenderen er afhængig af en god programteori, der bl.a. er karakteriseret ved at forstå de mekanismer, som indsatsen prøver at aktivere i en bestemt kontekst hos modtagerne. Modtagerens perspektiv er et andet. Nogen udefra henvender sig med "et tilbud" om at gøre noget for arbejdsmiljøet. Der skal være match mellem egne behov, værdier og den konkrete situation, for at tilbuddet opleves relevant, og der skal være et samspil med de mekanismer på arbejdspladsen, som kan omsætte tilbuddet til konkrete forandringer.

Foreløbig har vi kun introduceret begreberne. I de følgende to kapitler vil vi definere dem nærmere – først ved at udfolde de overvejelser, der er nødvendige i afsenderperspektivet og i det følgende kapitel ved at udfolde, hvordan man kan betragte implementeringspraksissen fra et modtagerperspektiv.

6. ARBEJDSMILJØINDSATSER FRA ET AFSENDERPERSPEKTIV

6.1 PROGRAMTEORI

Med udgangspunkt i Pawson & Tilley (1997) kan hovedbegreberne i dette og næste kapitel illustreres med følgende model:



Fi

Figur 5: Pawson & Tilley's model for virkningen af sociale programmer (her arbejdsmiljøindsatser)

Set fra afsendersiden – og det er synsvinklen i dette kapitel – bliver arbejdsmiljøindsatsen sammenfattet i en programteori. En programteori kan defineres som det samlede sæt af antagelser om hvad der kan føre til en ønsket effekt af et givent socialt program, fx en arbejdsmiljøindsats.

Programteorien indeholder antagelser om hvordan de konkrete indsatser, og de virkemidler, de betjener sig af, kan fremkalde de ønskede ændringer (se også Eriksen, 2008). De konkrete arbejdsmiljøindsatser vil på modtagersiden udfoldes gennem en række mekanisme som er de handlinger og beslutninger der afgør, om og hvordan ændringerne finder sted. En god programteori omfatter derfor også antagelser om de mekanismer der kan bære en indsats igennem. Mekanismerne udspiller sig altid i en kontekst som afgør, hvordan arbejdsmiljøet ændres som følge af indsatsen. Den interne kontekst er forholdene indenfor virksomheden, men også den eksterne kontekst i form af bl.a. marked og branchens normer og værdier har betydning.

I arbejdsmiljøsammenhæng er to yderligere begreber væsentlige. Det første er *indsatser* som vi allerede løbende har anvendt. Det er et udtryk for det sæt af aktiviteter som centrale arbejdsmiljøaktører samler sammen med henblik på at opnå et bestemt formål.

Indsatser iværksættes bl.a. som led i den politik som den (eller de) pågældende aktør(er) har for arbejdsmiljøet. Det andet begreb er *virkemidler* der er de mere konkrete aktiviteter som de centrale arbejdsmiljøaktører har mulighed for at iværksætte for at opnå bestemte formål.

Virkemidler er en samlet betegnelse for formålsbestemte aktiviteter der tilsammen udgør en arbejdsmiljøindsats, og de bygger på erfaringer med hvad der kan skabe forandring. Virkemidler er aktiviteter som iværksættes indenfor et afgrænset tidsrum. Det er aktiviteter som bygger på viden (erfarings- og/eller forskningsbaseret) om at de kan påvirke målgruppen til at ændre holdning, adfærd og prioritering eller øge vidensgrundlaget og handlekraften med henblik på at opnå et givet mål.

En arbejdsmiljøindsats bygger således på at en afsender foretager en prioritering af ressourcer og vælger af et eller flere virkemidler der kan påvirke arbejdsmiljøet hos en bestemt målgruppe.

Enhver indsats bygger på en række antagelser (programteorien): Når vi gør sådan og sådan, sker der sådan og sådan. Den kan måske være udtalt, usammenhængende, urealistisk eller fragmenteret for så vidt som forskellige aktører på aftagersiden har forskellige og mere eller mindre eksplicite forståelser af hvorfor indsatsen vil virke. I mange andre tilfælde vil antagelserne mere eksplicite, men måske alligevel præget af mangelfulde og usammenhængende antagelser – kort sagt inkonsistente antagelser. Grundlæggende set bygger alle indsatser på antagelser, og antagelserne sætter rammer for hvordan indsatsen sættes i værk og formodes at virke. Det er væsentligt for en effektiv arbejdsmiljøindsats at de antagelser arbejdsmiljøaktørerne gør sig, bliver så reflekterede som muligt, samt at disse refleksioner ekspliciteres i en sådan grad at de udgør en dækkende og sammenhængende programteori. Det er kort sagt en fordel at udvikle hvad vi vil kalde en *refleksiv programteori*.

Sandsynligheden for at opnå de ønskede effekter er meget større hvis programteorien er god. Det er samtidig nemmere at lave en god evaluering hvis der foreligger en refleksiv programteori med eksplicite antagelser. Det modsatte af en refleksiv programteori er en programteori hvori antagelserne er implicite og dermed ikke reflekterede i samme grad.

Det er dog vigtigt at understrege at programteorier, lige meget hvor eksplicite de er, aldrig er lig med det faktiske forløb af en indsats. Tilfældigheder, aktørers handlinger og uintenderede konsekvenser spiller selvsagt til stadighed en stor rolle i forandringsprocesser.

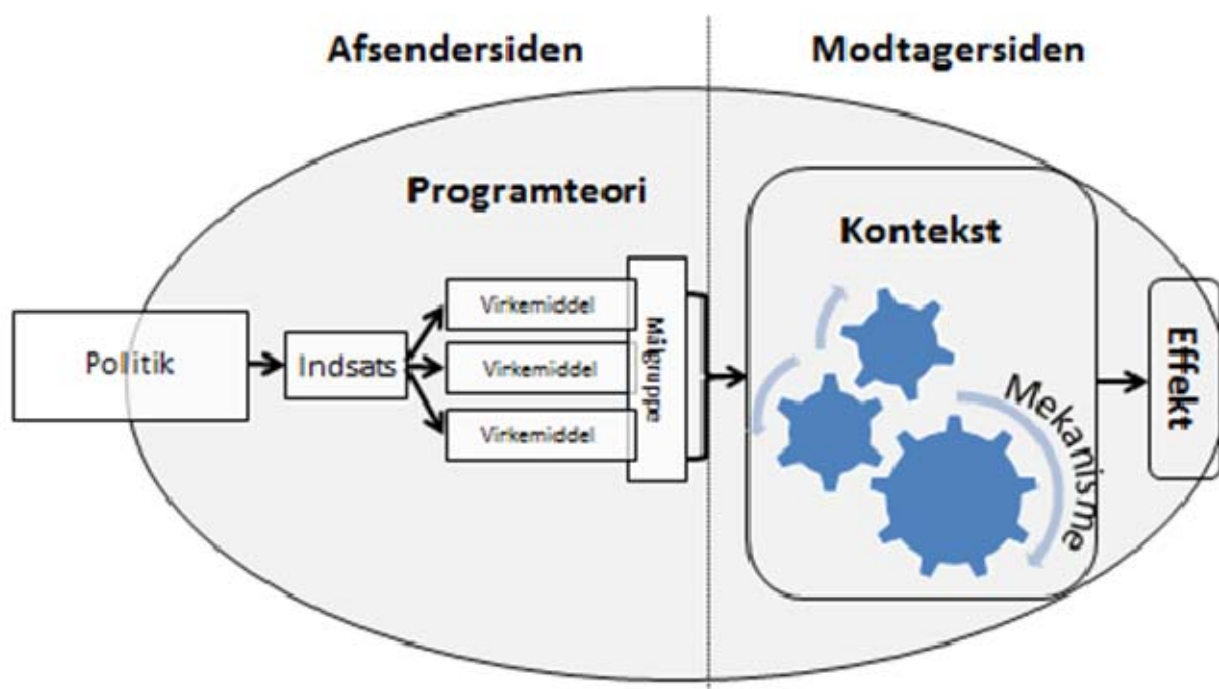
En god programteori vil bestå i en begrundet antagelse om hvordan kæden af eksterne aktører omsætter indsatsen på en sådan måde at der sættes en række mekanismer i gang i virksomhederne som medfører de ønskede ændringer i arbejdsmiljøet. I forhold til mejerieksemplet i forrige kapitel kan afprøvningen af en prototype og spredning af erfaringerne betegnes som en mekanisme der kan skabe ændringer. Mekanismen udspiller sig her i en særlig kontekst – netværket – som er afgørende for om ændringerne konkret kommer til udtryk. I fortællingen fra mejeriet "Solskin" er aftalen mellem den konkrete virksomhed og kunderne om typen af emballage et eksempel på hvordan den konkrete kontekst kan udgøre en barriere for at en ellers positiv ændring bliver realiseret.

I nedenstående tabel har vi sammenfattet definitionen af de centrale begreber vi nu har præsenteret. En nærmere uddybning af begreberne mekanisme og kontekst på modtagersiden følger i næste kapitel.

	Begreb	Definition
AFSENDERSIDEN	Programteori	Det samlede sæt af antagelser om hvordan en indsats vil virke. Heri indgår en fastlæggelse af mål, afgrænsning af målgruppen, fastlæggelse af virkemidler og antagelser om hvilke mekanismer som virkemidlerne kan få til at virke, og hvilken indflydelse den interne og eksterne kontekst har på mekanismerne. Programteorien beskriver hvorfor den givne indsats gør en forskel. Programteori kan være reflekteret hvis den er eksplicit og konsistent, og programteori kan være ikke-refleksiv hvis antagelserne er uformulerede og implicitte.
	Politik	Politik er de forskellige aktørers ønsker til hvordan arbejdsmiljøet skal udvikles. Den kommer til udtryk i lovgivningen, i aktørernes strategier, i aftaler og programmer. Politikker udpeger hvilke indsatser, der opleves relevante
	Indsats	Det samlede sæt af aktiviteter som skal løse den arbejdsmiljømæssige problemstilling man ønsker at gøre noget ved. Det kan være en paragraf af specifikke tidsmæssigt afgrænsede og sammenhængende aktiviteter som en eller flere arbejdsmiljøaktører iværksætter for at gennemføre sin politik på området. Det kan fx være en BAR-kampagne eller særlige tilsyn.
	Virkemidler	Tidsmæssigt afgrænsede og iværksatte aktiviteter til hvilke der knytter sig viden om, at de kan påvirke målgruppen til at ændre holdning, adfærd og prioritering eller øge vidensgrundlaget og handlekraften med henblik på at opnå et givet mål.
MODTAGERSIDEN	Mekanismer	Et samlet begreb for de forskellige aktiviteter (halvregulariteter) der gør det muligt for aktørerne på en arbejdsplads at omsætte en arbejdsmiljøindsats til konkrete ændringer af arbejdsmiljøet. Det kan bl.a. være valg, relationer og handlinger som forårsager bestemte udfald. Virkemidler aktiverer mekanismer, og et bestemt virkemiddel kan godt aktivere flere mekanismer i virksomheden. Mekanismebegrebet gør det muligt at sammenligne forløbet af arbejdsmiljøindsatser på tværs af virksomheder.
	Kontekst	De forhold og betingelser, som mekanismerne udspiller sig indenfor. Konteksten omfatter både de interne og eksterne betingelser som virksomheden påvirkes af.

Tabel 1: Grafisk fremstilling af de vigtigste begreber med relation til programteori

I Figur 6 viser vi hvordan et sådan sammenhængende system kan se ud. På afsendersiden er udgangspunktet den overordnede politik hos en arbejdsmiljøaktør (eller gruppe af aktører). Denne politik udmøntes i konkrete indsatser som er bygget op om en bestemt kombination af virkemidler rettet mod en målgruppe som kan spænde fra hele arbejdsmarkedet - som med AMO-reformen - til en afgrænset gruppe af virksomheder. Programteorien udgør afsendersidens forståelse af hvordan indsatsen og dens virkemidler vil fremkalde den ønskede effekt af indsatsen- programteorien ligger så at sige under både politik, indsats, virkemidler, mekanismer og kontekst. På modtagersiden udfolder der sig en praksis hvor indsatsen sætter en række mekanismer i gang der foregår i en bestemt kontekst, og som skaber effekten.



Figur 6: En model for politikens transformation fra afsender til modtager

Eksemplet nedenfor fra AMO-reformen kan anskueliggøre hvordan disse begreber kan anvendes. Reformen er udtryk for en omfattende indsats for at udvikle en mere effektiv organisering af virksomhedernes arbejdsmiljøarbejde, og et konkret eksempel fra et lille hjørne af denne indsats – ændringen af udtrykket sikkerhedsorganisation til arbejdsmiljøorganisation kan illustrere, hvordan de vigtigste begreber kan anvendes.

Programteorien bag ændringen fra sikkerheds- til arbejdsmiljøorganisation

Indsats: Udskiftning af begrebet sikkerhed med arbejdsmiljø i lovgrundlaget signalerer en bred tilgang hvor alle relevante problemer håndteres og arbejdsmiljøarbejdet bliver mere vedkommende.

Virkemiddel: Sikkerhed erstattes af arbejdsmiljø i lovgrundlag og knyttes sammen med formidlingsindsatser.

Målgruppen: Hele arbejdsmarkedet.

Mekanisme i virksomhederne: Medarbejdere og ledere bruger ordet arbejdsmiljø i daglig tale og gør dermed den bredere tilgang, som ligger i begrebet, til en integreret del af arbejdsmiljøforståelsen.

Kontekst: Den tidligere anvendelse af begrebet sikkerhed kunne indsnævre arbejdsmiljøforståelsen blandt medarbejdere og ledere i brancher hvor ulykker ikke er et stort problem. Omvendt kan ophøret af brugen af sikkerhed som styrende udtryk skabe uklarhed i brancher hvor ulykker er et vigtigt arbejdsmiljøproblem.

6.2 VIRKEMIDLER

Virkemidler er et nøglebegreb for arbejdsmiljøindsatsen, og det er derfor vigtigt at udfolde dette begreb nærmere. Det tilsvarende begreb i den internationale litteratur er udtrykket "policy instrument" som dog ikke er helt dækkende for begrebet virkemiddel i det danske sprog. I daglig tale anvendes begrebet virkemidler om mange forskellige typer aktiviteter, og det anvendes i flæng sammen med andre begreber som redskaber, værktøjer, kampagner, lovgivning, håndhævelse, politibetjentrolle og mange andre. Vi finder det som nævnt hensigtsmæssigt at skelne mellem indsatser og virkemidlerne. Som det fremgår af de eksempler der beskrives i det følgende, vil arbejdsmiljøindsatser ofte betjene sig af flere forskellige virkemidler, og indsatsernes resultater vil afhænge af hvordan modtagerne reagerer på virkemidlerne.

Der melder sig naturligt et ønske om at lave en oversigt over de eksisterende virkemidler. Da omfanget af indsatser og konkrete aktiviteter er stort, er det vanskeligt at lave en sådan udtømmende liste, selvom der er gjort mange forsøg. Vi holder os derfor til en mere overordnet systematisering af virkemidler, som man kan finde hos Evert Vedung (1998), som i sin typologi over 'policy instruments' skelner mellem kategorierne regulering (pisk), incitament (gulerod) og viden (prædiken) med de hyppigt anvendte metaforer i parentes. Disse begreber er et udtryk for den mekanisme som virkemidlerne forventes at påvirke modtagerne igennem.

Vedungs opdeling er altså en slags teori for virkemidler. Vedung når frem til opdelingen på baggrund af en gennemgang af en lang række forskellige forsøg på at lave typologier over indsatser og virkemidler. Han konstaterer at der ikke optræder en almen teori som baggrund for disse typologiseringer, men at de langt hen ad vejen bygger på en systematisering af empiriske data og observationer. Han finder, at der typisk optræder en opdeling af virkemidler efter de muligheder eller ressourcer afsenderen har til rådighed. Vedungs ambition er at udvikle et begrebsapparat som er komplet, og hvor kategorierne ikke lapper ind over hinanden.

Opdelingen i de tre generelle typer af virkemidler: regulering, incitament og viden er ganske anvendelige for overvejelser om, hvordan man bedst kan dosere og kombinere konkrete virkemidler eller indsatser, og den overordnede typologisering passer fint med en forståelse af, at spørgsmålet om, hvordan virkemidlerne virker, må være centralt, hvis man ønsker en eksplicit og refleksiv programteori.

Men selvfølgelig er Vedungs opstilling en idealtypificering som er hentet fra almen statskundskab og ikke som udgangspunkt er tænkt til arbejdsmiljøområdet.

