

”Ingen kan sige sig fri” – en begrebshistorie om konstitutive virkninger

Begrebet ”konstitutive virkninger” beskriver, uden at forudsætte optimistiske antagelser om vedvarende læring og optimering, hvordan evaluering former virkeligheden. Artiklen udlægger begrebets udvikling frem til i dag. Påstanden er, at det nu er påkrævet at reflektere over, hvordan vi alle er involveret i evaluering. Ingen kan sige sig fri for at medvirke til konstitutive virkninger.

Nøgleord: evaluering, brug af evaluering, konsekvenser, konstitutive virkninger

Introduktion

Begrebet om konstitutive virkninger har især de seneste ti år sat et præg på debatten om, hvilke konsekvenser evaluering har i praksis (Dahler-Larsen, 2014). Begrebet markerer en modbevægelse i forhold til den evalueringsoptimisme, der ellers har været en drivkraft i hele feltet (som også omfatter organisatoriske tiltag i retning af evalueringskultur og systemer til evaluativ information, uanset om de kaldes ”evaluering” i det daglige).

Evalueringsoptimismen indebærer, at man indbygger anvendelsen af evaluering i selve definitionen på evaluering, og at denne anvendelse naturligvis er ladet med plusord såsom udvikling og læring. Selvfølgelig arbejder vi ”læringsorienteret”. Vi har allerede foregrebet udfaldet af evalueringsprocessen.

Imidlertid nævner allerede Evert Vedung (1998), at anvendelsen er tilsigtet (og altså netop ikke følger med sikkerhed). Efterhånden som feltet udvikler sig og udvider sig, ser man, at positiv anvendelse af evaluering kræver mere opmærksomhed end et curlingbarn. Der skal konstant være opfølgning, læring, udvikling og opfølgning igen. Alligevel udviser evaluering utilsigtet adfærd (strategisk brug,

legitimerende brug, misbrug osv.). Men bemærk, at begrebet ”utilsigtet” stadig forudsætter begrebet ”tilsigtet” som målestok. Og man danner ikke ny teori ved at tilføje ”u-” til eksisterende tænkning.

Mens tilhængere af evaluering reparerer evaluering med evalueringskultur (daglig massage til de trætte praksismuskler), begynder skeptikere at spekulere over, om nogle af problemerne er indbygget i evalueringskulturen selv. Samtidig hører man stadig oftere, at ”evaluering skal jo give mening”. Bekymringen er tegn på sund skepsis, men forbliver analysesvag, så længe man ikke taler om, hvilken mening evaluering skal give for hvem, hvilken mening evaluering p.t. ikke giver, og hvad konsekvensen er, når den så netop ikke giver mening.

Man er nødt til at opgive
evalueringsoptimismen, slippe
solidariteten med evalueringsfeltets
selvforståelse og søge andre
inspirationskilder for at forstå
evaluerings betydning i samfundet.

Begrebet om konstitutive virkninger tilbyder et sådant alternativ. Ideen med begrebet er at undersøge verden, som den faktisk medproduceres (”konstitueres”) af evaluering. Hvis effekten er faktisk (det vil sige ”gjort”: ”faktisk” kommer af *facio* (latin) = ”jeg gør”), må man forholde sig til den, uanset om den var tilsigtet af andre – for længe siden, et andet sted, i store komplekse systemer, uden at de måske fortalte os direkte, hvad de faktisk tilsigtede, og uden, hvad der måske er endnu værre, at de, som planlagde evaluerings-systemerne, var skarpe nok til at forstå de besynderlige

logikker, de satte i gang. Redskabet til at opdage disse logikker hedder konstitutive virkninger.

I denne artikel vil jeg se på dette begrebs udvikling i sammenhæng med tre forskellige modeller for, hvad evaluering gør ved den praksis, der finder sted i organisationer i det daglige. Med model mener jeg et genkendeligt og typisk tanke-mønster baseret på grundforestillinger, der ofte er implicite. Jeg lægger vægt på de organisatoriske aspekter af disse tankemønstre, for det er organisationer, der gør evaluering magtfuld. Uden at påstå en klar lineær kronologi vil jeg skrive en dynamik frem i begrebets historie ud fra de tre organisationsmodeller. Der er ikke noget "objektivt" og uafviseligt grundlag for mit valg af modeller. De tre modeller er bare den måde, som jeg strukturerer en historie om begrebshistorien på – som et tilbud om, hvordan du som læser kan se en mening og en dynamik i den historie.

Dynamikken ligger ikke mindst i observationer i yderkanten af hver model, som baner vej for nye indsigter. Hvis vi tager de observationer alvorligt, kan den sidste og mest komplekse model medvirke til at opridse nogle dilemmaer og udfordringer, vi står over for i dag. Så hæng på.

MODEL 1

Evaluering som facade

I evalueringsfeltet så man set det som et centralt problem, især i de tidlige år, at der ikke blev fulgt op på de evalueringer, som blev sat i værk. Manglende brug er selvfølgelig et stort problem for evaluering, som påberåber sig at være nyttig.

