

Essay

Evaluering eller demokrati?

Af Peter Dahler-Larsen, Professor,
Institut for Statskundskab, Københavns Universitet

Magtudredningen 2.0



Evalueringssamfundet

Evaluering var egentlig et vidensbaseret redskab til at undersøge, om en given offentlig politik virker efter hensigten. I dag er evaluering dette og meget mere, både inden for og uden for den officielle styrekæde. Mange taler om en evalueringebølge og endda et evalueringssamfund.

Evaluering medvirker til en fremadskridende modernitet, hvor stadig større områder af livet bliver genstand for refleksivitet (Beck, 1992), men refleksiviteten er ikke ubegrænset. Den reguleres via politik og organisation (Power, 1997), og den er præget af samfundsmæssige forestillinger. Hvad evaluering er og kan, kommer til udtryk i meget forskellige *evaluation imaginaries* (Dahler-Larsen, 2012).

For nogle begrundes evaluering i forhåbninger om læring, udvikling og kvalitetsløft. Et ministerium er stadig officielt fortalere for en "stærket evalueringkultur." I den offentlige debat kritiseres evaluering for at øge offentligt ansattes dokumentationsbyrde og skabe mistrivsel blandt børn og unge. Det vidner om, hvor vidt udbredt evaluering er, og hvor vanskeligt det er at skille evaluering fra alt det, som den er viklet ind i.

Hvad taler vi om, når vi taler om evaluering?

Hos den svenske statskundskaber Evert Vedung (Vedung, 1998) defineres evaluering som en systematisk bedømmelse af indhold, udførelse og resultater af offentlig politik. Bedømmelsen tiltænkes at spille en rolle i praktiske handlingssituationer.

Systematisk er et nøgleord. Evalueringsmetoden forventes at være eksplicit og transparent. Den skal kunne efterprøves. Vurderingen forudsætter naturligvis værdier. Så man må være klar og skarp i sin bestemmelse af grundlaget for disse værdier.

Offentlig er et andet nøgleord. Evaluering handler om noget, som har betydning for et samfund eller et politisk fællesskab. Det er meningen, at evaluering skal bidrage til en bevidst og vidensbaseret håndtering af fælles anliggender. Endelig sigter evaluering evaluering (i modsætning til forskning i almindelighed) mod at gøre en praktisk forskel. Den udføres for, at nogen skal gøre noget med den.

Inden for disse hjørnesteene i Vedungs begrebsbestemmelse forekommer store variationer i evalueringers værdier, former og anvendelser.

En del beslægtede fænomener, såsom akkreditering, auditering, kvalitetsmåling og performance indikatorer falder inden for definitionen, uden at de dog i daglig tale kategoriseres som evaluering. I praksis bruges ordet *evaluering* både positivt og negativt også uden den systematik og begrebslige klarhed, som Vedung tilstræber.

De store temaer

I forhold debatten om demokrati og magt er nogle få temaer centrale:

- Evalueringers tilrettelæggelse. Hvordan går det til at bestemme formål, design, værdier og metoder i evaluering, og hvem er med?
- Evalueringsprocesser. Hvordan forløber evalueringsprocesser, hvem har muligheder for deltagelse på hvilke præmisser, og hvem har adgang til evalueringresultater?
- Konsekvenser. Hvilke konsekvenser drages af evaluering og af hvem?

Rundt om evalueringers tilrettelæggelse, processer og konsekvenser findes strukturelle og institutionelle rammebetingelser, som er under forandring.

Strukturforvandlinger i evalueringslandskabet

Udviklingstræk 1: Oprindeligt tog evaluering form som en forskningsbaseret undersøgelse af en offentlig politik. Spørgsmålet var, om politikken virkede i forhold til sine mål.

I de senere år er evaluering også blevet led i en politik. Så indgår evaluering ofte i en styringspakke. En uddannelsesinstitution under akkreditering forventes for eksempel at inddrage de studerende i løbende, systematisk evaluering af undervisningen. Under overskriften "evalueringkultur" har man som led i skolepolitikken indført kvalitetsrapporter, test og elevplaner. Denne institutionelle gennemlysning af skolen påvirker måden, som evaluering tænkes og anvendes på. Der kommer mindre fokus på, om en vedtaget national politik er effektiv, men mere fokus på den enkelte institutions performance.