Men opdelingen kan anvendes til at systematisere de aktiviteter, samfundet kan iværksætte for at fremme et godt arbejdsmiljø. En opdeling der også i nogen grad afspejles i de institutioner og strukturer, som er etableret, til at varetage arbejdsmiljøindsatsen som fx arbejdsmiljølov og tilsyn (regulering), BAR'erne (viden) og Forebyggelsesfonden (incitament). Opdelingen kan således også sætte fokus på hvorfor der er forskellige faglige vurderinger og politiske holdninger til hvordan arbejdsmiljøindsatsen skal prioriteres; noget der i den sidste ende vil udmønte sig i en bestemt – langt fra altid uanfægtet – fordeling af ressourcer til de tre typer af virkemidler.

6.3 TRE TYPER AF VIRKEMIDLER

Grundlæggende kan man sige at de tre typer af virkemidler peger på tre forskellige måder at motivere til handling på. Vedung lægger dog snittet mellem de tre typer et lidt andet sted end metaforerne kunne forlede en til at tro. Principielt beskriver tredelingen en forskel i forhold til med hvilken autoritativ kraft afsenderen påvirker modtageren.

Regulering er et spørgsmål om hvorvidt et givet forhold er erklæret tilladt (evt. under særlige betingelser) eller ulovligt, og om der i tilknytning hertil formuleres krav som målgruppen skal opfylde. I mange tilfælde er sådanne forhold jo understøttet af sanktioner, men langt fra altid.

Incitament er benævner Vedung også 'economic policy instruments' eller økonomiske virkemidler; som anvendes systematisk med henblik på at fremme den ønskede adfærd. Det vil typisk være gennem virkemidler af økonomisk karakter som fx støtte fra Forebyggelsesfonden eller favorisering af entreprenører med færre arbejdsulykker i forbindelse med en licitation, men kunne også være vejbumpe der motiverer lastbilerne til at køre væsentligt langsommere.

Viden som Vedung også benævner 'information', kan udnyttes gennem informationsspredning, målrettet rådgivning og vejledning; men indeholder ingen egentlig favorisering eller disfavorisering af nogle arbejdspladser i forhold til andre. Eksempler på denne kategori kunne fx være edutainment (som vi fx ser det i de reklamespots Videncenter for Arbejdsmiljø udsender på tv), og dokutræning (som fx tv-serien om arbejdsmiljø) eller anden form for formidling; men faktisk hører meget væsentlige dele af hele arbejdsmiljøsystemet fra Arbejdstilsynets fornyede vægt på dialog med arbejdspladserne - over den indsats som BST-systemet udgjorde tidligere - til en stor del af BAR'ernes aktiviteter og vejledninger under denne kategori.

Disse tre typer af virkemidler repræsenterer alle en asymmetrisk relation mellem afsender og modtager. Afsenderen har enten 1) magtbeføjelser til at fastlægge de legale normer 2) har mulighed for at omfordele ressourcer fra de 'uværdige' til de 'værdige', eller 3) iværksætte en informationsindsats, som går fra dem der har viden og indsigt, til modtagerne som ikke har denne viden.

Det er derfor interessant at diskutere, hvorvidt det særligt i forhold til det danske arbejdsmiljøarbejde og den danske arbejdsmiljølovgivning skaber mulighed for at tale om en fjerde type af virkemidler, der bygger på en mere symmetrisk relation. En sådan type virkemidler kunne fx være 'aftaler'. Den dækker netop over at flere parter enes om at gennemføre en indsats, fordi de alle mener at det vil gavne dem. Et eksempel herpå er de små mejerier som i et netværk indgik en aftale om at gennemføre en fælles indsats. Men også aftaler på brancheniveau hvor parterne i fællesskab støtter op om en konkret indsats, kan være udtryk for aftaler som en særlig type af et alment gældende virkemiddel.

6.4 SYNERGI MELLEM INDSATSER

Det særlige tilsyn er et eksempel på en indsats udformet og tilrettelagt af en arbejdsmiljøaktør (Arbejdstilsynet) som selv er i stand til at gennemføre en indsats med kombinationer af flere typer virkemidler. Ofte tilrettelægges indsatser dog således at flere aktører inddrages, og flere virkemidler og indsatser kombineres hvilket giver mulighed for at skabe synergi mellem indsatserne.

Et eksempel på dette er indsatsen fra BAR Byg i forhold til murerarbejde som beskrives i boksen nedenfor.

En indsats, der forandrer murerarbejdet

I en evaluering fra 2011 bliver det vist at en betydelig del af murerbranchen kender til og arbejder efter de nye principper om højst en meters løft ved opmuringarbejde. Dog er alle facetter af den nye arbejdsmetode ikke implementeret på alle byggepladser.

Indsats

Den samlede indsats består af en række elementer. Branchearbejdsmiljørådet for Bygge & Anlæg besluttede for over 10 år siden at sætte fokus på murerarbejdet. For mureren var der især erkendte muskelskeletbelastninger ved at mure under knæhøjde og over bryst/skulderhøjde, og for arbejdsmanden var der erkendte belastninger ved mange tunge løft, skub og vrid.

I et forarbejde blev det analyseret hvilke konsekvenser der ville være af at omlægge arbejdet sådan, at stilladset bliver flyttet hver gang der er muret en meter i højden frem for de gældende 1½ meter. Analysen viste at belastningen for murerne ville blive betydeligt reduceret ved en afskaffelse af opmuring under knæhøjde og over brysthøjde. Analysen førte frem til forslag om at mursten skal leveres på paller af ca. 25 centimeters højde og med færre mursten på hver palle, samt at der anvendes mørtelspande som ligeledes er løftet fra jorden. I forhold til arbejdsmanden er der fokus på at børen til transport af stenhoveder ændres.

Den primære indsats bestod i at BAR Byg udarbejdede en branchevejledning der beskriver principperne om udførelse af murerarbejdet. Som en del af indsatsen er der gennemført to forebyggelsesfondsprojekter som udvikler hjælpemidler, der gør det muligt at udføre arbejdet efter de nye principper. Derudover har branchens parter været aktive med at informere deres medlemmer, og Arbejdstilsynet anvender vejledningen som grundlag for tilsynsaktiviteter på området.

BAR systemet har en samfundsmæssig bestemt opgave som baserer sig på at oplyse ansatte og ledere om de til enhver tid mest fornuftige måder at udføre arbejdet på for at undgå/reducere nedslidning samt forebygge ulykker (viden). Det traditionelle virkemiddel i de konkrete indsatser er branchevejledninger som oplyser om principperne for en forsvarlig udførelse af arbejdet. Figuren nedenfor illustrerer en simpel og traditionel programteori for en branchevejledning.



Figur 7: Smpel programteori med implicitte antagelser

Som eksemplet ovenfor viser, har BAR Byg gennemført en væsentlig mere omfattende og mangefacetteret indsats hvor der netop har været satset på at opnå synergi mellem forskellige konkrete virkemidler (informationskampagne, forsøgsprojekter, tilsyn m.fl.) som hver især kan rubriceres under enten viden, incitament, regulering og aftaler, så alle typer af virkemidler kom i anvendelse.

Programteori – ”Opmuringsarbejde i højden”

BAR'ene er både formet af deres samfundsmæssige opgave og er et produkt af partssamarbejdet. Hvis ikke parterne slutter op om en vis grad af fornyelse og innovation, er det ikke muligt for BAR'en at formidle de mest fornuftige måder at udføre arbejdet på til den primære målgruppe. Derfor indgår virkemidler 'aftaler' som en mere eller mindre blind makker i BAR-samarbejdet. I det konkrete tilfælde har udarbejdelsen af branchevejledningen og de tilhørende udviklinger af tekniske hjælpemidler været en lang proces. Den endte med at parterne har givet hinanden håndslag på at stå bag den nye vejledning og aktivt arbejde for formidlingen af den. Det er primært sket ved at skrive om principperne i egne medlemsblade og på egne hjemmesider mv. samt ved at egne konsulenter har formidlet indholdet til ledere og medarbejdere. Virkemidlet er således fortsat oplysning, men oplysning som afsendes fra deltagerne i partssamarbejdet, og ikke nødvendigvis direkte fra BAR'en.

En del af motivationen for at udarbejde branchevejledningen er for parterne i branchen at man ved at ændre på opmuringarbejdet imødekommer Arbejdstilsynets krav om reduktion af belastninger. Pisker må derfor formodes at have været en driver for i det mindste arbejdsgiversiden i partssamarbejdet. Parallelt med at hjælpemidlerne blev færdigudviklede, og branchevejledningen blev publiceret og formidlet i branchen, har Arbejdstilsynet strammet sin tilsynspraksis op og fokuserer på om der mures efter de nye principper, og de tilsynsførende er begyndt at afgive reaktioner, hvis ikke det er tilfældet. Virkemidlet regulering (pisk) fungerer derfor i den konkrete case i synergi med virkemidlet oplysning (prædiken). Der kan således opstilles følgende programteori over denne indsats:

Målgruppe: Nedsættelse af muskelskeletbesvær for murere

Virkemidler:

1. Forundersøgelse (viden)

Mekanisme: at skabe enighed om problemets omfang og karakter

2. Forsøgsprojekter i konkrete virksomheder (incitament og viden)

Mekanisme: at vise at løsninger er mulige

3. Udarbejde og skabe enighed om en branchevejledning (aftale)

Mekanisme: at opbygge tillid til hinanden gennem respekt for alle parter legitime interesser

4. Formidling af branchevejledning (viden)

Mekanisme: at formidle kendskabet til vejledningen gennem mange kanaler

5. Tilsyn med efterfølgende reaktioner på manglende efterlevelse af vejledning (regulering)

Mekanisme 1: arbejdsgivere mister legitimitet gennem påbud

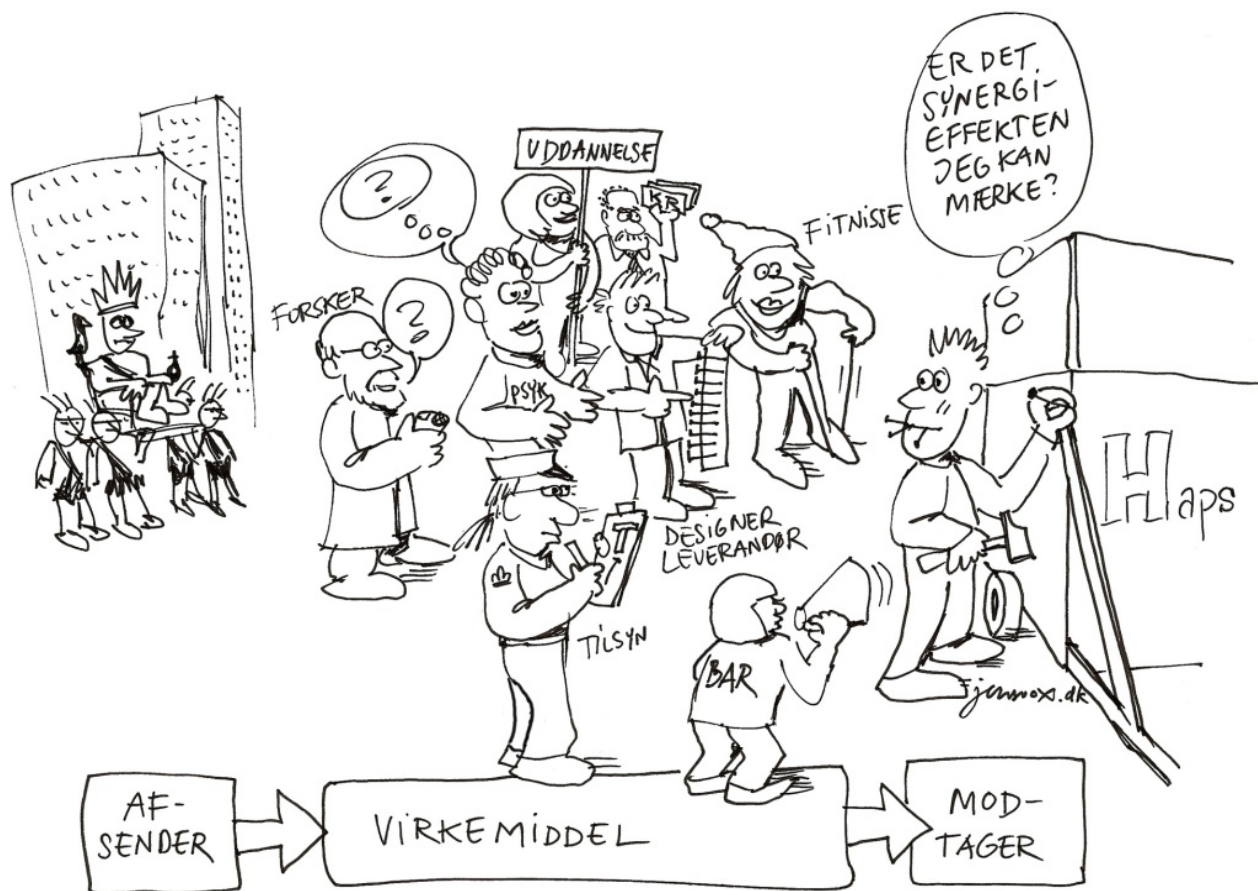
Mekanisme 2: myndighedens fokus understreger at vejledningens krav nu er almen praksis i branchen (denne mekanisme er også i spil i forhold til organisationernes formidling)

Kontekst: arbejdsgiverne og lønmodtagerne har tillid til egne organisationer og dermed fremmes effekten af mekanismen formidling

Murer-casen viser hvordan der er flere virkemidler og mekanismer i spil samtidig, og den understreger vigtigheden af at de spiller *sammen*. Hvis der fx ville være aktører som udsendte tvetydige signaler, er det let at forestille sig hvordan en ellers succesfuld indsats ville kuldsejle.

Synergimuligheden er alment anerkendt, og der er mange anbefalinger om at regulering, tilsyn og kontrol altid skal følges med bred information, oplysning og vidensdeling. Det fordrer i sig selv et samarbejde mellem fx myndighed og organisationer og det private marked for rådgivere, undervisere og vidensformidlere. De mest effektfulde indsatser bygger således på synergi mellem flere virkemidler der hver især har forskellige programteorier.

Det er dog ikke altid synergien opstår som led i en planlagt arbejdsmiljøindsats som kan ekspliciteres i en programteori. Der foregår en kompleks og på mange måder uforudsigelig udvikling i samfundet,



Figur 8: En illustration af synergieffekten mellem forskellige virkemidler

som ikke har sammenhæng med arbejdsmiljøområdet, men som får indflydelse på dannelse af holdninger og adfærd. Den sidste case i dette afsnit handler om alkohol på arbejdspladsen, og den viser hvorledes der kan ske store positive ændringer uden en klar arbejdsmiljøindsats og uden en entydig afsender og dennes programteori.

Alkohol

Gennem de sidste tre årtier har der været et voksende samfundsmæssigt fokus på alkohols negative betydning for sundheden. Talrige kampagner har haft til formål at nedbringe alkoholforbruget. Alligevel har det samlede alkoholforbrug været nogenlunde konstant (Kjøller, Juel, & Kamper-Jørgensen 2007). Der er

imidlertid i samme periode sket store ændringer i det individuelle drikkemønster. Således drikkes mere om aftenen og i weekenderne, hvorimod alkoholindtag på arbejdspladsen er faldet markant (Sundhedsstyrelsen 2008).

Sammenhængen mellem alkoholindtag og ulykkesrisiko er kendt og dokumenteret - primært i sammenhæng med trafik, men også på arbejdspladser. Nedbringelsen af alkoholforbruget, måske især indenfor byggeri og produktion hvor problemet var kendt, måtte således antages at kunne være en vigtig faktor for at begrænse antallet af arbejdsulykker. Indsatser der handler om alkohol på arbejdspladsen, fx opfordring til formulering af en alkoholpolitik, har imidlertid primært haft et individuelt sundhedssigte, fx ved at støtte medarbejdere med alkoholproblemer til at modtage afvænningsbehandling. Men de har ikke haft et egentligt arbejdsmiljøfokus (Sundhedsstyrelsen 2008). Der har således ikke været gennemført kampagner direkte rettet mod at nedbringe ulykker ved at reducere alkoholbruget på arbejdspladsen, men alligevel er det sket. Selvom langt de fleste virksomheder i dag har et forbud mod indtagelse af alkohol i arbejdstiden, skal årsagen til denne markante ændring formodentlig primært findes i en ændret samfundsmæssig holdning til at indtage alkohol og samtidigt kunne udføre sit arbejde. Ændrede samfundsmæssige normer og et stigende arbejdstempo, mere komplekse arbejdsopgaver og større effektivitetskrav ses således som den mekanisme der har skabt denne positive ændring.