Manglende brug kan forklares ved, at evaluering blot er facade eller "windowdressing" – svarende til det billede af afkoblede (Meyer og Rowan, 1977) og hykleriske (Brunsson, 1989) organisationer, som er leveret af nyinstitutionel teori. Teorien beskriver, hvordan organisationer lever op til normer og forventninger i deres omgivelser, men de gør det uden overbevisning og med beskyttelse af deres operationelle kerne. Kerneopgaven, hvad enten det nu er undervisning, forskning, pleje eller andet, forbliver så uforstyrret som muligt. Organisationsmodellen "evaluering som facade" forudsiger, at

evaluering vedrører tale, ikke handling. Papirer, ikke praksis.

Modellen er imidlertid blevet mere eller mindre forladt med tiden, idet den har vist sig at have nogle svagheder, og evaluering nogle styrker. Udviklingen starter med, at evaluering beslaglægger tid og ressourcer, altså penge, arbejdstid og opmærksomhed. Sidstnævnte er en af de knappest ressourcer i organisationer og stadig knappere i en digital tidsalder.

I en undersøgelse af resultatindikatorer i offentlig virksomhed blev det fundet, at nogle grupper af offentligt ansatte bruger i gennemsnit 4,44 timer om ugen på at dokumentere resultater (Dahler-Larsen og Pihl-Thingvad, 2014). Det er svært at tro, at man derefter har lige meget tid til kerneopgaven. (Andre bruger mindre tid, men tænker stadig meget på dokumentationen).

Man kan så indvende, at dokumentation i sig selv godt kan være en integreret del af kerneopgaven (se blot journalføring inden for sundhedsområdet). Enig. Men: Det er netop en svaghed ved forestillingen om den afkoblede organisation, at det antages, at en organisation altid beskytter sin operationelle kerne som en fast størrelse. Udviklingstræk ved organisationer de senere år peger på det modsatte. Tager man uddannelses- og forskningsinstitutioner, så synes en bred vifte af funktioner såsom koordinering, kommunikation, strukturændringer, reformer, kvalitetsudvikling, akkreditering og sidst, men absolut ikke mindst ledelse, ledelse og ledelse at beslaglægge stadig flere ressourcer, både hvad angår opmærksomhed, tid og penge.



Nogle institutioner opererer således med et overhead på f.eks. 150% på ekstern kursusvirksomhed. Dette tal er tillige et symbol på en slags mental overhead/overstyring i organisationen.

Derfor bruger en del gode professionelle deres kræfter på at komme over i organisationsudvikling, kvalitetsudvikling, ledelse eller evaluering. Disse funktioner styrkes, mens kerneopgaven tayloriseres og markedsføres. Så hele ideen om en isoleret og beskyttet ”kerneopgave” kollapse. Når evalueringskriterier og pengestrømme tillige sammenkobles, som det er sket i både undervisning og forskning, bliver forbindelsen mellem de eksterne krav og de interne processer således stærkere end forudset af nyinstitutionel teori.

Til at håndtere eksterne krav om kvalitet, evaluering, akkreditering osv. ansætter man konsulenter. De må sandsynliggøre, at de gør en forskel for organisationen (Mathison, 2011). De stiller krav om data og dokumentation, som de selv skal bruge udadtil, men de forsøger også at iscenesætte processer, der skaber forandringer i deres egen organisation.

Med Kjell Arne Røviks (2007) ord lever de op til ”virusteorien” om, at når foregangspersoner først er lukket ind, vil de også begynde at smitte andre. Hvor stærk effekten er, afhænger blandt andet af deres organisatoriske placering, deres ledelsesopbakning, af styrken af deres eksterne alliancer og af den magt, der ligger i at skabe dokumentation, der cirkulerer (Ferraris, 2013; Power, 2019).

Disse tendenser understøttes af en øget mobilisering og systematisering af evaluering i form af begreber som evalueringskapacitet, evalueringspolitik og evalueringskultur. Ikke mindst sidstnævnte er blevet et ofte anvendt begreb, også i officielle styringsdokumenter. Evaluatører er en invasiv art.

Fokuspunktet skubbes fra ”manglende anvendelse” til blandt andet ”konstitutive effekter” og til modeller, der betoner og forklarer sidstnævnte.

MODEL 2

Evaluerings som kolonisering

Model 2 kan man kalde ”evaluering som kolonisering”, fordi den betoner de mange magtkoloniseringer, der gør konstitutive effekter mulige og sandsynlige og til sidst dominerende.

Et ofte citeret eksempel er Wendy N. Espeland og Michael Sauder (2007), der viser, hvordan amerikanske law schools tvinges til at indrette sig efter offentliggjorte ranglister. Ranglister påvirker skolernes omdømme og søgningen til skolerne. Ingen leder eller administrator kan sidde effekten overhørig, også selvom man finder, at ranglisterne er uretfærdige eller bygger på mangelfulde kriterier. Nøglen til denne indflydelse ligger i de alliancer, der skabes mellem journalister, finansieringskilder, bestyrelsesmedlemmer, tidligere og ikke mindst kommende studerende. Med til at fuldbyrde effekten ligger en selvopfyldende profeti, hvor de bedste studerende søger hen til skoler øverst på ranglisten, hvilket pudsigt nok gør det lettere for skolerne øverst på ranglisten at fremstå som gode og desuden prestigefyldte.