Udviklingstræk 2: Oprindeligt mente man med "en evaluering" en proces afgrænset i tid og rum og et produkt i form af en evalueringsrapport på papir.

I dag spredes evaluering også online i tid og rum. Evaluering bliver vedvarende, gentaget med faste kadencer og rutinepræget. Det giver forudsigelighed. Når mennesker forventer bestemte former for evaluering, indretter de deres adfærd derefter. Når evaluering knyttes til konsekvenser for en institutions omdømme eller finansiering, kan det afstedkomme en omfattende organisatorisk, foregribende indsats for at sikre vedvarende opnåelse af de ønskede evalueringresultater.

Tillige udvides de sociale og politiske rum, hvor fælles evaluering finder sted. Inden for et land kan alle skoler og alle kommuner sammenlignes. Når Danmark deltager i PISA undersøgelser, sker det efter et fælles koncept sammen med 79 andre lande. Med større evalueringzoner og ensartet evaluering følger ofte et element af konkurrence.

Udviklingstræk 3: Evaluering vandrer således "opad og udad" til store nationale og internationale arenaer, men også "indad og nedad" til institutioner og organisationer. De opretter et kvalitetskontor med specialister, som kan svare på eksterne evalueringskrav, men også benytte lejligheden til at give evalueringsarbejdet et ekstra nøk (Mathison, 1991). Selv om evaluering i en organisation måske først er drevet af ydre grunde, kan den efterfølgende få reel indflydelse på flere niveauer (Power, 2019). Evaluering er ikke længere først og fremmest udført af forskere, men af mange aktører, der tjener en mangfoldighed af organisatoriske formål, værdier og interesser.

Udvalgte forskningsresultater

Er evalueringebølgen ligeligt fordelt?

Herhjemme slog evalueringebølgen uens igennem på forskellige sektorområder (Hansen, 2003). Der evalueres intenst inden for områder som undervisning, forskning, socialt arbejde, sundhed, beskæftigelse og udviklingshjælp, mens områder som jura, forsvar, kirke og kongehuset ikke oplever samme evalueringspres.

Større begivenheder som reformer i forvaltningen evalueres meget uens, eventuelt kun internt, eller slet ikke (Breidahl et al, 2015). Det samme gælder Danmarks deltagelse i krige.

Hvordan hænger evaluering sammen med problemløsning?

På nogle indsatsområder findes stabile, lovpligtige og obligatoriske evalueringssystemer. Et eksempel er arbejdspladsvurderinger (APV), der skal gennemføres på alle arbejdspladser mindst hvert tredje år, i øvrigt efter krav fra EU. APV fører i et vist omfang til konkrete indsatser, der sigter mod forbedringer. Disse indsatser lader sig forudsige ud fra ledelsens støtte til APV, medarbejdernes inddragelse og medindflydelse, samt en APV, der opfattes som troværdig, hvilket for eksempel fremmes ved både at bruge kvantitativ og kvalitativ metode i APV-arbejdet. Men den opfølgende brug af APV er ikke korreleret med, om arbejdspladsens arbejdsmiljøproblemer opfattes som alvorlige. Tillige stilles der ofte mange spørgsmål om problemer i arbejdsmiljøet, som faktisk rammer meget få personer, til gengæld for færre spørgsmål om meget udbredte problemer. Som element i offentlig politik formår APV således ikke at sætte mest ind der, hvor problemerne er størst (Dahler-Larsen et al., 2021). At gøre evaluering obligatorisk medfører ikke nødvendigvis målrettet løsning af problemer (Bromley og Powell, 2012).

Konsekvenser af evaluering

Internationale indikatorer og evalueringssystemer øver generelt en stor indflydelse på, hvordan problemer og tiltag konciperes (Rottenburg, 2015). For eksempel har PISA-undersøgelser sat dagsordenen for landenes skolepolitik (Grek, 2009).