6.5 FRA PROGRAMTEORI OG INDSATSER TIL PROJEKTER

Overvejelser om hvordan en indsats i praksis skal gennemføres indgår også i udviklingen af en programteori. Disse overvejelser omsættes normalt til et projektdesign hvor programteorien skal omsættes til konkrete aktiviteter med dertil knyttede ressourcer og tidsplaner. Der findes mange modeller for beskrivelse af projekter, og det gælder for lang de fleste at de fremstiller et projektforsløb som en række centrale og fortløbende aktiviteter som fx:

- Forberedelse: Udvikling af projektets resultatmål (ift. politik) og fastlæggelse af målgruppen
- Kontekstanalyse og program: Udvikling af en eksplicit programteori i forhold til indsatsen, udvælgelse af virkemidler og undersøgelse af sandsynlige mekanismer
- Design: Udvikling af plan, struktur og tidsforløb for implementering af virkemidler og gennemførelse af forløbet
- Gennemførelse af aktiviteter og opfølgning
- Forankring og evaluering

Der findes mange forskellige værktøjer til planlægning og vurdering af projekter. Et hyppigt anvendt og let tilgængeligt værktøj er Logical Framework Approach (LFA) som oprindeligt er udviklet til anvendelse indenfor udviklingsbistand (NORAD 1999). Fordelen ved LFA i forhold til mange andre projektværktøjer er dets fokus på konstant at vurdere om de grundantagelser som projektet hviler på, også holder i praksis. Det fremhæves at et projekt (i vores tilfælde en arbejdsmiljøindsats) skal vurderes i forhold til om de konkrete aktiviteter kan føre til de forventede outputs (se afsnit 7.3) i forhold til projektets formulerede mål. Væsentlige elementer er at skabe sammenhæng mellem mål, aktiviteter til at opnå målene, ressourcer til at gennemføre aktiviteterne, verificerbare indikatorer på at aktiviteter er gennemført og outputs leveret. Disse sammenhænge beskrives i programteorien og der er et særligt fokus på hvordan konteksten kan påvirke (positivt eller negativt) forløbet af projektet. I denne sammenhæng optræder begrebet "dræberantagelser" (killer assumptions). Dræberantagelser er de cen-

trale antagelser et projekt bygger på, og som hvis de ikke holder stik, betyder at projektet mislykkes eller kun i meget begrænset omfang når sit mål.

Der kan således hentes vigtig inspiration fra projektteori som kan styrke arbejdet med at udforme en holdbar programteori og efterfølgende omsætte den til konkrete aktiviteter der kan evalueres. Det er dog vigtigt at være opmærksom på at arbejdsmiljøindsatser forholder sig til komplekse forhold i virksomhederne, og at arbejdsmiljøet har en vis karakter af vilde problemer (se kap. 8). Arbejdsmiljøindsatserne og specielt resultaterne i virksomhederne følger derfor sjældent en retlinjet projektlogik. Ikke desto mindre er det udfordringen at udvikle så detaljerede forestillinger om hvordan en indsats kommer til at forløbe, som det nu er muligt med den tilgængelige viden og erfaring. I den sammenhæng kan LFA's bud på hvordan man kan systematisere sine antagelser, være meget inspirerende.

7. ARBEJDSMILJØINDSATSER FRA ET MODTAGERPERSPEKTIV

7.1 MEKANISMER

Mekanismer er et bredt begreb der kan anvendes til at beskrive forhold hvor noget påvirker noget andet. I vores sammenhæng har vi valgt at definere mekanismer som de forskellige forhold (handlinger, aktiviteter, relationer) der bringer aktørerne på en arbejdsplads til at omsætte en arbejdsmiljøindsats til konkrete ændringer. De kan bl.a. beskrives som valg, relationer og handlinger som forårsager bestemte udfald. Det vigtigste formål med mekanismebegrebet er at skabe mulighed for at sammenligne forløbet af arbejdsmiljøindsatser på tværs af virksomheder og kontekster. Mekanismebegrebet anvendes derfor til at undersøge om indsatser fører til iværksættelse af ændringer på bestemte måder som optræder i flere steder og dermed kan generaliseres og anvendes i nye indsatser.

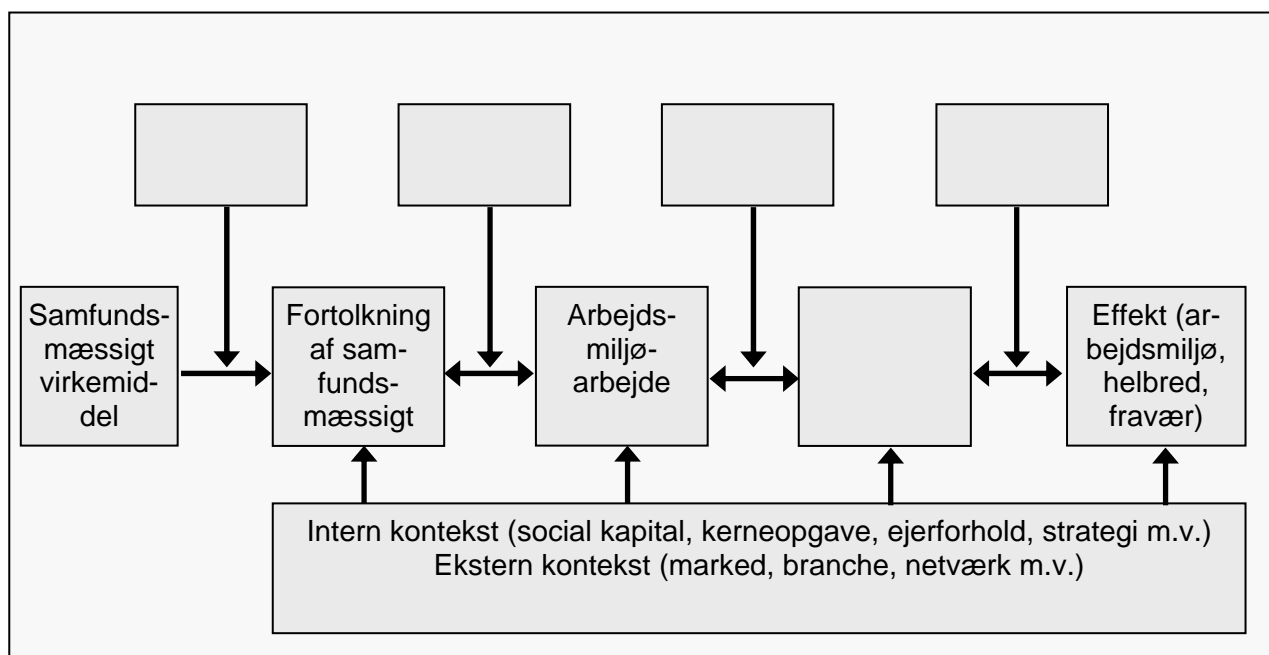
Går vi tilbage og overlader ordet til Pawson (2006, s. 23) som er udgangspunktet for vores aftapning af begrebet mekanisme så skriver han (i vores oversættelse og med vores fremhævning):

”Mekanismer er motorerne i forklaringerne ... Vi kan i grove træk forstå verden i kraft af dens halvregulariteter [altså det at den opfører sig på en genkendelig måde til trods for at der er masser af undtagelser]. Rytmerne i og koblingerne mellem naturlige og endda sociale systemer er konstante nok til at vi kan navigere i dem, skønt vi aldrig, som netop anført, er særligt overraskede når tingene ikke virker som forventet. Vi er afhængige af mekanismer til at fortælle os hvorfor sammenhænge skulle optræde. En række af begivenheder eller et mønster af adfærd forklares med at de er en del af et system, og mekanismen fortæller os *hvad det er ved systemet som genererer mønsteret*. Mekanismer forklarer årsags-virkningsrelationer ved at beskrive de i systemet iboende 'kræfter', hvad enten der er tale om systemets bestanddele (fx gasser og krudt⁶) eller aktører (fx eksaminatorer eller policy-makers) eller strukturer (som bureaukratier eller social programmer). I alle tilfælde er der noget ved den 'tilbøjelighed' ved systemet der forklarer den kausale regularitet. Mekanismen forklarer hvad det er ved systemet der får ting til at ske.”

Mekanismer er altså troværdige forklaringer på hvad der får ting til at ske på en rimeligt ensartet måde i flere end et tilfælde. Dog er det vigtigt at holde sig for øje, at begrebet 'kræfter' som Pawson anvender ikke betegner en mekanisk eller maskinel forandringskraft. Kræfter er betingede af aktørers handlinger og den givne kontekst. Man kan altså se de samme elementer gøre sig gældende i mange forskellige forandringsprocesser uden at kunne redegøre for den fuldstændige årsag til at det sker, og selvom resultatet ofte er nogenlunde det samme, er det ikke altid tilfældet. Der er netop tale om såkaldte 'halvregulariteter' og ikke naturlove hvor udfaldet altid er det samme.

En arbejdsmiljøindsats hvis den gennemføres og når sit mål, har været båret frem af et sæt af mekanismer. Til udbredelsen af en arbejdsmiljøindsats knytter der sig en række elementer som er illustreret i den følgende figur: Fra en samfundsmæssig prioritering til et bedre helbred. Hvert trin i en sådan udbredelse skabes af en række mekanismer, nogle kobler sig til de anvendte virkemidler, andre til de enkelte virksomheders evne og vilje til at omsætte indsatsen til interne arbejdsmiljøaktiviteter eller medarbejdernes inddragelse i den daglige implementering og andre igen kobler sig til konteksten.

⁶ Pawson bruger sammenhængen mellem krudt (svovl, salpeter og kulstof) og den reaktion der kan komme, hvis det antændes, til at illustrere en mekanisme.



Figur 9: **Sammenhængen mellem indsats, mekanismer og kontekst**

Mekanismebegrebet har i nogle tilfælde udløst den vel ikke helt urimelige indvending at det lægger op til en forståelse af de sociale systemer som samfundsmæssige indsatser udspilles i som en slags maskine. Selvom denne association ligger lige for, vil vi dog fremhæve at vores ambition med begrebet er at pege på alle de forhold der skaber regulariteter i et system. Det vil sige som det særlige der ofte får noget til at forløbe på samme måde i flere tilfælde. Det kan være både aktørdrevne og systemiske forhold der knytter sig til de sociale systemer som også arbejdsmiljøindsatser er en del af.

7.2 KONTEKST

Når arbejdsmiljøindsatsen sættes i søen, bliver den mødt af den kontekst der er til stede i den konkrete branche og enkelte virksomhed. Kontekst betegner alle de forhold som påvirker virksomhedens handlerum. Det kan både være eksterne faktorer som eksempelvis markeders udvikling, konkurrenter eller branchens generelle udvikling, og det kan være en intern kontekst der omhandler virksomhedens traditioner, ledelsesformer, sociale normer eller særlige fysiske rammer. Endelig kan der også tales om en politisk kontekst hvor man må tage højde for politiske beslutninger på området, arbejdsmiljøsystemet og arbejdsmarkedets parter. Dette møde med konteksten har afgørende betydning for i hvilken grad arbejdsmiljøindsatsen omsættes til konkrete ændringer der lever op til intentionerne med indsatsen. Mødet med modtagerens kontekst kan man anskue som den måde virkemidlet udfoldes på i praksis. Mekanisme og kontekst er altså et gensidigt begrebspar hvor den ene side ikke giver mening uden den anden.

For at kunne evaluere både årsager og effekt af en given arbejdsmiljøindsats, må man derfor få indblik i hvordan den eksisterende kontekst spiller sammen med indsatsen. Der er således behov for en kortlægning af praksis og kontekst hos modtageren af en arbejdsmiljøindsats. For at give et billede af hvad der er tale om i praksis, vender vi igen tilbage til casen fra mejerierne.

Mekanismer og kontekst - Ergonomi på mindre mejerier

I mejericasen er der flere integrerede virkemidler - herunder økonomiske incitament og netværksstyring som de mest fremtrædende. På et overordnet niveau er branchen presset af Arbejdstilsynet og truslen om påbud. Pisk som virkemiddel har dermed virket som motivationsfaktor både i forhold til de virksomheder der har fået afgørelser, men også indirekte på branchen som helhed gennem truslen om mulige påbud. Projekter udført i samarbejde med branchens store concern har skabt fokus på problemer og løsninger knyttet til muskelskeletbesvær i branchen og formidlet praktiske erfaringer. Opfordringen til at øge indsatsen er desuden fulgt op fra BAR Jord til Bord. Endelig åbner forebyggelsesfonden mulighed for økonomisk støtte til denne type af forebyggende indsatser. Branchen og temaet nedslidning forårsaget af tunge løft er således mødt med virkemidler af typerne regulering, viden og incitament, og disse virkemidler har dannet afsæt for den konkrete indsats som ergonomiprojektet udgør.

Selve projektets programteori tog udgangspunkt i muligheden for at opnå økonomisk støtte. Et andet afgørende virkemiddel bag programteorien var, at arbejdsmiljøindsatsen blev udført i regi af netværket af mindre virksomheder som gav hinanden "håndslag" på at arbejde sammen om den konkrete udvikling af redskaber og samtidig om at åbne dørene for hinanden. En grundlæggende motivationsfaktor var projektets målsætning om at løsningerne samtidigt med at kunne reducere belastningerne, også skulle leve op til at være praktisk anvendelige og økonomisk rentable.

Som projektet forløb, viste det sig at den indledende delvist eksplicite programteori i store træk var korrekt, idet det lykkedes med de gældende virkemidler at skabe en positiv udvikling såvel i netværket som i udviklingen af konkrete hjælpemidler. I denne case er det vigtigt at undersøge hvad der binder netværket sammen, og hvad der ligger til grund for, at konkurrenter gav hinanden "håndslag" på at arbejde med åbne døre. Det er således mekanismer, der knytter sig til anvendelsen af netværk som virkemiddel. Ligeledes er udviklingen fra prototype til udbredt anvendelse en proces der er afgørende for at skabe det ønskede slutresultat. Men det er samtidig en proces der for at kunne begribes, kræver kendskab til den kulturelle, organisatoriske og sociale kontekst der kendetegner de små mejerier.

En bred udbredelse af hjælpemidlet til så mange mejerier som muligt følger dog ikke umiddelbart af at selve projektet er succesfuldt gennemført. Nogle af de problemer der kan opstå, blev illustreret i den indledende historie.

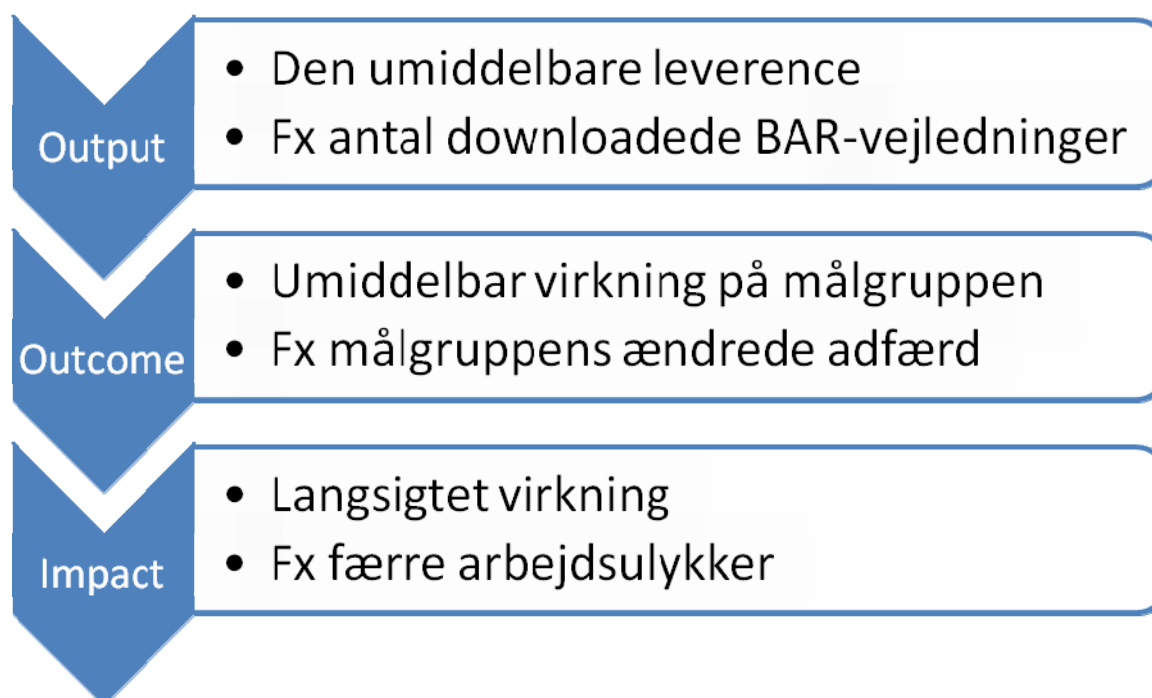
Der kan ikke indenfor en realistisk tids- og ressourceramme etableres en kausal sammenhæng mellem den direkte indsats og det langsigtede mål om at reducere nedslidning i den samlede branche. Derfor kan et sådant effektmål ikke anvendes som det eneste succeskriterium for indsatsen. I denne case er der også opstillet specifikke og lettere dokumenterbare delmål som fx biomekaniske mål for belastningsgraden når det enkelte udviklede redskab testes, samt omfanget af i hvor høj grad det udviklede redskab bruges i praksis, og hvor mange mejerier ud over testmejeriet der anskaffer redskabet. En simpel dokumentation af sådanne måltal giver dog ikke viden om, hvorfor og hvordan det lykkes at nå til et positivt resultat. For at kunne opnå en sådan viden, bør en evaluering inddrage analyse af de mekanismer der udfolder sig på virksomhedsniveau. Det kan fx være motivation til adfærdændring, implementeringsstrategiens omfang af forpligtelse og kontrol og anvendelsesgraden i forhold til den samlede produktionsopgave.