Med andre ord medvirker ranglister til at skabe en slags strategisk landskab, som man er henvist til at navigere i. Man kan kun i begrænset omfang sige sig fri. I en given ”valgsituation” må man allerede i øjeblikket tage i betragtning, hvilke konsekvenser et valg har i kraft af de refleksvirkninger, som evaluering kaster tilbage.

Man kunne have håbet på et ”mellemrum” med plads til at vælge til og fra, men det indskrænkes både i tid og rum. Denne effekt oplever vi i praksis som, at ”der var slet ikke tid til at sætte sig ned og drøfte, hvad al denne evaluering skal bruges til.” Det resonerer med Hartmut Rosas (2014) analyse af stigende acceleration i samfundet.

Man når ikke en vurdering af, om en given evaluativ information faktisk er metodisk valid, og om den giver et reelt billede af virkeligheden. Imidlertid er den vurdering ikke en logisk forudsætning for, at den evaluative information har funktionelle konsekvenser. For eksempel taler Elena Esposito og David Stark (2019) om, at en række aggregerede og automatiserede brugerevalueringer (for

eksempel på Tripadvisor, for eksempel af hoteller) sådan set fungerer som vejledning for nye brugere, også uden at man egentlig behøver at tro på, at brugerevalueringer er valide og objektive, og uden at man nogensinde får indsigt i de algoritmer, der bestemmer indsamling, opsummering og præsentation af resultaterne.

Det er foruroligende for den klassiske evaluator, der måtte mene, at først udvikler vi en valid evalueringsform, så aftaler vi dens brug, og så bruger vi den. I praksis ser vi den omvendte rækkefølge. I denne situation er det selvfølgelig højest relevant at sætte nærmere fokus netop på konstitutive virkninger.

De er påvist inden for forskning (de Rijcke og Wouters, 2016), sundhedsvæsen (Kerpershoek, 2016), undervisning (Hult og Edström, 2016), skoleledelse (Hanberger, 2016), global governance (Erkkila og Pironen, 2018), arbejdsmarkedspolitik (Andersen, 2020), filantropi (Williamson, 2017) og socialpolitik (Nordesjö, 2020).

Når konstitutive virkninger skal undersøges empirisk, kan man for eksempel opdele dem i områder som (Dahler-Larsen, 2014):

- Indhold
- Relationer
- Tid
- Verdenssyn

"Indhold" præges for eksempel, når temaer og undervisningsformer indrettes efter, hvordan undervisningen bedømmes evaluermæssigt. "Teaching-to-the-test" er et velkendt fænomen i undervisningen. Men effekten vedrører ikke blot, hvad der sker i den enkelte klasse. Også undervisningsplanlægning, lærebøger og overordnede mål for undervisningen påvirkes. En bestemt, institutionaliseret evalueringsform hører sammen med bestemte vinkler, temaer og kriterier. En enkel, men sigende iagttagelse er for eksempel, at PISA-undersøgelser ikke beskæftiger sig med historie og geografi. Man kan diskutere, om det skyldes, at disse fag ikke er vigtige. Men det er i hvert fald oplagt, at de er mere bundne til nationalitet og sted end for eksempel matematik. Omvendt kan man så sige, at når de ikke belyses i PISA, så bliver

de mindre vigtige. Voila: en konstitutiv effekt. De nationale test omfatter læsning, men ikke skrivning, for sidstnævnte er for svær at bedømme automatisk. Hvilken effekt har det på lærernes og elevernes måde at gå til faget på? Hvilke politikere ville finde sig i, at når de bruger dansk, må det kun være til læsning, aldrig til at sige eller skrive noget?

Under "relationer" indfanges der, hvordan evalueringssystemer påvirker menneskelige relationer. Hvis udsagn regnes for væsentlige? Hvis viden regnes for gældende? Hvem står til regnskab for hvem ud fra kriterier, man ikke selv har valgt? Selve afrapporteringsforholdet er et budskab i sig selv.

Under punktet "relationer" kan man også se på produktion og reproduktion af ulighed. Maria Akselvoll har for eksempel forsket i betydningen af IT i kommunikationen mellem skole og hjem. Mens velstillede forældre bruger deres kompetencer til at stille opfølgende spørgsmål til skolen, bliver forældre, der i forhold til kompetencer og udstyr er svagest stillede i forhold til IT, udfordret af, at de ikke har fulgt "godt nok" med i barnets udvikling i skolen (Akselvoll og Dahler-Larsen, 2021).

Under "tid" kan man beskæftige sig med, hvornår og hvor hurtigt man sætter ind med bestemte evalueringsskriterier, hvilket på sin side giver forskellige tidsrammer for arbejdet. Det kan være et problem, blandt andet fordi forskellige former for arbejde har forskellige indbyggede, stoflige "egentider" for at realisere bestemte effekter (Rosa, 2014). Når evalueringssystemer leverer data online, kan der også ske en acceleration af evalueringens betydning i forhold til andre sociale processer, for eksempel deliberative processer og politiske processer, fordi sidstnævnte simpelthen kræver mere tid. Drøftelsen er allerede overhalet af effekten.