På skoleområdet i Danmark har en evaluering fra VIVE vist, at 36% af lærerne i faget dansk målretter deres undervisning for at gøre resultaterne af de nationale test så gode som muligt i forhold til landsgennemsnittet. 28% af de adspurgte nedprioriterer de områder af faget, der ikke testes i (VIVE, 2020: 52).

Ansatte i den offentlige sektor reagerer på evaluering på meget forskellig måde. Et internationalt review (de Rijcke et al., 2016) viser, at forskningsevaluering kan have en betragtelig indflydelse på forskeres valg af emne, publikationsform og

udgivelsessprog. Men effekten afhænger af, hvordan det enkelte evalueringssystem er designet (Sivertsen, 2017).

Der er store forskelle i, hvor meget mental energi, offentligt ansatte bruger på at indrette deres arbejde efter evalueringskriterierne. Et internationalt forskningsreview tyder på udbredte konsekvenser for kognitiv dissonans, frustrationer og motivationssvigt blandt nogle offentligt ansatte (Dahler-Larsen og Møller, under udg.) Der er også store variationer i, hvor meget tid forskellige grupper af ansatte bruger på dokumentation. Det er ikke de samme personalegrupper, som oplever at evaluering beslaglægger arbejdstid, og som oplever, at evaluering udgør en udfordring for deres værdier og mentale kræfter (Dahler-Larsen og Pihl-Thingvad, 2014).

Evaluering og ulighed

Et forskningsreview inden for områder som skole, beskæftigelse og socialområdet har peget på, at evaluering baseret på resultatindikatorer kan have negative konsekvenser for de svageste brugergrupper (Møller, Andersen og Iversen, 2026). De risikerer at blive sorteret fra eller sendt videre i systemet. Forældre, som er dårligt stillede i forhold til IT, oplever også at få mindre udbytte af kommunikation med skolen om deres børns skoleresultater (Akselvoll og Dahler-Larsen, 2021).

Den offentlige indsigt i evalueringresultater er ikke jævnt fordelt

Går man på restaurant, kan man allerede på indgangsdøren se smileys, der angiver resultatet af myndighedernes vurdering af restaurantens fødevarerhåndtering. Hvis man søger ansættelse på restauranten, har man imidlertid ikke adgang til resultaterne af restaurantens arbejdspladsvurdering. Tilsvarende på universitetsområdet. Her har nutidige og potentielle studerende online adgang til spørgeskemadata om studiemiljøet, mens data om arbejdsmiljøet kun er tilgængeligt for allerede ansatte og ledelse. Det kan virke slående, fordi der i begge tilfælde er tale om spørgeskemadata, udført af samme eksterne firma, og opstillet i samme grafiske format. Et flertal af danskerne ønsker i øvrigt offentlig adgang også til arbejdspladsvurderinger (Dahler-Larsen, 2021).

Et indsatsområde kan have en betragtelig samfundsmæssig betydning uden at være genstand for et formelt evalueringssystem. Det gælder for eksempel filantropiske fonde, som årligt donerer hvad der svarer til 1% af det danske BNP til områder som forskning, sundhed, og kultur. Fondenes evalueringsskemaer er uens og kun i begrænset omfang genstand for offentlig indsigt.

Generelt er offentlig tilgængelighed af evaluativ information et demokratisk gode (Chelimsky, 2006), men andre hensyn kan trække i modsat retning. Offentliggjorte ranglister kan forstærke negative virkninger af evalueringseresultater (Espeland og Sauder, 2007). I Danmark har man politisk besluttet, at skolers testresultater ikke lægges offentligt frem.

Værdier og positioner i forskning og i debat

I studier af evaluering kan vinklingen betyde meget. Ikke mindst casestudier afspejler forskellige *evaluation imaginaries*. Nogle case studier fremstår som optimistiske illustrationer af, hvordan evaluering bidrager til læring og forbedring (Ejler et al., 2008). I andre tilfælde skildrer kritiske casestudier en magtudøvelse, hvor evaluering konspirerer ideologisk med New Public Management over for ansatte.