Anvendelsen af begreberne mekanisme og kontekst har til formål at gøre arbejdsmiljøaktører mere opmærksomme på hvordan indsatser konkret kan formodes at virke i praksis. Det næste spørgsmål bliver hvordan man så kan forstå effekter af indsatser. Hvornår og hvordan kan man udtale sig om, at noget har virket – har haft effekt – i komplekse arbejdsmiljøssammenhænge?

7.3 EFFEKT

I en evalueringssammenhæng er det almindeligt accepteret at det ofte er meget vanskeligt at sige noget om hvorvidt slutmålet af en given indsats fx i form af mindre nedslidning, færre arbejdsbetingede sygdomme og færre ulykker, er opnået. Arbejdsmiljøindsatser har typisk et langsigtet gennemslag og er påvirket af så mange andre faktorer at det i realiteten er meget vanskeligt at fastlægge effekten. Ofte vælges derfor andre evalueringsmål. I bedste fald kan der måles på arbejdsmiljøet. Medarbejdere kan spørges om det psykosociale arbejdsmiljø, eksponeringer kan måles, og skærme på rundsave kan registreres. Det kræver selvfølgelig at der er velbegrundede teorier om at de forhold i arbejdsmiljøet man måler, faktisk har betydning for sikkerhed og sundhed. Ofte er det imidlertid også vanskeligt at måle på arbejdsmiljøet; både fordi fx et BAR-projekt ikke kommer så tæt på den enkelte virksomhed, og fordi det ville blive for dyrt at måle.

Blandt andet fordi den egentlig effekt er svær at finde mål for, skelner man ofte mellem output, outcome og impact. *Output* er den umiddelbare 'leverance' af indsatsens kerneaktiviteter, *outcome* er de direkte virkninger indsatsen har overfor målgruppen, mens *impact* er den mere langsigtede virkning fx i form af holdbare forandringer. Lad os tage en BAR-vejledning som eksempel. Hvorvidt den er blevet spredt eller downloadet, eller om den er blevet læst, kan siges at være output. Outcome vil derimod være om folk kan huske hvad de har læst, om de har fået noget nyt at vide; og videre om de har ændret adfærd som følge af vejledningen. Impact dækker i denne sammenhæng over, hvorvidt BAR-vejledningen grundlæggende har forbedret arbejdsmiljøet og dermed de ansattes sikkerhed og sundhed på længere sigt.

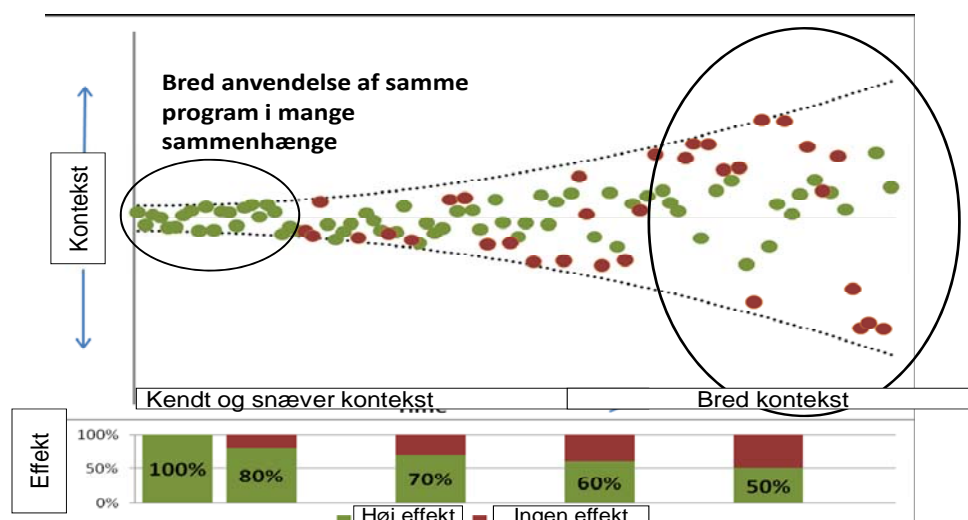


Figur 10: Definitioner på output, outcome og impact af indsatser

Af pragmatiske årsager bliver det typisk valgt at evaluere på output på aktivitetssiden. Det kan fx i forbindelse med informationsindsatser være det umiddelbare antal af downloads fra en hjemmeside eller antallet af distribuerede pjecer, men det kan også være output i næste led fx i form af andelen af en gruppe, som har kendskab til den information, som distribueres gennem hjemmesiden og pjecerne.

Disse output må imidlertid ikke forveksles med det ønskede resultat for indsatsen. Der er ikke nødvendigvis sammenhæng mellem kendskab til bestemt viden, og handlinger som forbedrer arbejdsmiljøet. En god programteori bruger viden fra andre compatible sammenhænge til at klargøre hvilke områder man indsamler data om, og hvordan disse områder gør det muligt at sandsynliggøre at projektet har effekt. Om det kan gøres meningsfyldt, afhænger af den sekundære videns generaliserbarhed – altså i hvilken udstrækning den kan generaliseres fra den kontekst hvor den er generet, til den anden kontekst hvor den anvendes i en programteori fx for en bestemt informationsindsats. Dokumentation for at en bestemt indsats kan forventes at have den ønskede effekt, er i mange sammenhænge et krav der stilles inden indsatsen bringes i bredere anvendelse fx når der udvikles nye medicinske behandlingsformer af sygdomme. Et sådant krav stilles sjældent til arbejdsmiljøindsatser. Erfaringer opsamles oftest efter en besluttet indsats er afprøvet. I forbindelse med bevillinger til projekter og indsatser som fx i Forebyggelsesfonden foreslås det ofte at der først gennemføres et pilotprojekt. Målet er at kunne justere den store og bredere indsats. En veludviklet programteori og efterfølgende evaluering kunne med fordel anvendes ved sådanne pilotprojekter for på den måde også at teste programteorien og justere de forventede resultater. Herved kan pilotprojekter anvendes til mere end blot at justere indsatsen på de indre linjer, men netop også til at justere indsatsen i forhold til et større kendskab til den kontekst man forventer at udbrede indsatsen i - herunder fx målgruppens variation, forskellige behov og reaktion på forskellige virkemidler.

I et pilotprojekt vil der ofte være flere ressourcer til de enkelte aktiviteter og en større bevågenhed på målgruppen som derved kan forventes at give et større gennemslag. Man kan derved opnå forskellige og måske meget bedre resultater fra pilotprojektet i forhold til den samlede indsats. Dette forhold er illustreret i den følgende figur:



Figur 11: Sammenhæng mellem evaluering af en indsats i en kendt og begrænset kontekst og i en bred kontekst (Inspireret af Parry 2010).

Figuren illustrerer hvordan evaluering af et forholdsvis begrænset antal cases som fx kunne være et pilotprojekt, kan bidrage med dokumentation for at indsatsen vil have en virkning. Når den udbredes til en større sammenhæng, vil der være både effektive og ikke effektive cases, og vurderingen af indsatsens virkning kan som i figuren falde til en effektivitet på 50 %. Gennemsnittet vil derfor pege på en ringe effekt, men det vigtige er imidlertid at der er 50 % af målgruppen som rent faktisk reagerer på

samme positive måde som målgruppen i pilotprojektet. Derved er der faktisk en positiv effekt, og det væsentlige bliver at finde frem til de mekanismer som gør at indsatsen virker for nogle virksomheder og ikke for andre. Et eksempel er anvendelsen af sociale medier som formidlingskanal overfor ejere af små virksomheder. Anvendes de overfor mediebranchen vil mange modtage og forholde sig til formidlingen, anvendes de derimod indenfor byggesektoren vil den del af målgruppen der nås være præget af unge mestre, hvorimod mange af de ældre næppe vil opdage en sådan kampagne. Evaluering af et pilotprojekt fortæller os således noget om, hvordan en indsats virker eller kan virke, hvorimod evalueringen af den samlede indsats fortæller os noget om, overfor hvem og i hvilken kontekst denne virkning kan forventes at indtræffe.

I konkrete arbejdsmiljøindsatser vil der typisk indgå forskellige blandingsforhold (konfigurationer) af Vedungs tre typer af virkemidler. Blandingsforholdet er et udtryk for en given sammensætning af konkrete virkemidler som fx tilsyn og informationsformidling. Det sker typisk med udgangspunkt i en forforståelse af den målgruppe som indsatsen er rettet mod. Nedenfor viser vi et eksempel på hvorledes forskellige virkemidler kan kombineres, og hvilken betydning det har for forventningerne om resultater, når målgruppens kontekst tænkes med ind i indsatsen. Arbejdstilsynets implementering af det særlige tilsyn er anvendt som eksempel på hvordan en indsats kan tilrettelægges, således at den kombinerer forskellige virkemidler.

Arbejdstilsynets særlige tilsyn - analyse af programteori

Arbejdstilsynet har indenfor de senere år udviklet en ny tilsynsstrategi der omtales som "det særlige tilsyn". Denne tilsynsform indebærer at virksomhederne efter et besøg af Arbejdstilsynet får en tidsfrist til at forbedre påpegede problemer i stedet for et egentligt påbud, og de får samtidig vejledning i hvad de skal gøre for at løse problemet. Programteorien bag denne indsats er i kort form at virksomhederne gennem muligheden for selv at løse problemet vil være mere motiverede til at forbedre arbejdsmiljøet, end hvis de får et direkte påbud. Denne programteori udspringer af Arbejdstilsynets egne mangeårige erfaringer med tilsynsbesøg.

Programteoriens forståelse af hvordan indsatsen vil virke, vil imidlertid kun kunne opfyldes, hvis nogle grundlæggende forudsætninger i konteksten er opfyldt. Arbejdstilsynet skal have politisk opbakning til at arbejde med denne kombination af kontrol og vejledning (pisk og prædiken), og de udførende tilsynsførende skal have kompetencer der gør dem i stand til at udfylde rollen. På de interne linjer skal virksomheden have vilje og evne til at efterleve den vejledning, de modtager fra den tilsynsførende. Det kan man imidlertid ikke forudsætte at alle virksomheder har. Programteorien kan således kun forventes at blive opfyldt i en afgrænset del af målgruppen.

Det vil derfor være væsentligt i en evaluering af en sådan tilsynsindsats at opnå et kendskab til hvordan indsatsen og dens virkemidler påvirker forskellige segmenter i målgruppen. At opnå et sådant kendskab forudsætter kendskab til de tilsynsførendes praksis og til hvordan forskellige typer af aktører i virksomhederne reagerer på denne tilsynsform.

Dette eksempel viser at også den enkelte praktikers implementering af en indsats får væsentlig betydning for hvordan den fungerer i praksis. I dette tilfælde gennem de tilsynsførendes villighed og kompetencer til at påtage sig en ny rolle som adskiller sig fra det traditionelle tilsyn. Denne problemstilling er beskrevet i den internationale forskning som betydningen af 'street level bureaucrats'.

Implementeringen afgøres af praktikerne

Det har i mange år stået klart at praktikerne har en afgørende indflydelse på implementering. Inden for myndighedsområdet kaldes praktikerne *'street level bureaucrats'* (SLB) (Lipsky 1971), og det har vist sig at uanset hvor stramt man fra toppen forsøger at styre sine *'street level bureaucrats'*, vil de altid forsøge at etablere en praksis som de opfatter som meningsfuld. Det er vigtigt for *'street level bureaucrat'*ernes eget arbejdsmiljø, men det har også en god begrundelse i forhold til at få størst mulig effekt ud af udøvelsen af myndighed. På arbejdsmiljøområdet har Black og Baldwin (2010) således gennemført en undersøgelse af tilsyn med arbejdspladser. De viser hvordan den tilsynsførende bliver en afgørende formidler af den programteori, som myndigheden (Arbejdstilsynet) har fastlagt på baggrund af lovgivningen. Den tilsynsførende er en aktør der uanset tilsynsreglernes klarhed, udøver en række skøn gennem sin praksis. Skøn der baseres på den tilsynsførendes faglige viden, erfaring, egne performance mål og etableringen af den konkrete relation mellem virksomhed og tilsynsførende. Denne analyse er et godt eksempel på hvor komplekst det er at beskrive og forstå en indsats som arbejdstilsyn i praksis. Pointen er at afvejningen af om man bør foretrække "pisk", vælge en "gulerodsstrategi" eller fokusere på at "prædike" ikke kan afgøres på forhånd. Det er i mange tilfælde en afgørelse der udvikles gennem den konkrete interaktion og relation mellem tilsynsførende og virksomhed. Myndigheden vil ifølge Black og Baldwin derfor ofte være i en konstant dialog med sig selv og med omverdenen om vægtningen mellem en kontrol-straf-strategi og en 'hjælp til selvhjælp'-strategi. Dialog bør derfor heller ikke tolkes som uklarhed, men som en nødvendig løbende undersøgelse og justering af de valgte virkemidler og af forståelsen mellem tilsynsførende og virksomhed. En justering som er nødvendig for i den konkrete situation at opnå det bedst mulige resultat.

Vi har i de to foregående kapitler udviklet et begrebsapparat til forståelse af arbejdsmiljøindsatser. Når vi arbejder med arbejdsmiljøindsatser, har vi allerede fremhævet at en række arbejdsmiljøproblemer kan opfattes som vilde problemer. Det betyder bl.a. at der sjældent kan etableres en udtømmende forståelse af hvordan en arbejdsmiljøindsats har påvirket de virksomheder som indgår i målgruppen. Endsige opstilles sikre forestillinger om hvordan en kommende indsats vil virke. Ikke desto mindre vil vi hævde at der gennem det begrebsapparat vi har præsenteret, er skabt en mulighed for at blive klogere på hvordan og i hvilke sammenhænge de mange virkemidler kan virke. Ligesom vi gennem mekanismebegrebet har fået en ny synsvinkel til at anskue og beskrive de mange forskellige måder som virksomheder opfører sig på, når de udsættes for arbejdsmiljøindsatser. Hensigten er at blive klogere på de forhold der kan genkendes på tværs af indsatser og på tværs af virksomheder. Det kan lade sig gøre på baggrund af en generel forståelse af arbejdsmiljøområdets mange facetter, og med et begrebssæt der kan håndtere denne kompleksitet. Herved kan der opnås en bedre forståelse af, hvad der virker og hvorfor.

I det næste kapitel vil vi se på hvordan vi med dette udgangspunkt kan prioritere en evalueringsindsats ud fra de til rådighed stående evalueringsteorier.

8. FRA EVALUERING TIL HANDLING

Evaluering kan som nævnt defineres som en systematisk bedømmelse af udfald, præstationer og organisering af samfundsmæssige aktiviteter. Som tidligere nævnt er denne bedømmelse rettet mod fremtidige handlingssituationer og evalueringer har derfor et anvendelsessigte (Dahler-Larsen 2010; Vedung 1998). Det grundlæggende formål med evaluering er således at opnå viden som kan anvendes til at træffe beslutninger om fremtidige handlinger. Det betyder at evalueringer har et klart læringsmål; vi skal blive klogere således at vi kan træffe bedre beslutninger. I praksis knytter der sig dog et sideformål hertil; legitimering af indsatsen. Såvel bevillingsgiver som den udførende aktør har behov for at legitimere at ressourcerne til indsatsen er anvendt forsvarligt og med et positivt udbytte. De to formål kan i nogen grad være sammenfaldende når den opnåede viden både kan anvendes til at kvalificere den fremtidige beslutning og samtidig vise om pengene er givet godt ud. Men afhængigt af resultatet kan forskellige aktører have interesser, som ikke nødvendigvis understøtter læringsformålet. Det er derfor nødvendigt at overveje hvorledes disse to formål med evaluering hænger sammen, og det gør vi i slutningen af dette kapitel.