Kategorien "verdenssyn" er opstillet for at vise, at konstitutive effekter i meget omfattende forstand kan betegne, hvad "det gælder om", og hvad "meningen er" med en given indretning af en institution eller et professionelt felt. Når sundhed gøres op i kvalitetsjusterede leveår, som så igen kan omsættes til økonomiske størrelser, så er der allerede sket en hel del med opfattelsen af sundhed.



Når forskning vurderes alene ud fra sine produkter (publikationer), så sidestilles den med produktionsvirksomhed. Allerede Martin Heidegger (1951) var inde på, at viden og videnskab i det moderne samfund ikke blot producerer ”resultater”, men også et helt ”verdensbillede”, en kosmologi.

Evaluering realiserer denne effekt helt ned i detaljen. Men Heidegger var også inde på, at mennesket (i al sin mangfoldighed og sammensathed) ikke har tænkt sig selv med ind i det moderne verdensbillede.

En af de mekanismer, der i stedet ”klemmer” mennesker ind i evalueringens billede, kaldes institutionel lock-in (Osterloh og Frey, 2010). Ved lock-in forstærker en institution, ved hjælp af sanktioner, et handlingsspor, som allerede er valgt. For eksempel kan man betegne personer, som kritiserer evalueringssystemet, for at være modstandere af forandring eller kværlanter, eller endnu bedre: Man kan antyde, at deres kritik skyldes, at de ikke selv er dygtige nok til at opnå gode evalueringssystemer. I tilgift giver udskiftning af personale mere og mere accept af et givet evalueringssystem. Det bliver sværere at modvirke det. Til sidst kan ingen sige sig fri for at blive ramt af dets virkninger. Til gengæld vil færre og færre opleve det som problematisk, hvis ellers den institutionelle lock-in

virker. Man må også tage evalueringssystemet til sig som et faktum, selvom man i princippet er uenig i dets præmisser. Det gælder for eksempel, når evalueringssystemer knyttes til bevillinger og ansættelser.

Mekanismerne til at forstærke et evalueringssystem og dets konstitutive effekter kan være komplekse. Af og til evaluerer man et evalueringssystem ud fra præmisser, som systemet selv har sat, og succes bliver selvopfyldende. Det sker for eksempel, hvis man evaluerer det nationale BFI-system ud fra, at det med tiden har givet en stigning i BFI-point.

Betragt ligeledes indførelsen af et karaktersystem, hvis officielle filosofi er baseret på graden af målopfyldelse. Så sikrer man så at sige i anden potens, at mål og evalueringssystemer indgår i et gensidigt stigningsforhold, hvorved man yderligere styrker de potentielle konstitutive virkninger. Det er ikke blot målene, der styrkes og fastholdes. Det er også målstyringen som filosofi og som menneskeopfattelse. For eksempel oplever vi nu efter nogle år med syvtrinsskalaen, som er baseret på vurdering af, hvor mange fejl og mangler der trækker ned fra den tilsigtede målopfyldelse, at studerende i høj grad har tilegnet sig tankesættet. De argumenterer selv for, at hvis de har opfyldt de opstillede mål, så skal de have høje karakterer. De kan også begynde at opfatte al undervisning, der ikke er målrettet, for tidsspilde.

Når studerende kritiserer lærere for at ”gå uden for pensum”, så er det verdenssyn, der er pløjet ind i karaktersystemet, konstitutivt bestyrket. Som lærer frygter jeg nogle gange, hvordan mine studerende efterfølgende vil forholde sig til fremtidige arbejdsopgaver, hvis de gennem mange år har lært, at det eneste, der skal til, er at leve op til formelt fastsatte målsætninger, hvis disse en dag omfatter deportation af befolkningsgrupper eller afskaffelse af demokratiet. For den installerede mentalitet – ”jeg har krav på at blive evalueret alene ud fra min lydige realisering af på forhånd fastsatte målsætninger og på højeste udmærkelse, hvis jeg ikke begår fejl” – vil ikke udgøre noget bolværk for de studerende, den dag de stilles over for horrible målsætninger.

Observationer under de fire kategorier, indhold, relationer, tid og verdenssyn, er ikke adskilte, men flettet sammen – ligesom replikker, karakterer, plot og skæbner i et drama.

Netop fordi man kan fortælle stærke, sammenhængende narrative om konstitutive virkninger altomfattende greb om virkeligheden, kan der også være grund til at opstille nogle skeptiske indvendinger mod dette i øvrigt brugbare begreb.

De opstillede domæner for konstitutive effekter er så brede og altomfattende, at der ikke rigtig udelukkes noget fra at være en mulig konstitutiv effekt. Det gør det også svært at konkludere i en given empirisk undersøgelse, at der ikke var nogen konstitutive effekter af en given evalueringsform. I praksis kan man måske finde en eller anden konstitutiv effekt, hvis man leder efter den. Vi har ikke at gøre med et eksakt empirisk begreb, men med et sensitizing concept, som først fyldes ud med materiale, når nogen gør det gældende på en bestemt måde i en konkret undersøgelse. Begrebet lever og ånder sammen med en problematisering.