En tredje type af casestudier er brudt frem de senere år. De viser mange aktører implicerede i frembringelsen af konsekvenser af evaluering. For eksempel har nogle universitetsledelser forstærket de lokale virkninger af et nationalt forskningsevalueringssystem (Lind, 2019). Nogle forskere benytter også selv indikatorer, som de ellers er uenige i, så længe disse indikatorer stiller forskerne i et positivt lys (Fochler og de Rijcke, 2017). Disse effekter er uens på forskellige områder, alt efter forskernes prestige og sociale kapital og forskningsområdets omdømme (Rowlands og Wright, 2021; Aagaard, 2015). At lokale aktører er impliceret i udmøntningen af effekter af evaluering, er ikke ensbetydende med, at konsekvenserne er magtfrie eller ligeligt fordelt.

Når professionelle minglerer med evaluering, skyldes det ikke altid egeninteresse. For eksempel kan en kreativ eller fleksibel brug af patientkategorisering på sundhedsområdet give ekstra tid og ressourcer til patienter med lidelser, som er svære

at diagnosticere (Kerpershoek et al., 2016). Inden for forskning bruger nogle af de højest rangerende forskere deres gode citationstal til at købe sig forskningsfrihed på innovative og risikable forskningsområder, hvor evalueringresultaterne er usikre (Young et al., 2017).

Nuancerede casestudier viser således, hvordan effekter af evaluering skabes i et samspil mellem aktører i et tvetydigt terræn. Terrænets kortlægning er ikke uafhængig af perspektivet. De mest nuancerede resultater kommer frem, hvis man ikke abonnerer på firkantede *evaluation imaginaries*.

Et eksempel på diskussionernes sammensathed

International forskning peger på alvorlige problemer i SET (Student Evaluation of Teaching) i uddannelsessystemet. Rutinemæssige, spørgeskemabaserede studenterevalueringer er i ringe grad eller slet ikke er korreleret med de studerendes læringsudbytte (Uttl et al., 2017). De er i stedet slet skjulte popularitetsmålinger, som afspejler og styrker en forbrugeragtig holdning til undervisning (Clayson, 2022). Tilmed er disse målinger biased til ulempe for undervisere, der er kvinder, især yngre kvinder, eller minoriteter (Mengel et al., 2018).

Men praksis består. Nogle ledelser, kvalitetsafdelinger og studerende på videregående uddannelser, såvel som nogle lærere med hidtil pæne resultater, har begrænsede interesser i forandring af evalueringrutinerne. I mellemtiden henvises der til akkreditering som begrundelse, selv om der fra et akkrediteringsperspektiv alene forlanges en "systematisk" og "velovervejet" inddragelse af de studerende, ikke en væg-til-væg evaluering, som tilmed i praksis hjemsøges af lave svarprocenter. Det medvirker til evalueringstræthed, at flere instanser (både på institutionelt og nationalt niveau) fortsat udsender spørgeskemaer om undervisningskvalitet til de studerende (DAI, 2023).

Eksemplet viser, at når evalueringssystemer først findes, og er viklet ind i krydsende organisatoriske og institutionelle logikker, kan de være vanskelige at ændre.

Forskningen fremover

Der er behov for forskning i de lange udviklingslinjer i evaluering for at revidere og kvalificere fremherskende *evaluation imaginaries*, der ofte ser evaluering næsten automatisk udvikle sig til enten det bedre eller det værre. Der er behov for at nuancere og specificere de komplicerede sammenhænge mellem, hvad der sker på de store institutionelle arenaer og i de lokale konstruktioner af evaluering og dens konsekvenser.

Disse mangfoldige konsekvenser bør studeres, ikke blot nedspilles og uskyldiggøres som såkaldte "utilsigtede konsekvenser". Konstitutive virkninger kan være et bedre samlebegreb for de ringe i vandet, som evaluering skaber for praksisser og for relationer mellem mennesker (Dahler-Larsen, 2014).