Grundlaget for overhovedet at anvende evalueringresultater til noget som helst er tillid til at evalueringen frembringer viden som vi kan stole på. Der rejser sig derfor et spørgsmål om, hvordan viden skal forstås i evalueringssammenhæng, og hvornår vi kan stole på den. Dette spørgsmål har desværre ikke et enkelt svar fordi arbejdsmiljøindsatsen har bevæget sig fra at fokusere på konkrete afgrænsede arbejdsmiljøproblemer med enkle løsninger til komplekse problemer som fx stress og sygefravær som også har komplekse løsninger. I evalueringsteorien er det netop beskrevet som henholdsvis 'tamme' og 'vilde' problemer. De tamme problemer kan løses af en entydig intervention med en sikker effekt. Det gælder eksemplet med selve afskærmningen af rundsaven. De vilde problemer er derimod komplekse, og de kan vanskeligt beskrives og afgrænses. Effekten bliver dermed stærkt kontekstafhængig og mere eller mindre uforudsigelig. I bedste fald kan der derfor tales om en sandsynlig effekt af en arbejdsmiljøindsats. Selvom det er muligt at fremskaffe mere og mere viden om de vilde problemer og om mulige indsatser som kan bidrage til at øge sandsynligheden for en positiv effekt, vil de typisk ikke kunne reduceres til tamme problemer, hvor effekten bliver sikker og forudsigelig. De stadige forandringer af konteksten vil betyde at udfaldet vil blive mere eller mindre forskelligt hver gang.

Det betyder at den traditionelle effektevaluering har svært ved at forklare de resultater som den frembringer. Det betyder imidlertid ikke at den skal forkastes, selvom arbejdsmiljøet i dag i høj grad er domineret af vilde problemer. Der er stadig behov for at skaffe viden om kvantificerbare ændringer. Hvis der ikke er tegn på positive ændringer, vil der næppe være belæg for at fortsætte med en bestemt indsats, selvom denne indsats også er stærkt kontekstafhængig, og andre faktorer end indsatsen har haft indflydelse. Det er blot vigtigt at være opmærksom på at en positiv effekt godt kan optræde, selvom et niveau er uændret eller negativ. Det kan fx være, at ændringer i konteksten medfører væsentlige forværringer, som indsatsen faktisk forebygger en betydelige del af – uden dog at kunne opnå en positiv udvikling. Netop dette var tilfældet i et projekt der undersøgte sygefravær i københavnske plejehjem. Her steg sygefraværet betydeligt i løbet af projektperioden på alle plejehjem, men i interventionsplejehjemmene var sygefraværet steget signifikant mindre (Kristensen & Smith-Hansen 2003). Der skal desuden tages hensyn til at effekt på arbejdsmiljø og helbred typisk optræder som resultat af mange forskellige indsatser, som spiller sammen. Det vil således sjældent være muligt at finde direkte effekter af enkeltstående indsatser, og en del af målet med evalueringen må derfor være at analysere hvordan forskellige forhold har indflydelse på de forandringer som man søger at opnå med indsatsen.

Erkendelsen af at de fleste arbejdsmiljøindsatser optræder som vilde problemer hvor der ikke kan forudsiges en sikker effekt, betyder naturligvis ikke at man skal undlade at handle. Vi ved der findes væsentlige arbejdsmiljøproblemer som ganske vist er vanskelige at løse, men som har alvorlige konsekvenser for både individer, virksomheder og samfund. Det er under alle omstændigheder nødvendigt at forsøge at forbedre arbejdsmiljøet, og det afgørende bliver derfor at afklare hvordan vi får den bedst mulige viden som grundlag for at træffe beslutning om fremtidige handlinger. I den sammenhæng er evidensbegrebet centralt.

8.1 EVIDENS

Evidens som begreb må forstås som den bedst tilgængelige viden om hvordan indsatser skal designes for at opnå den størst mulige sandsynlighed for en høj grad af effekt. Den tilgængelige viden skal anvendes her og nu, samtidig med at der arbejdes så systematisk som muligt med at skaffe yderligere viden, og heri spiller fortsatte evalueringer en væsentlig rolle i forøgelsen af vores vidensgrundlag og forståelse af mekanismer og dermed udvikling af reflekterede programteorier.

Evidens er blevet et modebegreb, og dermed også et begreb som bruges og misbruges til mange forskellige formål. Det er i stigende grad et krav at alle beslutninger om nye aktiviteter i den offentlige sektor skal baseres på evidens. Formålet er i så høj grad som muligt at sikre at aktiviteterne fører til de ønskede resultater. Nye forslag kan derfor skydes ned med henvisning til at der ikke foreligger tilstrækkelig evidens, eller diskussionen kan drejes til at være en teknisk diskussion af gyldigheden af den foreliggende evidens. Håndteringen af dette spørgsmål vender vi tilbage til lidt senere. Indledningsvis undersøger vi her evidensbegrebet lidt nærmere.

Ifølge Nudansk Ordbog⁷ betyder evidens: 'En oplysning som støtter en antagelse' eller med Oxford Advanced Learner's Dictionary's⁸ formulering (oversat til dansk af os): 'Information som giver en stærk begrundelse for at tro noget eller bevise noget'. Ifølge denne betydning omfatter evidens altså enhver oplysning som understøtter en antagelse. Evidens behøver således ikke at være resultatet af forskning eller videnskab. Men evidens – oplysninger – giver også anledning til at stille to centrale spørgsmål: Kan man overhovedet stole på oplysningen? Og er oplysningen nu relevant for den pågældende antagelse? Når man anvender begrebet evidens, indebærer det derfor altid en diskussion af gyldighed og relevans.

Et eksempel på relevansen kan være asbestsagen i 1980'erne. Der blev i en panikagtig løsning anvendt flere milliarder kroner på at fjerne loftsplader i dagsinstitutioner og skoler, men den sundhedsmæssige betydning heraf kan i høj grad diskuteres. Spørgsmålet er om der til trods for omhyggelige sikkerhedsforanstaltninger ikke alligevel ville være flere fibre i luften efter en renovering, og under alle omstændigheder ville den sundhedsmæssige effekt af renovering sammenlignet med forseglede loftsplader formodentlig være marginal. En væsentlig årsag til at bruge de mange penge var en udtalelse af en læge i fjernsynet: Én asbestfiber er tilstrækkeligt til at fremkalde kræft. Mens denne oplysning ud fra en biologisk betragtning må antages at være korrekt, er den næppe relevant for vurderingen af risikoen ved henholdsvis at fjerne loftspladerne eller ved at forsegle og bevare loftspladerne. Fjernelsen af loftspladerne kan måske i virkeligheden bedre ses som en politisk handling der skulle signalere at man gjorde noget ved problemerne.

⁷ Politikens Nudansk Ordbog med etymologi. København, 2000.

⁸ Oxford Advanced Learner's Dictionary. Oxford 1995. Opslagsordet er *evidence*.

Men ud over at være information der giver en stærk begrundelse for noget, er evidens også blevet en del af en institutionaliseret praksis for hvordan viden og evalueringer på samfundsplan bruges til at udvikle politik og indsatser "så vi er sikre på, at de virker". Som nævnt taler man om Cochrane-tilgangen inden for det medicinske felt og Campbell-tilgangen inden for bredere samfundsmæssige områder. Inspirationen til at tænke i evidensbaseret politik kommer fra det medicinske område. Generelt står forestillinger om meget skarpe metodekrav til hvad der overhovedet tæller som evidens, uhyre stærkt i både Cochrane- og Campbell-tilgangene. Begge favoriserer de såkaldte Randomiserede Control Trials (RCTs) hvor det er 'randomiseret' hvem der modtager behandlingen eller indsatsen, og hvor effekten måles både hos modtagerne af behandlingen/indsatsen, og hos en kontrolgruppe som ikke modtog den. Begge tilgange ser tilsvarende bort fra den information som stammer fra kilder der ikke lever op til sådanne skrappe metodekrav. Det betyder i praksis at der ofte ikke er meget evidens at basere sin politik på når den gyldige viden skal sammenvejes. Derved opstår der et valg hvor man kun kan vælge indenfor virkemidler som det er muligt at undersøge med de snævre metodekrav, uanset om de passer til den relevante kontekst. Andre virkemidler som potentielt kunne være mere effektive, vælges derfor fra fordi anvendelsen af disse virkemidler ikke kan randomiseres eller på anden måde tilpasses metodekravene.

Andre forståelser giver forslag til hvordan man kan arbejde med evidens der er tættere koblet til praksis. Pawsons (2006) realistiske tilgang til evidensbaseret politik har vi allerede refereret flittigt til. Mere bredspektret har man, fx på uddannelsesområdet, set en række forsøg på at omfavne og måske samtidig også opløde en evidensbaseret tilgang til udviklingen af politikker og praksis (se fx Davies et al. 2012; Hargreaves 1997). Endeligt kan man pege på Rieper og Foss Hansen (2007) der problematiserer den evidensstænkning som placerer RCT-produceret evidens i toppen af hierarkiet. Deres bud er at det er mere relevant at tale om evidensstypologier frem for om et evidenshierarki, for så vidt som forskellige typer af metoder producerer viden som har forskellig relevans afhængigt af den sammenhæng den producerede viden skal bruges i.

8.2 VALIDITET

Gyldighed af en oplysning eller validiteten, som det kaldes i forskning, er som nævnt udover relevansen det andet af to kriterier for evidens. Validiteten deles traditionelt op i to typer (Krogstrup 2011): intern og ekstern validitet. Den interne validitet handler om hvorvidt den anvendte metode (teoretisk og praktisk) faktisk er i stand til at påvise de sammenhænge som man vil undersøge. I epidemiologiske undersøgelser er et væsentligt spørgsmål for den interne validitet således at kontrollere for alle alternative årsager til den årsags-virkningssammenhæng som undersøges. Mens den eksterne validitet handler om mulighederne for at generalisere resultatet til en bredere sammenhæng. Det vil sige om den effekt som er fundet i en undersøgelse, også kan forventes at opstå i en anden sammenhæng – altså i en anden kontekst.

Et eksempel fra epidemiologisk forskning i det psykosociale arbejdsmiljø kan illustrere denne problemstilling. Det er en nærmest fastslået kendsgerning at høj indflydelse på arbejdet har en positiv indflydelse på helbredet. Det er påvist i mange undersøgelser og opfattes på mange måder som en relativt kontekstfri sammenhæng. En nylig finsk undersøgelse af skovarbejdere (Joensuu et al. 2012a) stiller imidlertid spørgsmålstegn herved. I et forløbsstudie fandt forskerne at skovarbejdere med høj beslutningsmulighed i arbejdet havde en forøget hjerte-kar-dødelighed. Forklaringen på resultatet kan på grund af undersøgelsesdesignet ikke direkte udledes af den gennemførte forskning, men den skal formodentlig netop findes i konteksten (Joensuu et al. 2012b; Rugulies 2012). De finske skovarbejdere blev i 1990'erne udsat for en betydelig omstrukturering af branchen, og det er måske de skovarbejdere

med størst beslutningskompetence som har været hårdest ramt af denne omstrukturering. Dette eksempel illustrerer netop at selvom den interne validitet af de mange undersøgelser som påviser den positive sammenhæng mellem indflydelse og helbred, er høj, så er den eksterne validitet usikker fordi der åbenbart er sammenhænge hvor høj indflydelse kan føre til øgede belastninger.

I forbindelse med diskussionen om validitet skal det også nævnes at der i forskningskredse også opereres med begrebet reliabilitet – eller pålidelighed – når man taler om validitet i forbindelse med en vurdering af forskningskvalitet (Krogstrup, 2011). Dette begreb refererer til muligheden for at andre forskere er i stand til at gentage undersøgelsen og nå frem til de samme resultater. I samfundsmæssige indsatser er det stort set aldrig muligt at gentage den samme indsats i en uændret form, og det er derfor ikke muligt at operere med reliabilitet i traditionel forstand. I stedet kommer reliabiliteten til at handle om en omhyggelig og kvalificeret metodebeskrivelse. Den giver andre forskere, evaluatore og beslutningstagere mulighed for at vurdere om en undersøgelse metodemæssigt er gennemført på en kvalificeret måde.

Hovedformålet med evaluering er altså at fremskaffe den bedst mulige evidens for beslutninger om fremtidige handlinger. Det er i den sammenhæng problematisk at anvende udtrykket: "Der er evidens for..." Det er vanskeligt – udover banale eksempler – at forestille sig at der for virkemidler i arbejdsmiljøindsatsen fremskaffes den endegyldige evidens for, hvornår noget virker eller ikke virker. Ovenstående eksempel fra de finske skovarbejdere viser netop hvordan en ret generaliserbar viden om de positive aspekter af indflydelse i de konkrete tilfælde også skal vurderes i deres egen konkrete sammenhæng.

I realiteten vil beslutningstagere (politikere, myndigheder, organisationer m.v.) derfor altid skulle træffe beslutninger på et ufuldstændigt vidensgrundlag. Et krav om fuld og sikker viden vil kun skabe handlingslammelse. Et eksempel er børneopdragelse: Forældrene står principielt i en vanskelig situation. De vil gerne give deres børn den bedst mulige opdragelse, men selvom de læser alt hvad der er skrevet om børneopdragelse, vil de aldrig have tilstrækkelig sikker viden til altid at træffe de rigtige beslutninger i en konkret situation. Samtidig er de nødt til hver eneste dag at træffe beslutninger. Handlingslammelse vil under alle omstændigheder være næsten den værst tænkelige løsning på problemet.

Viden genereret ved evaluering har imidlertid det indbyggede problem at den i princippet altid er bagudrettet (selvom procesevaluering kan have en vis samtidighed). Beslutningen tages, indsatsen sættes i gang, og først når den er afsluttet, vil der foreligge et evalueringsresultat. På det tidspunkt har den oprindelige beslutning nærmest kun historisk interesse. Beslutningstagerne står over for nye beslutninger om konkrete indsatser som i sagens natur endnu ikke er evalueret.

Alt for ofte ender evalueringsresultater derfor med ikke at blive anvendt. De kommer blot til at udfylde et sekundært formål med at legitimere den oprindelige beslutning og påvise at beslutningstagerne optræder forsvarligt ved at gennemføre evaluering af deres indsatser.

Selvom der handles på et ufuldstændigt grundlag, er der stadig både behov for og mulighed for at anvende eksisterende viden. Der vil typisk være en stor mængde viden til rådighed som både omfatter tidligere evalueringer, forskning og ikke mindst beslutningstageres praktiske erfaringer både med tidligere indsatser og med at få indsatser tilpasset til noget som er politisk realiserbart. Det er netop ved beslutninger på et ufuldstændigt grundlag at en eksplicit og reflekteret programteori kan spille en vigtig rolle. Gennem arbejdet med at sætte ord på programteorien bringes viden og erfaringer i spil samtidig med, at man over for sig selv og andre begrundet, hvorfor man antager, at indsatsen vil virke. Der vil være elementer i programteorien, der er baseret på relativt sikker viden; mens den tilgængeli-

ge viden for andre elementer vil være mere usikker eller manglende. Her må man gennemføre et kvalificeret skøn for at bygge bro mellem de områder, hvor viden er mere sikker (Pawson, Wong, & Owen 2011 & Hasle et al 2012).

8.3 LEGITIMITET OG LÆRING

Vi fremhævede i indledningen af dette kapitel at hovedformålet med evaluering af arbejdsmiljøindsatser må være at blive klogere, således at de næste indsatser kan kvalificeres, og muligheden for en væsentlig effekt dermed kan øges. Formålet er altså læring. Men det er vigtigt at huske at evalueringer i realiteternes verden også har en anden funktion: At legitimere de beslutninger der tages, således at både politik, indsatser og virkemidler, samt den dertil knyttede ressourceanvendelse fremstår som både fornuftige og rimelige. Især i en tid hvor evalueringer får et omfang som det vi ser i disse år, kan man nogle gange få den mistanke at legitimeringen har fået overtaget i forhold til læringen – at evalueringen bliver et mål i sig selv. Selvom det ikke skulle være tilfældet, er det værd at have i baghovedet at evalueringer – på godt og ondt – tjener begge formål.