Her er det vigtigt at være omhyggelig. Det er ikke nok at fastslå, at virkningen er utilsigtet. For hvorfor være solidarisk med nogle bagvedliggende intentioner, som måske både er politisk problematiske og desuden sværere at kortlægge? Man kan jo heller ikke automatisk og på forhånd vide, at konstitutive virkninger er "dårlige" eller forkastelige. For eksempel peger Emiel Kerpershoek et al. (2016) på, at sundhedspersonale laver fiksfaktserier med registreringer i diverse dokumentationssystemer – ikke til egen fordel, men til fordel for patienten, for eksempel ved at give plads og rum til nogle yderligere undersøgelser i situationer præget af usikkerhed.

Indlysende nok må man jo medgive, at det ikke er alle former for overvågning, registrering og kontrol, som per definition har dårlige virkninger. Vi synes jo ofte som professionelle, at der skal være en vis kontrol med alle de professioner, som vi ikke selv tilhører. At en konstitutiv effekt for eksempel kan indebære en form for magtudøvelse alene ved at sætte bestemte afrapporteringsrelationer i værk, er jo ikke nødvendigvis dårligt, medmindre det går ud over os selv. Vi kan drøfte i et demokrati, hvilke

afrapporteringsrelationer, der skal være, klart nok. Men det er svært at forestille sig et demokrati, hvor ingen har krav på afrapporteringer fra andre. Med andre ord er en interesse i at pege på bestemte konstitutive virkninger forbundet med en eller anden forestilling om, hvad der er ret og rimeligt, og hvordan samspillet mellem mennesker i et samfund bør fungere.

Man må derfor erkende, at der til enhver erkendelse af en konstitutiv effekt hører en potentiel diskussion af erkendelsesinteressen og af, hvordan forskellige interessenter måtte have det med den pågældende effekt og hvilke demokratiske implikationer den måtte have. Det gode ved begrebet "konstitutive virkninger" er til gengæld, at det ikke stiller sig tilfreds med, hvad forskellige interessenter erklærer som det stipulerede formål med evaluering. De må også forholde sig til, hvad evaluering faktisk har af effekter.

Bemærk også, at evaluering som praktisk og politisk virkemiddel i dag de facto har forladt evalueringsfeltets oprindelige ide om, at evaluering skulle foretages netop og kun netop i de situationer, hvor man kunne sandsynliggøre en konkret og positiv anvendelse såsom "læring". Det krav stiller man ikke længere. Når man forlanger evalueringskultur, er den nærmest blevet et gode i sig selv. Når man argumenterer for, at alle elever skal have en elevplan, er det ikke, fordi alle kan have en påvist gevinst af den. I nogle elevplaner står der jo blot, for eksempel i dansk: "Du arbejder godt. Men du kan blive bedre til mundtlig og skriftlig dansk." Det forestiller de færreste sig vel har nogen konkret læringseffekt. Men det behøver det heller ikke at have. Elevplanen er nemlig blevet en rettighed. Evaluering i praksis er spredt så meget ud, at den ikke længere retfærdiggøres ved evalueringslitteraturens interne normer.

Så man udfordres på, hvilke normer man så skal lægge til grund for betragtninger af konstitutive effekter. Samtidig er også opsporingen af netop "effekt" en udfordring. Forskningen har i de senere år stadig tydeligere vist, at produktionen af konstitutive effekter er kompleks og interaktiv og dermed ikke blot et resultat af et givet evalueringsystem i sig selv.

MODEL 3

Evalueringsystemer som skrøbelig forandring af verden

Jeg har lånt ordet "skrøbelig" fra videnssociologen Nico Stehr, der taler om det moderne videnssamfund som skrøbeligt, fordi det konstant er påvirket af ny viden som produktionskraft. (I skrivende stund holder jeg øje med dagens smittetal for at finde ud af, om vi snart må lukke samfundet lidt op igen).

Det er samfundets skrøbelighed, der muliggør konstitutive virkninger af evaluering, men samtidig – og det er så pointen i model 3 – er selve viden, evaluering og dermed konstitutive virkninger også selv skrøbelige. Sagt mere konkret afhænger evalueringssystemers virkninger af en lang række handlinger og valg af mange instanser i meget komplekse systemer. For eksempel kan samme evalueringssystem have to ganske forskellige effekter på offentlige og privatiserede arbejdspladser (Dahler-Larsen og Kjær Foged, 2018). Ligeledes har samme nationale evalueringssystem (BFI) ganske forskellige virkninger på forskellige universiteter, hvor ledelsen via lokale ledelsessystemer enten styrker eller svækker virkningen af det nationale system (Lind, 2019; Mouritzen et al., 2018).

Tillige kan professionelle, som selv er genstand for evaluering, bruge evaluering mere eller mindre aktivt i egen interesse (Fochler og de Rijcke, 2017). De oplever måske ikke dette som et selvstændigt valg, givet det strategiske landskab, som defineres af evaluering- og finansieringssystemer. For eksempel har jeg interviewet gymnasielærere, som var meget skeptiske over for elevtilfredshedsvurderinger som kvalitetsmål. Men omvendt kunne de ikke dy sig for at bruge positive tilfredshedsvurderinger som markedsføringsmateriale over for nye elever, hvis et fald i elevtal ellers ville medføre fyring af to lærere efter næste sommerferie (Dahler-Larsen og Pihl-Thingvad, 2014). Mere ressourcestærke professionelle kan opleve at have indflydelse på evalueringssystemers design og dermed indirekte på deres effekter (Dambrin og Robson, 2011), også selvom enhver sådan indflydelse kan være tvetydig (Jensen, 2009).