Konsekvenser af evaluering kan være meget sammensatte. Et eksempel er sammenhængen mellem evaluering og trivsel. Evaluering kan medføre overvågning og kontrol i et præstations-samfund, hvilket skaber mere og mere mistrivsel. Evaluering kan modsat være med til at udpege miljøer, hvor der er ringe muligheder for trivsel. Men selve det, at der i stigende grad måles på trivsel og mistrivsel kan også fremme tvivl og tilbyde rubrikker for at italesætte mistrivsel, som samtidig gøres mere og mere normal, jo oftere, der spørges, og jo flere, der fanges i spørgeskemakategorierne. Disse sammenhænge er komplekse. Man kan forske i selvforstærkende sammenhænge mellem måling af et problem og problemets omfang uden at negligere, at følelser af mistrivsel samtidig kan være fuldt ud reelle. Der er behov for i højere grad også at forstå selve målingen som en aktør i det store samfundsmæssige spil.

Store, institutionaliserede evalueringssystemer øger afstanden mellem de specialister, der planlægger evalueringssystemer, og de mennesker, som oplever konsekvenserne. I nogle tilfælde er det uklart, hvem der har ansvaret, og hvem der har mulighed for at forandre et evalueringssystem. Det udgør et problem for den demokratiske samtale.

Der er behov for sammenlignende studier af, hvem der er involveret i evalueringsdesign og evalueringsprocesser og under hvilke betingelser, der er offentlig adgang til evalueringresultater og muligheder for demokratisk deltagelse i

evaluering og dens anvendelse. Der er behov for forskning i, hvordan store evalueringssystemer forandres i et demokratisk perspektiv.

Det vil være frugtbart med eksperimenter, hvor aktører bringes sammen med henblik på erfaringsdeling og fælles demokratisk deliberation om evalueringssystemer. At studere og samle erfaringer med sådanne eksperimenter vil udfylde et hul i forskningen og revitalisere det demokratiske perspektiv i evaluering.

Litteratur

- Akselvoll M og Dahler-Larsen P (2021) Evaluation people and real people in home-school cooperation. In: Dahler-Larsen P (ed) *A Research Agenda for Evaluation* Cheltenham E Elgar, pp.105-127.
- Beck U (1992) *Risk Society - Towards a New Modernity*. London: Sage Publications.
- Breidahl, K. N., Gjelstrup, G., Balle Hansen, M., & Hansen, H. F. (2015). Evaluering af politisk-administrative reformer: En komparativ analyse af evalueringen af de største reformer i Danmark og Norge. *Økonomi og Politik*, 88(1), 34-45. [6].
- Bromley P og Powell WW (2012) From smoke and mirrors to walking the talk: Decoupling in the contemporary world. . *Academy of Management annals* 6(1): 483-530.
- Chelimsky E (2006) The purposes of Evaluation in a Democratic Society. In: Shaw I, Greene JC and Mark MM (eds) *The Sage Handbook of Evaluation*. London: Sage Publications, pp.33-55.
- Clayson D (2022) The student evaluation of teaching and likability: what the evaluations actually measure. *Assessment & Evaluation in Higher Education* 47(2): 313-326.
- Dahler-Larsen P (2012) *The Evaluation Society*. California: Stanford University Press.
- Dahler-Larsen P (2014) Constitutive Effects of Performance Indicators: Getting beyond unintended consequences. *Public Management Review* 16(7): 969-986.
- Dahler-Larsen P (2021) Bør arbejdspladsvurderinger offentliggøres? Om cirkulation af evaluativ viden på arbejdsmiljøområdet. *Tidsskrift for Arbejdsliv* 23(1): 84-92.
- Dahler-Larsen P og Pihl-Thingvad S (2014) *Resultatmålinger og stress*. Odense: Syddansk Universitetsforlag.
- Dahler-Larsen P, Sundby A og Boodhoo A (2021) How and how well do workplace assessments work? Using contextual variations in a theory-based evaluation with a large N. . *Evaluation: The International Journal of Theory, Research and Practice* 27(3): 326-346.