Opfyldelsen af læringsformålet kan således blive hæmmet af de interesser som knytter sig til evalueringen. Bestilleren af en evaluering og/eller bevillingsgiver kan have interesse i at legitimere egen indsats og evaluator en interesse i at stille sin kunde tilfreds. Forskellen mellem læring og legitimering er ikke klar og entydig, men skal snarere opfattes som en analytisk skelnen. I praksis vil der være flydende grænser med større eller mindre overlap mellem de to.

Normalt lægger man vægt på at få en uvildig evaluering, men det er i praksis mere et ideal end en realitet. En evaluering bliver typisk sat i gang af de samme beslutningstagere som er ansvarlige for den indsats der skal evalueres. Mens dette sammenfald kan give relativt gode muligheder for at anvende evalueringsresultater, fordi det er den primære bruger som bestiller evalueringen, er der også en bagside. Beslutningstagerne vil næppe have interesse i evalueringsresultater som peger på, at det var en forkert beslutning at gennemføre indsatsen, og at pengene er spildt. Samtidig er evaluator typisk en konsulentvirksomhed som lever af at lave evalueringer, og den har måske ikke interesse i at præsentere kritiske resultater som risikerer at bestilleren af evalueringen bliver utilfreds. Omvendt er evaluator selvfølgelig afhængig af at bevare sit faglige renommé.

Selvom såvel bestiller som evaluator formodentligt er pålidelige institutioner og enkeltpersoner som gør deres bedste for at en evaluering bliver brugbar, så kan interesseløshed ikke undgå at præge evalueringen. Det kan ske i større eller mindre grad. I nogle tilfælde er der ikke så mange potentielle problemer knyttet til en evaluering, mens der kan være mere kontroversielle spørgsmål i andre. Men overordnet set må man antage, at evalueringsprocessen i et vist omfang påvirkes af de involverede aktørers forståelser og interesser. I beskrivelsen af evalueringsopgaven vil beslutningstageren typisk nedprioritere spørgsmål som potentielt kan stille de centrale beslutningstagere i dårligt lys fx ved at fokusere på andre aktørers rolle frem for egen rolle. Tilsvarende vil evaluator være tilbøjelig til at nedtone negative resultater, hvis man får indtryk af at de giver problemer for bestilleren. Det vil ikke ske ved at forfalske resultater, men ved at fremhæve nogle resultater frem for andre og fx ved at forklare negative resultater med uforudsete kontekstuelle faktorer fx målgruppers økonomiske problemer eller andre aktørers manglende engagement. Det vil sjældent være forhold som er forkerte, men der sker let en vægtning af materialet, som aflaster bestillerens ansvar. Det samme hensyn kan gøre sig gældende over for andre betydningsfulde aktører som har været inddraget i den evaluerede indsats.

Problemet bliver større jo mere omfattende – og dyrere – indsatsen er og dermed også den tilsvarende størrelse af evalueringen, men ved de større evalueringer kan man forsøge at modvirke denne tendens. Det kan fx ske ved at inddrage forskningsinstitutioner som må antages at være mindre afhængige af bestilleren, selvom forskningsinstitutioner til stadighed bliver stadig mere afhængige af ekstern finansiering, og de kan derfor også have en interesse i at bevare gode relationer til bestilleren. Denne problemstilling kan man forsøge at imødegå ved at inddrage paneler af uafhængige forskere/eksperter som har begrænsede økonomiske interesser i den pågældende evaluering og i bestilleren. Der kan dog aldrig findes nogen personer eller institutioner som er helt interessefrie, og samtidig gør den slags ekstra tiltag evalueringen dyrere og mere kompliceret. Det vil derfor typisk være ved mere omfattende evalueringer af store programmer at sådanne omhyggelige forsøg på at imødegå risikoen for interessebias vil være relevant.

Afsluttende kan vi opsummere at evaluering som aktivitet er præget af en række udfordringer. Først og fremmest i forhold til hvordan evalueringen skal gribes an, så den kan vise det den angiver at vise. Men også i forhold til at evalueringer nogle gange har fået et næsten rituelt tilsnit (Dahler-Larsen 1998). At gøre en evaluering til mere læring end legitimering kræver en indsats fra alle involverede – beslutningstagere, de udførende, indsatsens modtagere og evaluatorene.

I det næste kapitel vil vi vise hvordan man kan skalere forskellige typer af indsats med forskellige tilgange til evaluering. Derigennem bliver det muligt at opnå det mest optimale evalueringsdesign og undgå nogle af faldgruberne ved pragmatisk evaluering.

9. EVALUERINGSSCENARIER

Beslutningstagere som skal iværksætte en evaluering, står altså med en vanskelig opgave. Programteorien om den arbejdsmiljøindsats der skal evalueres, må forventes at være mere eller mindre ufuldstændig, der skal vælges mellem forskellige måder at gennemføre evalueringen, midlerne til at gennemføre den er begrænsede, og så har både beslutningstagere og evaluator større eller mindre interesser i visse udfald af evalueringen. Der skal derfor findes pragmatiske løsninger hvor 'det bedste ikke skal blive det godes værste fjende', samtidig med at de mulige pragmatiske løsninger ikke ender med at være så meget 'quick and dirty' at resultaterne af evalueringen bliver misvisende og risikerer at gøre mere skade end gavn.

Der vil sjældent entydigt være tale om direkte skadelige evalueringer, men et par tænkte problematiske eksempler kan anvendes som illustration: Hvis man fx måler på output, hvilket er et billigt og ofte anvendt mål (fx antal hits på en hjemmeside), kan man måske få et billede af at en bestemt indsats har vakt interesse og dermed er godt på vej. Imidlertid er de mange klik måske blot udtryk for at brugere alligevel er inde på en hjemmeside som de ofte anvender, og i den forbindelse blot undersøger om en nyhed om indsatsen er interessant, og det konstaterer de måske at den faktisk ikke er. Et andet eksempel kan være et tilfælde hvor man gennemfører en række pilotforsøg. Her kan en billig evalueringsmetode bestå i at ringe til de projektansvarlige for at undersøge hvordan det er gået med forsøgene. Men i det tilfælde vil de projektansvarlige typisk have en relativ positiv holdning, netop fordi de selv har et medansvar for at gennemføre forsøget, og man risikerer dermed at få et positivt billede af fremdriften som ikke nødvendigvis er retvisende.

Det væsentligste bliver at få tilpasset evalueringssønskerne til de tilgængelige ressourcer på en sådan måde at en tilstrækkelig høj kvalitet kan opretholdes, og det skal ske samtidig med at der tages hensyn til den politiske kontekst. Hovedformålet er læring som skal anvendes til at forbedre den næste indsats. Praksis vil imidlertid være meget afhængig af hvilken type indsats der er tale om. Der vil være store forskelle imellem den begrænsede indsats og den omfattende indsats, og i dette kapitel åbner vi yderligere for denne diskussion gennem en diskussion af mulighederne for at gennemføre evalueringer som på den ene side giver relevant viden, og på den anden side forholder sig til de konkrete ressourcemæssige og politiske muligheder. For at kunne gøre dette er det nødvendigt at skelne mellem forskellige typer af arbejdsmiljøindsatser, som hver for sig har forskellige muligheder og begrænsninger. De fire typer af evaluering vi diskuterer, er:

- Evaluering af den begrænsede indsats
- Evaluering af den omfattende indsats
- Tværgående evaluering
- Evaluering af et program

9.1 FØR INDSATSEN

For alle fire typer er der en række overvejelser som med fordel kan ske før den konkrete evaluering designes. Det er en væsentlig forudsætning for en god evaluering at indsatsens formål og design er klare og gennemskuelige, og at begge dele bygger på en eksplicit og relevant programteori som har klare og troværdige antagelser om hvordan indsatsen vil virke. Samtidig er det et væsentligt kriterium at evalueringen skal være så anvendelig som mulig at deltagerne i indsatsen hvilket bl.a. kan sikres ved at inddrage deltagerne i tilrettelæggelsen af evalueringen. Evalueringen starter derfor på mange måder ved beslutningen om og planlægningen af indsatsen, og det gælder uanset om der er tale om en

begrænset eller en omfattende indsats. Der vil selvfølgelig være forskel på, hvor omfattende denne planlægning vil være afhængigt af forskellen, men den vil under alle omstændigheder være relevant. Selv i den begrænsede informationsindsats hvor der skal lægges nye oplysninger på en hjemmeside, vil det være relevant at gøre sig overvejelser om målgruppen: fx spørgsmål om informationen opfylder den pågældende målgruppes behov, og om målgruppen kan finde frem til informationen.

Og her kan der måske være behov for at indsparke et lille forbehold. Grundlæggende er det vores anbefaling at evalueringsaktiviteten medtænkes allerede i planlægningen af indsatsen, fordi det forøger chancerne for en god evaluering. Men det må ikke forstås på den måde at det er behovet for en evaluering som ender med at definere indsatsens aktiviteter. I nogle sammenhænge er det netop beskrevet i forskningen hvordan behovet for at registrere og især kvantificere i sig selv omformer den aktivitet, der skal registreres (Power 1996). Men det er jo et dilemma hvorvidt man skal gennemføre den indsats, man er sikker på virker i et vist men begrænset omfang, fordi der er god evidens for det, eller om man skal gennemføre den indsats man tror på virker meget bedre, men hvor evidensen er svagere, eller den nyskabende indsats som man aldrig har prøvet før, og som derfor slet ikke har nogen evidens overhovedet. Den forøgede opmærksomhed på evalueringer og evidens vil tendentielt skubbe løsningerne i den forsigtige retning - og favorisere aktiviteter som kan tælles.

I forhold til at medtænke evalueringen i planlægningen af selve indsatsen kan det være en idé at fastlægge procedurer for udvikling og beskrivelse af de projekter som skal gennemføres som led i en arbejdsmiljøindsats. Der bør tages hensyn til omfanget af indsatsen, således at der ikke anvendes for mange ressourcer til planlægning og dokumentation af små aktiviteter. Væsentlige projektelementer der bør beskrives som led i sådanne procedurer, kunne være:

- Formål
- Målgruppe
- Aktiviteter (indsats – virkemidler)
- Ressourceforbrug
- Succeskriterier
- Programteori
- Og dertil kommer: Evalueringen.

En sådan praksis vil styrke muligheden for at få et højt udbytte af de evalueringer der gennemføres – også når ressourcerne er små, og den systematiske planlægning vil også forøge indsatsens kvalitet. Vi har igennem denne publikation fremhævet vigtigheden af at eksplicite programteorien både for at kvalificere indsatsen og for at kunne tilrettelægge evalueringen. Men også det at indtænke evalueringen på forhånd er vigtigt: Det vil fx ofte være billigere fra starten at lave en løbende registrering som kan indgå som evalueringsdata, frem for at gennemføre en efterfølgende retrospektiv dataindsamling. Det er selvfølgelig også nyttigt på forhånd at fastlægge omfanget af evalueringen så det er kendt af alle relevante interessenter; det kan være en stor frustration for projektdeltagere når de på bagkanten bliver bedt om at indrapportere en række informationer som de stort set aldrig har lige ved hånden.

Efter at have peget på behovet for at planlægge evalueringen samtidig med selve indsatsen, vil vi nu gå over til at diskutere evalueringer i forhold til forskellige typer af indsats.

9.2 EVALUERING AF DEN BEGRÆNSEDE INDSATS

En stor del af specielt BAR-aktiviteter må opfattes som ret begrænsede indsatser. Indsatser til hvilke der bevilges ret små beløb, og hvor de fleste ressourcer i realiteten stilles til rådighed af de organisati-

oner som leverer den arbejdskraft som indgår i at tilrettelægge og gennemføre indsatsen. Typiske eksempler er informationsaktiviteter af skriftlig karakter som fx en BAR-vejledning, en informationspjece og et nyt tema på en hjemmeside. Det kan også være andre ting som et informationsmøde hvor der præsenteres nye forskningsresultater eller udveksles erfaringer mellem relevante aktører.

Ressourcerne til at anvende på evaluering af denne type indsatser vil naturligvis være ganske små, og det skal derfor nøje overvejes hvordan de få ressourcer kan anvendes på en måde som faktisk bidrager med relevant viden. Mange registrerer løbende hits på hjemmesiden, afsætning af trykte materialer og antal deltagere på informationsmøder. Sådanne registreringer kan gennemføres med ret lave omkostninger og kan være anvendelige til dels at sikre sig at der overhovedet er interesse for det man laver, dels at skabe legitimitet om sine aktiviteter. Hvis ingen benytter sig af dem, bliver legitimiteten naturligvis lav. Det er derfor under alle omstændigheder vigtigt at forsætte med denne praksis, men samtidig at være opmærksom på at sådanne registreringer ikke fortæller om materialet bliver brugt til noget relevant, om brugerne kan forstå indholdet, og om der er informationer som mangler.

Derudover kan det også være relevant at overveje, om de potentielle evalueringsressourcer måske bedre kan anvendes i en grundigere forberedelse af indsatsen hvor der dermed bliver plads til at udvikle en mere eksplicit programteori. Her kan eksperter og forskere i nogle tilfælde måske med fordel anvendes til rådgivning i denne tidlige fase frem for i en efterfølgende evaluering som bliver for overfladisk til at tilføre en virkelig værdi.

En egentlig evaluering med ekstern evaluator er ikke mulig ved de begrænsede indsatser. Udover ovennævnte vil den mest relevante tilgang derfor være selvevaluering hvor de aktører som gennemfører indsatsen, også evaluerer hvordan den er forløbet. Hovedformålet vil være at korrigere indsatsen mens den er i gang samt opbygge viden som kan anvendes i den næste tilsvarende indsats. Selvevalueringer kan også være relevante i forhold til de andre typer af evalueringer vi kommer til nedenfor, men vi omtaler dem her, fordi de nok er mest relevante i forhold til de begrænsede indsatser.

En selvevaluering vil i denne sammenhæng bestå i to elementer – dataindsamling og kritisk refleksion. Evalueringen giver kun mening hvis der er indsamlet data som understøtter den kritiske refleksion. Men data skal ikke forstås for snævert: Det drejer sig for det første om de løbende registreringer af hits, efterspørgsel, afsætning og så videre der kan belyse om der er en grundlæggende interesse for den pågældende indsats, og dermed om der eventuelt er behov for større ændringer. Årsagerne til at det går godt eller mindre godt, kan dog kun dårligt belyses gennem et sådant materiale. Der er derfor for det andet behov for at indsamle et mere kvalitativt datamateriale. Her er det de involverede aktører som selv indsamler erfaringer og materiale. Der er ikke ressourcer til at bede andre om det, og derfor må de involverede selv opsøge feedbackmulighederne. Det vigtigste element ved de små indsatser er at de involverede benytter de lejligheder der opstår i forbindelse med andre aktiviteter, til at efterspørge feedback. Når de deltager i møder udenfor eget kontor, kan de spørge de personer de møder, om de har set materialet, om de kender problemstillingen for materialet, om de synes den er vigtig, hvordan de selv arbejder med den osv. Den kritiske refleksion kan fx tage form af et refleksionsmøde hvor de involverede aktører samles. Den feedback som de forskellige personer har indsamlet, præsenteres, og det overvejes i fællesskab hvordan det skal fortolkes, og hvilke konsekvenser det kan få for den igangværende indsats. Skal materialet fx revideres eller suppleres. Mødet afsluttes med at vurdere om der er læringer som kan anvendes i fremtidige lignende indsatser. Resultaterne af mødet skrives ind i et referat som dels kan anvendes operationelt til at fastholde beslutningerne, og dels kan dokumentere at evalueringen er gennemført.

9.3 EVALUERING AF DEN OMFATTENDE INDSATS

Grænsen mellem en mindre indsats og den omfattende indsats vil ikke ligge fast, men jo mere omfattende en indsats bliver, jo lettere vil det andet lige være at skaffe ressourcer til en evaluering. Men i langt de fleste tilfælde vil ressourcerne stadig være en begrænsende faktor. Normalt anvendes der under 10 % af en bevilling til evaluering, og man skal derfor op på en million kroner for at få blot 100.000 kr. til evaluering, og selv for dette beløb er der stadig ikke plads til særlig ambitiøse evalueringdesign. Eksempler på omfattende indsats kan i BAR-regi være BAR SoSus social kapital projekt, I-BARs Road show indenfor maskin- og metalindustrien og BAR Bygs kampagne rettet mod små virksomheder i forbindelse med screeningen i 2005/6.