Tillige bevæger de fleste professionelle sig i dag i komplekse systemer, hvor mange evalueringssystemer, delvist overlappende, delvist konfliktende, forsøger at præge hverdagen (Pollock, 2018). Det kan give et vist spillerum for ageren, om end stadig under usikkerhed. Nogle oplever også at være løbende involveret i forhandlinger om konsekvenser af evalueringssystemer, som kun giver utilstrækkelige og partielle billeder af virkeligheden, hvilket kun øger betydningen af forhandlinger, hiven og halen i det daglige (Dambrin og Robson, 2011).

Evalueringsystemer kan have ganske reelle konsekvenser, selvom de involverede aktører udmærket er klar over, at de ikke giver et dækkende billede af virkeligheden. Hvad eksperter sagde i gamle dage om ikkevalide evalueringer, nemlig at "de kan ikke bruges", er i dag med fuld kraft overhalet af virkeligheden.

Der er tilstrækkeligt mange stærke mekanismer til at omgå validitet for i stedet at gå direkte til konsekvenser. En medvirkende forklaring er, at i en verden med mange virkelighedsopfattelser vil "validitet" og "reliabilitet" alligevel være mere eller mindre subjektivt defineret og kan derfor ikke være en generel forudsætning for praktisk handling. I en nyere undersøgelse har jeg fundet, at arbejdsmiljørepræsentanters vurdering af, om en arbejdspladsvurdering (APV) giver et reelt billede af arbejdsmiljøet, ikke er korreleret med, om samme APV bliver brugt som dokumentation på arbejdspladsen. Om man selv synes, at billedet er dækkende, og om andre overbevises om, at billedet er troværdigt, er således to adskilte funktioner. Begge er i øvrigt korreleret med, at der faktisk foretages handling som opfølgning på APV'en. Så der er flere veje til handling. Ikke alle går gennem reliabel evaluering.

Tværtimod virker en hel del evaluering altså netop, ganske uafhængigt af om vi synes, at den er metodisk god og dækkende for vores værdier. Desuagtet kan evaluering etableres som fakticitet uden at afhænge af reliabilitet og kulturel kongruens.

Michael Power (2019) har udviklet en teoretisk model for, hvordan man kan forstå etableringen af en sådan fakticitet trin for trin. Med udgangspunkt

i de spor, som evalueringsdokumentation sætter sig, kan effekten blive institutionaliseret, når dokumentationen indgår i organisatoriske rutiner, når interne evalueringsfunktioner bliver magtfulde, når disse opnår alliancer med eksterne parter osv. Der er naturligvis til dels tale om en påvirkning af tankemåder, diskurser og mentalitet. Men de konfigurationer, der skaber effekter ud af evaluering, virker også gennem dokumenter, netværker, medier og ressourcestrømme. Og de kan som sagt virke uden først at være "godkendt" af bevidstheden og normerne.

Tanken her minder om en pointe hos Slavoj Žižek (2008). Han taler om, at en "bevidstgørelse" ikke hjælper over for ideologi. Vi tager gerne retorisk afstand fra forkastelige ideologier og undertrykkende systemer. Samtidig er vores handlinger indflettet i selvsamme. Hvilket vi godt ved. Det er vores handlinger, der er ideologiske, ikke vores bevidsthed. En vending af vendingen i denne artikels titel lyder derfor, at man ikke kan sige sig fri for konstitutive effekter.

Pointen kan bruges til at understrege, at i vore dages komplekse organisatoriske systemer virker evalueringer ikke altid "imod" os. De virker også "gennem" os, gennem vores handlinger. Også selvom vi ganske godt ved, at evalueringssystemer er ufuldstændige, imperfekte, hverken valide eller reliable osv. Det er altså værre fat, end at evalueringsoptimismen stadig har tag i os. Selv når vi opgiver evalueringsoptimismen i vores tanker, så har de konstitutive virkninger allerede tag i vores handlinger. Med den tredje model, "evaluering som skrøbelig forandring af verden", kan vi skitsere en slags "løsning" på det problem, Heidegger skitserede. Mennesker passer faktisk ind i evalueringernes verdensbillede. Mennesker er aktivt medvirkende, men altså stadig uden et gennemtænkt forhold mellem menneskebilledet og verdensbilledet.

Hvis man nu erkender, at konstitutive effekter af evaluering er reelle nok, men samtidig skrøbelige, fordi de afhænger af tusind aktørers handlinger i komplekse systemer, kan man stille nogle relevante spørgsmål til nogle af disse aktører. Det er jo sjovt nok os selv, blot i forskellige roller. Til konstruktører af evaluering kan vi spørge, om de

ikke godt vil tage noget mere praktisk medansvar for de konstitutive virkninger af de systemer, de bygger op. Det må være tiden til at betragte det som et forældet svar, at "det har selvfølgelig aldrig været intentionen, at systemet skulle have disse virkninger" (som en ministeriel repræsentant sagde, da vedkommende hørte, at BFI-optællinger på nogle universiteter bliver brugt på individniveau i et for mig at se bemærkelsesværdigt forsøg på at sige sig fri for sit ansvar).