- Dahler-Larsen, P. og Møller, A.L. (under udg.): Constitutive effects of performance indicators: A review of a new stream of research. Work in progress. Institut for Statskundskab, Copenhagen.
- DAI (Danmarks Akkrediteringsinstitution) (2023). Står udbyttet mål med indsatsen? Kvalitetsperspektiver på studenterevalueringer på de videregående uddannelser.
- de Rijcke S, Wouters PF, Rushforth AD, et al. (2016) Evaluation practices and effects of indicator use-a literature review. *Research evaluation* 25(2): 161-169.
- Ejler N, Seiding HR, Bojsen DS, et al. (2008) *Når måling giver mening, resultatbaseret styring og dansk velfærdspolitik i forvandling*. København: Jurist- og Økonomforbundets Forlag.
- Espeland W og Sauder M (2007) Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology* 113(1): 1-40.
- Fochler M og de Rijcke S (2017) Implicated in the indicator game? An experimental debate. *Engaging Science, Technology, and Society* 3: 21-40.
- Grek S (2009) Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy* 24(1): 23-37.
- Hansen HF (2003) *Evaluering i Staten. Kontrol, Læring eller Forandring?* København: Samfundslitteratur.
- Kerpershoek E, Groenleer M og 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 JK (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 (1991) What Do We Know About Internal Evaluation? *Evaluation and Program Planning* 14: 159-165.
- Mengel F, Sauermann J og Zölitz U (2018) Gender Bias in Teaching Evaluations. *Journal of the European Economic Association* 17(2): 535-566.

- Møller, M. Ø., Andersen, V. N., & Iversen, K. (2016). Review af resultatbaseret styring: Resultatbaseret styring på grundskole-, beskæftigelses- og socialområdet. KORA
- Power M (1997) From Risk Society to Audit Society. *Soziale Systeme* 3: 3-21.
- 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.
- Rottenburg RaM, Sally Engle (2015) A world of indicators: the making of governmental knowledge through quantification In: Rottenburg RM, Sally Engle; Park, Sung-Joon & Mugler, Johanna (ed) *The World of Indicators* Cambridge: Cambridge University Press pp.1-33.
- Rowlands J og Wright S (2021) Hunting for points: the effects of research assessment on research practice. *Studies in Higher Education* 46(9): 1801-1815.
- Sivertsen G (2017) Unique, but still best practice? The Research Excellence Framework (REF) from an international perspective. *Palgrave Communications* 3.
- Uttl B, White CA og Gonzalez DW (2017) Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation* 54: 22-42.
- Vedung E (1998) *Utvärdering i Politik och Forvaltning*. Lund: Studentlitteratur.
- VIVE (2020). Anvendelsen af de nationale test. Delrapport 5: Evalueringen af de nationale test.
- Young M, Sørensen MP, Bloch C og Degn L (2017) Systemic rejection: political pressures seen from the science system. *Higher Education* 74(3): 491-505.
- Aagaard K (2015) How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy* 42(5): 725-737.

Magtudredningen 2.0: Essay-serien

En central del af Magtudredningen 2.0 er en bred inddragelse af forskere, hvis forskning kredser om et eller flere af magtudredningens temaer. Som led i projektet blev der i foråret 2024 afholdt 15 forskerworkshops, hvor oplægsholdere efterfølgende blev inviteret til at omarbejde deres oplæg til et essay. Essay-serien er forfatternes perspektiver på centrale tematikker for en dansk magtudredning og har forfatterne som afsender.

Dette essays er en del af tema 4 i Magtudredningen 2.0's forskningsplan: Hvad er vilkårene for det lokale og regionale demokrati i det 21. århundrede?

Siden den første magtudredning er der sket store ændringer i det lokale demokrati og der opstår i for mange tilfælde uklarhed i den offentlige debat om hvorvidt beslutninger træffes på kommunal, regionalt eller statsligt niveau. Dette tema fokuserer på strukturelle forskydninger og statens styring af kommunerne (eksempelvis i forbindelse med budgetloven), kommunernes implementering af vedtagne politikker, innovation i det lokale demokrati og nye rammer for den lokale velfærd, f.eks. gennem frikommuneforsøg og samarbejdsformer der går på tværs af sektorer og faggrænser. Desuden behandles de afledte konsekvenser for borgernes oplevelse af det lokale demokrati.