Selv ved større projekter vil ressourcerne altså også være begrænsede, og der skal derfor ske en prioritering. Der kan typisk være ønsker om den pragmatiske evaluering med 'lidt fra alle hylder'. Lidt procesevaluering for at korrigere undervejs, lidt effektevaluering for at få viden om hvorvidt indsatsen flytter noget, og lidt virkningsevaluering for at få indblik i hvad der virker. I nogle tilfælde kan et sådan design måske lade sig gøre - En procesevaluering, nogle simple output-mål og en evaluering af afgrænsede dele af programteorien som det vurderes har særlig interesse. Men det er også her, hvor ambitionerne kan skyde over målet, og resultatet blive for overfladisk til at bidrage med anvendelig viden.

Der vil i sagens natur, fordi der skal bevilges flere penge, blive stillet større krav om udarbejdelse af en omfattende projektbeskrivelse som følger en systematisk projektstruktur som beskrevet i afsnit 6.5. For at sikre en god evaluering er der behov for klare programteorier som kan danne grundlag for at udpege de centrale antagelser og indikatorer, som skal prioriteres i en evaluering. Også her vil det være fornuftigt at planlægge evalueringen på projektdesign tidspunktet. Derved kan det sikres, at de nødvendige midler til evalueringen er til stede. Hvis der skal anvendes en ekstern evaluator, kan vedkommende udpeges på et tidligt tidspunkt og dermed starte dataindsamlingen. Samtidig kan implementeringen af indsatsen tilrettelægges således at den er gennemskuelig, og data fra starten løbende indsamles på en omkostningseffektiv måde. Fx vil det være lettere at vurdere udbyttet af et Road show som i maskin- og metalindustrien hvis man løbende har registreret alle markedsførings tiltag og efterfølgende kan sammenholde dem med det faktiske fremmøde.

Selv ved større projekter vil midlerne til ekstern evaluering som nævnt være begrænset. Det må derfor forventes at de involverede personer også i dette tilfælde selv må stå for en stor del af dataindsamlingen. Det kan være lettere i dette tilfælde, fordi man ofte har mere kontakt med det praktiske felt indsatsen gennemføres i. Ved Road showet og møderækker for små virksomheder indenfor byggeriet har BAR-konsulenter og repræsentanter for organisationerne fx deltaget i en lang række møder hvor de som deltagerobservatører har kunnet iagttage, hvordan deres målgruppe har reageret på indsatsen. Viden opnået på denne måde anvendes selvfølgelig umiddelbart af de involverede til både at vurdere den igangværende indsats og i udviklingen af nye indsats, men den kan også indgå som led i en systematisk evaluering. Med læring som et klart hovedformål med evalueringen er det ud fra en ressourcemæssig betragtning ikke nødvendigvis den bedste brug af en ekstern evaluator at gennemføre den størst mulige adskillelse mellem indsats, de ansvarlige og evaluator. Det bliver hurtigt tidskrævende for evaluator at sætte sig ind i hele implementeringsforløbet og i den forbindelse at gennemføre en længere interviewrække. Det kan både være mere omkostningseffektivt og udbytterigt for læringen at evaluator kun udfører en begrænset selvstændig dataindsamling som primært anvendes til at udfordre de indsatsansvarliges programteori og fortolkning af forløbet. Det kan ske på en eller flere workshops hvor programteori og implementering diskuteres igennem. De ansvarlige har som deltagerobservatører et dybt indblik i forløbet, og evaluator har som en ekstern person mulighed for at udfordre de forudfattede meninger. Det samlede resultat kan dokumenteres af evaluator således at der forelig-

ger en systematisk evalueringsrapport, der kan anvendes til at fastholde læringen og samtidig dokumentere at der er gennemført en evaluering.

9.4 TVÆRGÅENDE EVALUERING

En særlig mulighed som i dag anvendes i et vist – omend beskedent – omfang, er forskellige former for tværgående evalueringer. Der er netop ved at blive igangsat en større evaluering af BAR-systemet som kan tjene som et eksempel herpå, men man kunne også tænke sig sådanne evalueringer anvendt i større omfang i mere fokuserede udgaver. Her kunne man i en BAR eller flere BAR'er tage initiativ til at gennemføre tværgående evalueringer af sammenlignelige virkemidler eller indsatser. Det kan dreje sig om at samle flere aktiviteter med væsentlige fællestræk som kan omfatte interaktive hjemmesider, eller indsatser som handler om møde- og netværksaktiviteter. En sådan tilgang kunne give flere klare fordele. For det første kan der samles flere ressourcer sammen, således at der bliver mulighed for at foretage en mere ambitiøs evaluering. For det andet kan regelmæssige tværgående evalueringer erstatte behovet for at gennemføre evaluering af alle indsatser, hvor der også bruges evalueringsressourcer på små rutinebaserede aktiviteter hvor udbyttet vil være begrænset. For det tredje vil evaluering af flere parallelle indsatser i forskellige kontekster give mere generaliserbar viden om mekanismer og kontekst i denne form for indsats.

I nogle tilfælde kunne det være relevant at iværksætte det man kan kalde en metaevaluering (Pawson 2006), hvor resultaterne fra et større antal allerede gennemførte evalueringer analyseres og sammenlignes på tværs. Det kan fx være evalueringer som er udarbejdet efter de ideer der er skitseret i de to foregående afsnit om de begrænsede og de mere omfattende indsatser. Ved en metaevaluering produceres ikke ny empiri, men der bygges på allerede eksisterende resultater fra mere projektnære evalueringer. En sådan tilgang til evaluering vil give gode muligheder for at få dybere indsigt i de anvendte virkemidler, deres mekanismer i samspil med kontekst hos modtagerne og ikke mindst deres generaliserbarhed. Et aktuelt eksempel kunne være Forebyggelsesfondens støtte til en lang række projekter der anvender forskellige former for "barfodsaktører" – medarbejdere der melder sig til i en tidsafgrænset periode at arbejde for at udbrede en særlig forebyggelsesindsats. Det kan fx være forflytningsvejledere, stress coaches, mentorer eller trivselsambassadører. En tværgående evaluering af denne type projekter kunne give en værdifuld viden om, hvornår denne indsatsstrategi virker, og hvad der kan begrænse dens udbytte.

Større tværgående evalueringer er ofte sat i gang fra centralt politisk hold. Det gjaldt både den tidligere Capacent evaluering af BAR-systemet og den kommende BAR-evaluering. Der er også eksempler på indsatser, der udvikles og gennemføres af en række BAR'er i fællesskab og derved kan danne grundlag for en tværgående evaluering. Et eksempel er kampagnen: "Gravid med job" som havde til formål at forebygge at gravide forblev i arbejdsfunktioner der kunne skade dem selv eller fosteret. Kampagnen var iværksat af seks BAR'er (BAR Byg, I-BAR, BAR jord til bord, BAR for service og tjenesteydelser, Grafisk BAR og BAR transport og engros).

Mens vi for de to tidligere typer af indsatser har argumenteret for at det er en god idé at forberede evalueringen på forhånd, vil det typisk være anderledes med den tværgående evaluering. Det kan gennemføres i tilfældet med en tværgående kampagne som fx 'gravid i job', men i mange tilfælde vil behovet for den tværgående evaluering først opstå, når det viser sig at flere forskellige aktører anvender sig af sammenlignelige virkemidler. Fx er det først efter det i praksis har vist sig, at der er mange der har ansøgt om projekter i Forebyggelsesfonden med et væsentligt element af barfodsaktører, der opstår en mulighed for at lave en værdifuld tværgående evaluering af dette tema. Den tværgående evaluering

giver i de tilfælde kun mening hvis der kan opstilles en reflekteret programteori, som er fælles for væsentlige elementer af de indsatser som skal indgå i den tværgående evaluering. I tilfældet med barsktsaktører kunne programteorien bestå i at uanset om der er tale om en forflytningsvejleder eller en stresscoach, vil det have en bedre effekt at det er en kollega, som formidler indsatsen frem for en ekstern ekspert.

Den tværgående evaluering vil derfor tage udgangspunkt i at formulere den fælles programteori og på grundlag heraf prioritere evalueringen således at der ikke anvendes ressourcer på at evaluere problemstillinger, som kun er relevante for en mindre del af de indsatser som indgår i den tværgående evaluering, og som derfor ikke kan bidrage til at opbygge en tværgående viden om virkningsmekanismerne.

9.5 EVALUERING AF ET PROGRAM

Et program er en stor samlet indsats – oftest baseret på en politisk fastlagt prioritering og i nogle tilfælde en afsat økonomisk ramme. De specifikke aktiviteter som gennemføres som led i programmet, er oftest koordineret på forskellige måder og forløber over en længere periode. Et konkret eksempel er 2020-arbejds miljøstrategien hvor flere af de enkelte aktiviteter i planen nærmest kan beskrives som et program i sig selv. Det gælder fx indsatsen rettet mod børn og unge. Et andet eksempel er at tilbyde specielt designede forebyggelsespakker til mindre virksomheder indenfor en række nedslidningstruede brancher. Pakkerne består af konkrete aktiviteter, som har til formål at forebygge nedslidning og udstødning på den enkelte virksomhed. Denne aktivitet indgår som en del af programmet for Forebyggelsesfonden som er en konkret udmøntning af et politisk program om at begrænse nedslidningen og udstødningen fra arbejdsmarkedet.

Programmer implementeres normalt igennem mange forskellige aktiviteter, som forløber over en længere periode, og som er organiseret i en række indsatser, som hver især anvender flere forskellige virkemidler. Mens både begrænsede og omfattende indsatser sjældent bygger på en forestilling om at de har en direkte målbar effekt på sundhed og trivsel, er det imidlertid anderledes med programmer. Her vil man ofte have sådanne effektmål. Det gælder ikke mindst for den samlede 2020-arbejds miljøstrategi, men selvom programmet normalt vil indeholde en bred vifte af omfattende aktiviteter, vil det også her være vanskeligt at etablere sådanne direkte kausale sammenhænge.

Evaluering af programmer bliver derfor i sagens natur en kompliceret opgave der må omfatte flere forskellige evalueringstilgange og metoder. I det følgende giver vi et eksempel på hvordan det blev gjort i forbindelse med evalueringen af Forebyggelsesfondens aktiviteter.

I 2011 formulerede Forebyggelsesfonden i et udbud kravene til en evaluering af Forebyggelsesfondens virksomhed. Man ønskede en evaluering af samtlige støttede projekter inden for Fondens fire hovedformål. Desuden skulle den ny tilskudsform 'forebyggelsespakkerne' evalueres. I udbuddet var der redegjort for at man ønskede både proces- og effektevaluering, og med hensyn til effektevaluering skulle evaluator specifikt redegøre for at der kunne gennemføres en sådan evaluering med brug af kontrolgrupper. Man ønskede desuden at evalueringen blev lærende, sådan at evalueringresultaterne undervejs kunne anvendes til at forbedre Fondens aktiviteter.

Evalueringen kunne tolkes således at hovedformålet var at lære hvad der virker godt, og hvad der ikke virker i de gennemførte projekter. Heri skulle indgå hvilke virkemidler, metoder, koncepter og tilgange projekterne har valgt at anvende, hvilke rammebetingelser projekterne har haft, og hvilke effekter der kan måles i de enkelte projekter. Evalueringen forventede man på den ene side kunne bidrage til at

spredte gode eksempler til andre arbejdspladser og jobgrupper, og på den anden side – og måske vigtigst – skulle den kvalificere Forebyggelsesfondens arbejde i forhold til fremtidige udbud af støtte til projekter. Et væsentligt element heri er at kunne bidrage til at udvikle større og mere præcise krav til projekternes egen evaluering. Endelig indeholder evalueringsskemaet et krav om at evalueringen parallelt med de andre formål kan klargøre om man kan identificere samfundsmæssige betydninger af Forebyggelsesfondens indsats. På mange måder er der derfor tale om en egentlig programevaluering.

Hovedvægten i evalueringen ligger således på evalueringen af de talrige enkelte projekter og de mulige tværgående analyser. Selve programevalueringen – den overordnede evaluering af om Forebyggelsesfonden gennem de støttede projekter og forebyggelsespakker har formået at have en samfundsmæssig positiv virkning på arbejdsmiljøet – må ifølge udbuddet antages at få en mindre placering i evalueringsskemaet.

Evalueringen opbygges således at evaluator gennem evalueringen af de mange enkelte projekter efterfølgende bliver i stand til at kunne gennemføre en række analyser på tværs af de mange evaluerede projekter, og herved nå frem til generaliserede konklusioner om hvad der virker og ikke virker i projekterne. Udfordringen som står tilbage, er at sikre at evalueringen af en så stor mængde meget forskellige projekter bidrager til en større forståelse af, hvilken samlet virkning på arbejdsmiljøet Forebyggelsesfonden har.

Denne evaluering er kendetegnet ved at den blev igangsat flere år inde i Forebyggelsesfondens levetid. Det vanskeliggør åbenlyst mulighederne for at lave effektevaluering. Aktiviteterne har her et så stort omfang og med så store økonomiske ressourcer at det principielt ville være muligt mere systematisk at operere med kontrolgrupper, men i praksis er det vanskeligt når den skal igangsættes midt i forløbet, og denne type effektevaluering kommer derfor, så vidt det er oplyst, heller ikke til at indgå i evalueringen af Fonden. Historien om Forebyggelsesfonden er også et godt eksempel på de vanskelige vilkår evalueringer ofte må leve med når der er tale om evaluering af store programmer. Fondens overordnede formål og aktiviteter er nemlig for nyligt blevet ændret kun halvandet år efter at evalueringen blev sat i gang, og før der forelå brugbare resultater.

I forbindelse med 2020-arbejdsmiljøstrategien er der igangsat et overvågningsprogram som gennemføres i samarbejde mellem Arbejdstilsynet og NFA. Her igennem bliver det muligt at følge udviklingen i sundhed og trivsel samt i virksomhedernes arbejdsmiljøindsats, men det bliver vanskeligt at koble denne udvikling sammen med de aktiviteter som gennemføres som led i 2020-arbejdsmiljøstrategien. For at øge mulighederne for at foretage sådanne koblinger samt ikke mindst at foretage korrigerende handlinger undervejs, kunne det være en fordel at igangsætte en løbende registrering af de enkelte indsatsområder som sættes i gang som følge af planerne, på så tidligt et tidspunkt som muligt. Det kunne mest hensigtsmæssigt ske i tilknytning til de enkelte dele af planen hvor nogle dele er så omfattende at de nærmest udgør et program i sig selv. Det gælder fx for den del af planen som er rettet mod børn og unge. Her er der tænkt i flere forskellige indsatsområder som både retter sig mod virksomheder og uddannelsesinstitutioner og samtidig skal spænde over en meget bred vifte af brancher. Der er derfor også en meget bred vifte af aktører som skal inddrages. Det vil derfor være vigtigt for at udvikle denne indsats undervejs således at der skabes et overblik over hvem der deltager i aktiviteter, hvad de foretager sig, og hvad der kommer ud af det. Et led heri er at opstille indikatorer for den udvikling som skal sættes i gang. Fx kan målgruppen variere betydeligt i forhold til om man ønsker at nå de unge selv, lærerne på uddannelsessteder, og/eller de virksomheder hvor de unge arbejder. En programteori vil herefter kunne anskueliggøre hvor indsatsen kan føre til resultater, og den kan dermed danne grundlag for en analyse af udviklingen i aktiviteter og eventuelle behov for at korrigere undervejs.

9.6 AFSLUTTENDE BEMÆRKNINGER

Vi har i ovenstående lagt vægt på at skelne mellem de forskellige situationer som skabes af evalueringssituationernes forskellige omfang og karakter. Det er dog altid ønskeligt at indtænke evalueringen i forberedelsen af den egentlige aktivitet, fordi det alt andet lige vil give den bedste anvendelse af ressourcerne; især hvis det lykkes at undgå at evalueringen bliver et efterfølgende overraskende krav til de involverede om supplerende dataindsamling, møder, analyser, dokumentation og så videre.

Men ellers er det forskellen mellem aktiviteternes omfang, og hvad det betyder for evalueringen der er i fokus. Ved den begrænsede indsats er evalueringen af beskedent omfang. På den kvantitative side er det typisk kun output-mål der kan være relevante, og selve evalueringen vil ofte med fordel kunne gennemføres som en selvevaluering hvor det er de samme som står for aktiviteten, der med en vis systematik evaluerer den. Det er netop nærheden til aktiviteten som skaber gode betingelser for læring.