Man kunne som konstruktør af evalueringssystemer være mere nysgerrig på, hvilke virkninger et sådant system i praksis har, og dernæst forholde sig i handling til det, man opdager. Det gælder også politikere. Når man for eksempel politisk vælger, at en institutions evne til at tiltrække private forskningsmidler indgår i tildelingen af offentlige ditto, så forstærker den politik jo de konstitutive virkninger af private aktørers indflydelse på forskningen gennem en mekanisme, hvor de jo også bestemmer over de offentlige midlers fordeling. Man kunne i stedet have valgt, at de offentlige midler skulle have andre virkninger, måske kompenserende virkninger i forhold til de private.

Også ledere må kunne udfordres på, om de dæmper eller forstærker evalueringssystemer internt i deres organisationer. Empirisk forskning tyder i hvert fald på, at de har en medvirkende rolle at spille.

Det har vi også som professionelle. Hvis professionalisme er at udøve dømmekraft inden for en ramme, så må vores medvirken til realisering af konstitutive virkninger af evalueringssystemer også kunne bedømmes og debatteres.

Som nævnt tyder forskning på, at konstitutive virkninger realiseres gennem handling, de påføres ikke kun. Som minimum kan man starte med at undlade at gange dem op. For eksempel kan ældre forskere gange virkningen af bibliometriske indikatorer op, når de fortæller unge forskere, at det eneste, der tæller, er publicering i ganske bestemte tidsskrifter. Det gælder i hvert fald, hvis den opskrift faktisk er strammere fortalt, end hvad der faktisk sker i det pågældende forskningsmiljø. En gang imellem kan vi som forskere skrive i tidsskrifter, hvor vi tror, der foregår noget interessant,

fordi folk læser vores artikler i stedet for blot at tælle dem.

Og hvornår, hvorfor og hvordan skal vi sommetider indrette undervisningen, så den ikke svarer til de studerendes forventninger og dermed spolerer tilfredshedsmålingerne, for simpelthen derved at sætte nye logikker i gang? Et farligt, underliggende spørgsmål er imidlertid det, som stilles af Bénédicte Vidaillet (2021): Er vi afskåret fra sådanne eksperimenter, simpelthen fordi vi er blevet for psykologisk afhængige af gode evalueringssystemer som en vej til anerkendelse?

Konstitutive virkninger og demokrati?

Når vi alle på forskellige måder medvirker til at skabe konstitutive virkninger, kan man selvfølgelig rejse spørgsmålet, om disse virkninger kunne gøres til genstand – ikke blot for individuelle krumspring (og psykologisk tolkning), men for politisk tolkning og fælles demokratisk håndtering? Det er et vigtigt og presserende spørgsmål. Men vanskeligt, for, som vi har set, er konstitutive virkninger vanskelige at afgrænse og vanskelige at påvise. De er relative til betragte-

ren, relative til normer og værdier, og vi er ofte for meget viklet ind i dem til at erkende dem.

På den måde forbinder begrebet konstitutive virkninger på ny evaluering med en drøftelse og belysning af værdier, som var en oprindelig ambition i evaluering, men som på en eller anden måde er gået tabt i mellemtiden.

En interessant udfordring er, hvordan man skaber demokratiske institutioner, hvor feedback om konstitutive virkninger på ny kan nå frem til beslutningstagere, og hvor disse virkninger ikke blot afvises blankt som ”utilsigtede”, men i stedet tages alvorligt som nogle konsekvenser, der bør håndteres. De er resultater af den fælles skæbne, som kommer af, at mange aktører er viklet ind i de samme evalueringssystemer.

Et mareridtsscenario er, at mange egentlig er ligeglade med, hvordan evalueringer egentlig virker, bare man selv redder sig en position, hvor man konstruerer evaluering, men ikke oplever dens effekter. Et beslægtet mareridtsscenario er en situation, hvor et topstyret evalueringssystem skaber en række konstitutive virkninger, som alle i systemet kan se, men som dem i toppen ikke behøver at forholde sig til. Den første position hedder specialist. Den anden politiker.

For at bryde dødvandet tænker jeg personligt en del over, om det hidtidige forløb omkring de nationale test kan tjene som en slags model eller i det mindste som et pilotprojekt. Der har været plads til en omfattende debat og en grundig evaluering. Der har været en rådgivningsgruppe med mange interessenter, som efter fælles drøftelser har givet input til politikerne om, hvordan fremtidens test kunne se ud, hvis man ønsker at opretholde nationale test.

Det lykkedes faktisk til sidst forligskredsen at ændre det nationale testsystem. Men alle æg skal selvfølgelig ikke lægges i den kurv. Når det kommer til skabelse af konstitutive virkninger, så gælder det i mere end én forstand, at ingen kan sige sig fri.