Ved den omfattende indsats vokser kravet til programteorien indsatsen. Selve evalueringen vil ofte kunne være større og også med brug af ekstern evaluator. Men igen gør behovet for fornuftig allokering af ressourcerne det nødvendigt at tænke løbende og ubureaukratisk dataindsamling og gerne med brug af interne datakilder. Det kan netop være i mødet mellem de involveredes egen (refleksive) programteori og den eksterne evaluators kritiske udfordring af denne at den største mulighed for læring ligger. Den endelige evaluering skal selvfølgelig også vende sig ud mod et bredere publikum med henblik på formidling af både læring og legitimering.

En særlig mulighed ligger i de tværgående evalueringer hvor erfaringerne fra sammenlignelige, men adskilte indsats, bringes sammen. Gjort rigtigt kan tværgående analyser afløse behovet for evaluering af de enkelte projekter, men også en tværgående evaluering baseret på allerede foretagne evalueringer – en såkaldt metaevaluering – kan være af stor værdi. Det er især mulighederne for på baggrund af forskellige kontekster at vurdere programteoriene og de mekanismer der trækkes på, som gør tværgående evalueringer interessante.

Evalueringer af programmer er indlysende den mest omfattende og komplicerede opgave. Men absolut med sine egne muligheder, fordi det netop er den samlede og måske synergiske virkning, produceret af de mange mere eller mindre sammenhængende indsats, der er i fokus. Samspillet mellem kvantitative og kvalitative data spiller bl.a. derfor en langt vigtigere rolle i sådanne evalueringer. Evalueringer af programmer indenfor arbejdsmiljøområder fylder mere og mere i de her år: Som eksempel er Forebyggelsesfonden blevet gjort til genstand for evaluering, og med 2020-arbejdsmiljøstrategien er der lagt op til flere evalueringer på programniveau. Ideelt set skulle sådanne store evalueringer kunne bygge på evalueringer af indsats på de lavere niveauer; men der kan være indbygget dilemmaer mellem læringsbehovene på de forskellige niveauer. Hvordan vi samfundsmæssigt sikres det fulde udbytte af de planlagte evalueringer står vel i et vist omfang stadigt åbent, for så vidt som det først er nu vi får så mange af dem, men der er ingen tvivl om, at politikdannelsen på arbejdsmiljøområdet – på linje med politikdannelsen på en lang række andre områder – vil blive påvirket af de mange evalueringer.

10. PERSPEKTIVERING

Vi har i denne publikation præsenteret en række overvejelser over evaluering af arbejdsmiljøindsatser. Perspektivet er at finde veje til at udvikle en evalueringskultur i arbejdsmiljøindsatsen som ikke bliver til et tidskrævende kontrolregime, men kommer til at tjene til en løbende læring af erfaringerne. En sådan læring giver mulighed for effektivisering af samfundets arbejdsmiljøindsats, hvad enten den gennemføres af de enkelte arbejdsmiljøaktører alene – fx Arbejdstilsynet, Arbejdsmiljørådet, BAR og organisationerne – eller i et samarbejde på tværs af de forskellige aktører.

I udviklingen af en evalueringskultur skal der trækkes på flere forskellige former for viden. Det drejer sig om evaluering som en både teoretisk og praktisk disciplin hvor både viden om arbejdsmiljøet, arbejdsmiljøindsatser og det omkring liggende samfund inddrages. I denne udvikling kan det være en fordel at koordinere evalueringspraksis gennem udarbejdelse af fælles retningslinjer.

Det er imidlertid væsentligt at være opmærksom på at krav om evaluering også i sig selv kommer til at præge indholdet af den indsats man ønsker at iværksætte. Når der skal evalueres, vil der være en risiko for at fokusere på det som relativt let kan evalueres. Det kan betyde at mere komplekse indsatser som har et væsentligt større potentiale for forbedring af arbejdsmiljøet, risikerer at blive nedprioriteret. Det vil formentlig særligt gælde indsatser som involverer mange aktører og måske går på tværs af institutioners resortområder, og hvor man integrerer arbejdsmiljøindsatsen i andre aktiviteter. Her kan potentialet være stort, men det kan også være svært at bevise i en evaluering hvilken effekt netop arbejdsmiljøbidraget har haft på udfaldet. I forbindelse med auditering af certificeret arbejdsmiljøledelse har det fx vist sig at der er risiko for en fokusering på de forhold, som forholdsvist let kan auditeres – at der er styr på papirarbejdet – mens forhold som vanskeligere kan iagttages af auditøren, nedprioriteres (Hohnen & Hasle 2011).

Det er i den sammenhæng væsentligt at stimulere de faglige diskussioner af hvordan god evaluering gennemføres af arbejdsmiljøindsatser. I forlængelse heraf vil en bedre forståelse af de virkningsmekanismer som får virksomhederne til at gennemføre konkrete arbejdsmiljøforbedringer, være et vigtigt element. Diskussioner som dem, der er gennemført på de tre workshops om evaluering, som er blevet organiseret af Arbejdsmiljørådet i samarbejde med CAVI kunne være en måde at stimulere udviklingen af evalueringskulturen. Et øget samarbejde mellem praktikere som organiserer indsatser, og forskere kunne i det hele taget være vigtig for at få sat den viden som allerede er opbygget gennem praksis ind i en generel sammenhæng. Ligesom forskerne gennem sådanne diskussioner kunne få en bedre forståelse for hvor der er behov for øget viden.

I det hele taget handler det om at evalueringer skal anvendes i praksis. Det sker alt for ofte at evalueringer bliver til symbolske øvelser som gennemføres fordi det nu en gang er kravet fra bevillingsgiver, men hvor ingen i sidste ende interesserer sig for resultatet. Derved bliver evalueringerne blot til noget som opfattes som overflødig og som tidsspilde, og værdifulde muligheder for at lære af evalueringerne går dermed tabt.

Kunsten bliver at gennemføre evaluering som kan stimulere til refleksion på flere niveauer. På det udførende niveau har man behov for evalueringsresultater som kan bidrage til en hands-on-læring, hvor aktørerne bliver klogere på hvordan de skal tilrettelægge den næste sammenlignelige indsats, eller skal bruge elementer i den sidste indsats til at integrere i andre indsatstyper. Det kræver typisk kvalitative data med kontekstnærhed som kan bruges til organisatorisk læring. På det bevilgende/politiske niveau er der også behov for refleksion og læring, men her drejer det sig om aggregerede data på et højere abstraktionsniveau, hvor der er en større grad af samfundsmæssig læring som kan

danne grundlag for beslutninger om politikker, der forholder sig til overordnede indsatser. Her er behovet for kvantitative resultater derfor større – samtidig med at de kvantificerbare udfald vanskeligt kan kobles direkte på indsatsen. Det bliver derfor vigtigt at evaluatore og de udførende aktører kan fremlægge overbevisende programteorier som kan forklare, hvorfor bestemte indsatser fører til de ønskede resultater. Her skal trækkes på de direkte resultater af evalueringerne, men også på den forskningsmæssige viden der successivt opbygges om virkemidler og deres mekanismer i virksomhederne.

LITTERATUR

Advice A/S 2008, Evaluering af videncenter for arbejdsmiljø – Centerets virksomhed 2005 til primo 2008, Videnscenter for Arbejdsmiljø.

Arbejdsmiljøgruppen af 1972 1973, Arbejdsmiljø: skader, omkostninger, målsætninger, nærdemokrati, planlægning, Arbejdsmiljøgruppen, 1.

Arbejdsmiljøgruppen af 1972 1974, Arbejdsmiljø: dårlige rygge, stress, høreskader, Arbejdsmiljøgruppen, 2.

Arbejdsmiljøgruppen af 1972 1975, Arbejdsmiljø: grænseværdier, arbejdstiden, meningsfyldt job, udstødningen, Arbejdsmiljøgruppen, 3.

Arbejdsmiljøgruppen af 1972 1976, Arbejdsmiljø: medbestemmelse, arbejdsmiljøtjeneste, uddannelse og forskning, planlægning, Arbejdsmiljøgruppen, 4.

Arbejdstilsynet 2005, Fremtidens Arbejdsmiljø 2010.

Arbejdstilsynet 2007a, 10 år med Rent Arbejdsmiljø 2005, Arbejdstilsynet.

Arbejdstilsynet 2007b, 10 år med Rent Arbejdsmiljø 2005, Arbejdstilsynet.

Arbejdstilsynet 2010a, Evaluering af arbejdsmiljøreform - Slutrapport, Arbejdstilsynet.

Arbejdstilsynet 2010b, Fagligt grundlag for prioritering af arbejdsmiljøindsatsen - Fremtidens arbejdsmiljø, 2020.

Arbejdstilsynet 2011, Evaluering af Arbejdsmiljøfonden, Arbejdstilsynet.

Arbejdstilsynet 2012, Overvågning af arbejdsmiljø og arbejdsmiljøindsats 2006-2010 - overvågningsrapport 2010, Arbejdstilsynet

Beskæftigelsesministeriet 2011, Nye veje til et bedre arbejdsmiljø – Regeringens strategi for arbejdsmiljøindsatsen frem til 2020.

Black, J. & Baldwin, R. 2010, "Really Responsive Risk-Based Regulation", Law & Policy, vol. 32, no. 2, pp. 181-213.

Briner, R. B. & Reynolds, S. 1999, "The costs, benefits, and limitations of organizational level stress interventions", Journal of Organizational Behavior, vol. 20, no. 5, pp. 647-664.

Capacent & Arbejdstilsynet 2007, Arbejdstilsynet Effektevaluering af BAR-projekter, Arbejdstilsynet.

Dahler-Larsen, P. 1998, Den rituelle refleksion: om evaluering i organisationer Odense Universitetsforlag, Odense.

Dahler-Larsen, P. 2003, Selvevalueringens hvide segl Syddansk Universitetsforlag, Odense.

Dahler-Larsen, P. 2010, "Evaluering," in Kvalitative metoder - en grundbog, S. Brinkman & L. Tanggaard, eds., Hans Reitzels Forlag, København, p. 165.

Dahler-Larsen, P. & Krogstrup, H. K. 2004, Nye veje i evaluering System Academic.

Davies et.al 2012, Using Evidence: How Research Can Inform Public Services.

- Eriksen, M. 2008, Den logiske model - et værktøj til at planlægge, gennemføre og evaluere sociale indsatser. KREVI, Århus
- Frick, K. & Wren, J. 2000, "Reviewing Occupational Safety and Health Management," in Systematic Occupational Health and Safety Management, K. Frick, P. L. Jensen, & T. Wilthagen, eds., Pergamon, Oxford, pp. 17-42.
- Gustavsen, B. 1992, Dialogue and development : theory of communication, action research and the restructuring of working life Van Gorcum.
- Hargreaves, D. 1997, "In Defence of Research for Evidence-based Teaching: a rejoinder to Martyn Hammersley", British Educational Research Journal, vol. 23, no. 44.
- Hasle, P., Broberg, O., Hvenegaard, H., & Jensen, U. B. 2000, Arbejdsmiljøindsatsen i 25 år - succes eller fiasko?, CASA, København.
- Hasle, P. & Zwetsloot, G. 2011, "Editorial: Occupational Health and Safety Management Systems: Issues and challenges", Safety Science, vol. 49, no. 7, pp. 961-963.
- Hasle, P., Kvorning, L. V., Rasmussen, C. D. N., Smith, L. H., Flyvholm, M.-A. (2012), A model for design of tailored working environment intervention programmes for small enterprises. Safety and Health at Work, 3, 181-193.
- Health and Safety Executive 2012, Health and Safety Executive - Business Plan 2010/2011, Health and Safety Executive.
- Hohnen, P. & Hasle, P. 2011, "Making work environment auditable - A 'critical case' study of certified occupational health and safety management systems in Denmark", •2 (USE2013) som afholdes i Nelson, New Zealand, februar 2013., vol. 49, no. 7, pp. 1022-1029.
- Joensuu, M., Kivimäki, M., Koskinen, A., Kouvonen, A., Pulkki-Råback, L., Vahtera, J., Virtanen, M., & Väänänen, A. 2012a, "Differential Associations of Job Control Components With Mortality: A Cohort Study, 1986-2005", American Journal of Epidemiology, vol. 175, no. 7, pp. 609-619.
- Joensuu, M., Kivimäki, M., Koskinen, A., Kouvonen, A., Pulkki-Råback, L., Vahtera, J., Virtanen, M., & Väänänen, A. 2012b, "Joensuu et al. Respond to "Structure and Context Matters"", American Journal of Epidemiology, vol. 175, no. 7, pp. 625-626.
- Kjøller, M., Juel, K., & Kamper-Jørgensen, F. 2007, Folkesundhedsrapporten Danmark 2007, Copenhagen: National Institute of Public Health.
- Kristensen, T. S. 2005, "Intervention studies in occupational epidemiology", Occupational and Environmental Medicine, vol. 62, pp. 205-210.
- Kristensen, T. S. & Smith-Hansen, L. 2003, Det udviklende arbejde: Helbred, stress og kvalifikationer. Resultater fra SARA-projektet Frydenlund, København.
- Krogstrup, H. K. 2011, Kampen om evidens - resultatmåling, effektevaluering og evidens Hans Reitzels Forlag, København.
- Lipsky, M. 1971, "Street-Level Bureaucracy and the Analysis of Urban Reform", Urban Affairs Review, vol. 6, no. 4, pp. 391-409.
- Møller, N., Jensen, P. L., & Broberg, O. 1988, Arbejds måder i sikkerhedsgruppen, Arbejdsmiljøfondet, København.

Nielsen, K. A. & Svensson, L. 2006, Action Research and Interactive Research Shaker Publishers, Maastricht.

Nielsen, K. T., Meldgaard Hansen, A., Lund, H. L., & Starheim, L. 2010, Har Arbejdstilsynet et godt øje til arbejdsmiljøet?, Landsorganisationen i Danmark.

NORAD 1999, Logical Framework Approach: handbook for objectives-oriented planning.

Olsen, K., Legg, S., & Hasle, P. 2012, "How to use programme theory to evaluate the effectiveness of schemes designed to improve the work environment in small businesses", *Work: A Journal of Prevention, Assessment and Rehabilitation*, vol. 41, no. 0, pp. 5999-6006.

Parry, G. "Building and Assessing Evidence through the Science of Improvement".

Pawson, R. 2006, Evidence-based policy: A realist perspective Sage.

Pawson, R. & Tilley, N. 1997, Realistic evaluation Sage, Los Angeles, London, New Delhi, Singapore, Washington DC.

Pawson, R., Wong, G., & Owen, L. 2011, "Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy", *American Journal of Evaluation*, vol. 32, no. 4, pp. 518-546.

Pedersen, L. M., Nielsen, K. J., & Kines, P. 2012, "Realistic evaluation as a new way to design and evaluate occupational safety interventions", *Safety Science*, vol. 50, no. 1, pp. 48-54.

Power, M. 1996, "Making things auditable", *Accounting Organizations and Society*, vol. 21, no. 2-3, pp. 289-315.

Rieper, O. & Foss Hansen, H. 2007, Metodedeibatten om Evidens, AKF Forlaget.

Rigsrevisionen 2000, Arbejdstilsynets styring af tilsynsfunktionen , Rigsrevisionen.

Robson, L., Clarke, J., Cullen, K., Bielecky, A., Severin, C., Bigelow, P., Irvin, E., Culyer, A., & Mahood, Q. 2007, "The Effectiveness of Occupational Health and Safety Management Systems: A Systematic Review", *Safety Science*, vol. 45, pp. 329-353.

Rugulies, R. 2012, "Invited Commentary: Structure and Context Matters: The Need to Emphasize "Social" in "Psychosocial Epidemiology"", *American Journal of Epidemiology*, vol. 175, no. 7, pp. 620-624.

Semmer, N. K. 2006, "Job stress interventions and the organization of work", *Scandinavian Journal of Work Environment & Health*, vol. 32, no. 6, pp. 515-527.

Stranddorf, J., Møller, N., & Jensen, P. L. 1992, Arbejds miljøarbejdet i udvikling, Arbejds miljøfondet, København.

Sundhedsstyrelsen 2008, Fokus på alkohol, alkoholpolitik og alkoholproblemer på arbejdspladsen København.

Vedung, E. 1997, Public policy and program evaluation Transaction Publishers, New Brunswick.

Vedung, E. 1998, "Policy instruments: typologies and theories," in Carrots, sticks & sermons - policy instruments and their evaluation, M.-L. Bemelmans-Videc, R. C. Rist, & E. Vedung, eds., Transaction Publishers, New Brunswick, NJ, pp. 21-58.