Litteratur

- Akselvoll, M. & Dahler-Larsen, P. (2021). *Evaluation people and real people in home-school cooperation*. I Dahler-Larsen, P. (red.). *A Research Agenda for Evaluation*. Northampton, MA: Edward Elgar Publishing.
- Andersen, N.A. (2020). *Fra opkvalificeringseffekt til motivationseffekt: Effektevalueringers rammesætning af den aktive arbejdsmarkedspolitik*. I Klindt, M.P., Rasmussen, S. & Jørgensen, H. (red.). *Aktiv Arbejdsmarkedspolitik*. København: Djøfs Forlag.
- Brunsson, N. (1989). *The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations*. Chichester: John Wiley.
- Dahler-Larsen, P. (2014). *Constitutive effects of performance indicators: Getting beyond unintended consequences*. *Public Management Review*, 16(7), 969-986.
- Dahler-Larsen, P. & Kjær Foged, S. (2018). *Konkurrence svækker jobtilfredshed*. *Politiken*, 25. august.
- Dahler-Larsen, P. & Pihl-Thingvad, S. (2014). *Resultatmålinger og stress*. Odense: Syddansk Universitetsforlag.
- Dambrin, C. & Robson, K. (2011). *Tracing performance in the pharmaceutical industry: Ambivalence, opacity, and the performativity of flawed measures*. *Accounting, Organizations and Society*, 36, 428-455.
- de Rijcke, S., Wouters, P.F., Rushforth, A.D., Franssen, T.P. & Hammarfelt, B. (2016). *Evaluation practices and effects of indicator use – a literature review*. *Research Evaluation*, 25(2), 161-169.
- Erkkilä, T. & Piironen, O. (2018). *Rankings and Global Knowledge Governance*. Basingstoke: Palgrave Macmillan.
- Espeland, W. & Sauder, M. (2007). *Rankings and reactivity: How public measures recreate social worlds*. *American Journal of Sociology* 113(1), 1-40.
- Esposito, E. & D. Stark (2019). *What's observed in a rating? Rankings as orientation in the face of uncertainty*. *Theory Culture and Society*, 36(4), 3-26.
- Ferraris, M. (2013). *Documentality*. New York, NY: Fordham University Press.
- Fochler, M. & de Rijcke, S. (2017). *Implicated in the indicator game? An experimental debate*. *Engaging Science, Technology, and Society*, 3, 21-40.
- Hanberger, A. (2016). *Evaluation in local school governance: A framework for analysis*. *Education Inquiry*, 7(3), 217-236.
- Heidegger, M. (1951). *The age of the world view*. *Boundary 2*, 4(2), 341-355.
- Hult, A. & Edström, C. (2016). *Teacher ambivalence towards school evaluation: promoting and ruining teacher professionalism*. *Education Inquiry* 7(3), 305-325.
- Jensen, C.B. (2009). *Making lists, enlisting scientists: The bibliometric indicator, uncertainty and emergent agency*. *Science Studies*, 24(2), 64-84.
- Kerpershoek, E., Groenleer, M. & de Bruijn, H. (2016). *Unintended responses to performance management in Dutch hospital care: Bringing together the managerial and professional perspectives*. *Public Management Review*, 18(3), 417-436.
- Lind, J.K. (2019). *The missing link: How university managers mediate the impact of a performance-based research funding system*. *Research Evaluation*, 28(1), 84-93.
- Mathison, S. (2011). *Internal evaluation – historically speaking*. *New Directions for Evaluation*, 132, 13-23.
- Meyer, J.W. & Rowan, B. (1977). *Institutionalized organizations: Formal structure as myth and ceremony*. *American Journal of Sociology*, 83, 340-363.
- Mouritzen, P.E., Opstrup, N. & Pedersen, P.B. (2018). *En fremmed kommer til byen. Ti år med den bibliometriske forskningsindikator*. Odense: Syddansk Universitetsforlag.
- Nordesjö, K. (2020). *The constitutive effects of social investment evaluation*. *Evaluation*, 27(2), 210-228.
- Osterloh, M. & Frey, B.S. (2010). *Academic Rankings Between the 'Republic of Science' and 'New Public Management'* (arbejdsrapport). Zürich: CREMA (Center for Research in Management, Economics and the Arts).
- Pollock, N., D'Adderio, L., Williams, R. & Leforestier, L. (2018). *Conforming or transforming? How organizations respond to multiple rankings*. *Accounting, Organizations and Society*, 64, 55-68.
- Power, M. (2019). *Modelling the micro-foundations of the audit society: Organizations and the logic of the audit trail*. *Academy of Management Review*, 46(1), 6-32.
- Rosa, H. (2014). *Fremmedgørelse og acceleration*. København: Hans Reitzels Forlag.
- Røvik, K.A. (2007). *Trender og translasjoner: ideer som former det 21. århundrets organisasjon*. Oslo: Universitetsforlaget.
- Vedung, E. (1998). *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
- Vidaillet, B. (2021) *Victims or accomplices? Our strange appetite for evaluation*. I Dahler-Larsen, P. (red.). *A Research Agenda for Evaluation*. Northampton, MA: Edward Elgar Publishing.
- Williamson, A., Leat, D. & Scaife, W. (2017). *Narratives of performance measurement in philanthropic foundations*. *Voluntary Sector Review*, 8(3), 273-298.
- Zizek, S. (2008). *Ideologiens sublime objekt*. København: Hans Reitzel.