

Evaluering som udvikling

Udvikling er det grundlæggende formål med evaluering. Al evaluering handler om feedback, det vil sige tilbageføring af information om styring og organisering med henblik på at stimulere til forbedringer af denne styring og denne organisering. Udtrykket "evaluering" må absolut ikke afgrænses, så denne idé om feedback falder uden for begrebet. Udviklingsorienteret evaluering kan have mange former som for eksempel selv-vurdering, bedømmelse af kolleger og samarbejdspartnere samt brugerevaluering. Den kan være formativ og summativ. To problemer med udviklingsorienteret evaluering er tilbøjeligheder til Potemkin-kulisser og forekomster af sociale fælder. Ud over udvikling kan evaluering også have kontrol og grundviden som formål.

Kaizen er et ledelsessystem i Japan, som indebærer kontinuerlig forbedring. Kaizen er to ord på japansk: Kai og Zen. Kai kan oversættes med "gør" og Zen med "bedre", "gør bedre" (kanji: 改善, "forbedre"). Disse to ord sat sammen står for det, der er blevet en berømt doktrin inden for japansk management. Systemet er udviklet af Taiichi Ohno (大野 耐一), fader til Toyotas produktionssystem. I Sverige anvendes ordet især om en måde at forholde sig til kvalitet på inden for en virksomhed, hvor aktiviteterne stadig forbedres via mange små ændringer. Kaizen indebærer altså stræben efter kontinuerlig forbedring og handler ikke om at ændre metode radikalt. Målet med kaizen er at mindske sløseri inden for virksomheden, det vil sige at eliminere momenter, som indebærer omkostninger uden at tilføje noget af værdi. Foranstaltningerne inden for kaizen begrænses ikke til den enkelte virksomhed, men omfatter også såvel leverandører som personalets sundhed og velbefindende uden for arbejdspladsen.¹

Evaluering hviler på en uhørt *enkel idé*. Evaluering handler om at kortlægge ordentligt og bedømme både resultatet af noget, samt hvordan man er kommet frem til det, det vil sige metoden. Metoden er kolossalt vigtig i evaluering; evaluering er ikke bare en bedømmelse af et resultat, metoden må og skal inddrages. Det, der bedømmes og kortlægges i evaluering, er, hvad der foregår i den offentlige sektor. Selvfølgelig kan man også evaluere private virksomheder, eller hvad der foregår i den frivillige sektor. Men eva-

luering er opstået som en reformbevægelse i den offentlige sektor. Og i denne artikel begrænser jeg mig til denne. Den store pointe med alt dette er feedback, altså tilbageføring af information. Hensigten er at fremskaffe et godt grundlag, som kan føres tilbage til beslutningstagere og interessenter, så de kan træffe beslutninger om ansvarlighed eller forbedringer. Forbedringsfunktionen kalder vi også *udvikling*.

Kaizen plejer ikke at blive sat i forbindelse med udvikling gennem evaluering. Men der findes lighedspunkter, for eksempel understregningen af vigtigheden af kontinuerlige, små forbedringer. Og det er tankevækkende, at disse to systemer, der begge reflekterer over det, man gør, har set dagens lys og er blevet iværksat næsten samtidigt, men i helt forskellige dele af verden. Længere end dette vil jeg ikke trække parallellen mellem udviklende evaluering og kaizen. Jeg vil gå videre med udviklende evaluering og lade det sagte tjene som øjenåbner.

Hvad ordet "evaluering" betyder

Formelt kan ordet "evaluering" (engelsk: evaluation) defineres som omhyggelig (ordentlig, systematisk) slutbedømmelse af resultater, slutpræstationer, forvaltning og beslutningsindhold samt organisering af offentlig virksomhed, som menes at spille en rolle i praktiske beslutningssituationer (Vedung 1998: 20ff).

Den, der læser dette, bør huske, hvad jeg mener med en definition. En definition er blot et udsagn om, hvad et ord betyder. Min definition herover angiver *ikke*,

¹ <http://sv.wikipedia.org/wiki/kaizen> (20081102).



Evert Vedung
 Professor
 Statsvidenskab, Uppsala Universitet
 Oversat af Astrid Holtz

hvad der er god eller dårlig evaluering. Jeg forsøger blot stilfærdigt at påpege, hvordan jeg selv vil bruge udtrykket. Desuden mener jeg, at ovenstående er en *minimaldefinition*. Med dette mener jeg omtrent følgende: Hvad er det allermindste, man kan forlange af en aktivitet eller et produkt af en aktivitet, for at man kan kalde det "evaluering"?

Evaluering er en afsluttende bedømmelse. Med dette mener jeg en bedømmelse af aktiviteter, der er indført, som foregår, eller som evt. er afsluttede. På den ene side er denne begrebsbestemmelse snæver. Den udelukker evaluering ex ante, det vil sige konsekvensanalyser af påtænkte og overvejede interventioner, som foretages, før interventionerne besluttet og sættes i gang. På den anden side er den bred. Den begrænser sig ikke kun til beslutningernes og aktiviteternes effekt på resultatniveauet (det vil sige ude i samfundet eller i naturen), men inkluderer også slutpræstationer, forvaltningsprocesser og beslutningsindhold.

"som menes at spille en rolle i praktiske beslutningssituationer." Dette lille indslag i definitionen er vigtigt. Det betyder følgende: For at jeg skal kalde noget for en evaluering, må hensigten med den også være, at den skal bruges i kommende beslutningssituationer og danne grundlag for fremtidige afgørelser. Den behøver ikke faktisk at komme i brug, det er tilstrækkeligt, at det tilsigtes, at den tages i brug. Hvis jeg sætter pegefingeren på en rapport og siger: "Dette er en evaluering," så betyder det blandt andet, at det må være meningen med rapporten, at den skal spille en rolle i praktiske beslutningssituationer.

Jeg siger ikke i definitionen, hvad disse praktiske beslutningssituationer er. Jeg vil holde definitionen smal. Det kan være alle slags beslutningssituationer i den offentlige sektor. Det kan være beslutningssituationer, som handler om ansvarlighed. Det kan være beslutningssituationer, der handler om udvikling. Jeg har i definitionen ikke villet specificere, hvad der menes med beslutningssituationer, for at lade stå åbent, at det kan handle om alle mulige beslutningssituationer, så længe det handler om beslutningssituationer.

Til sidst bør det måske for tydelighedens skyld nævnes, at evaluering ifølge min definition er beregnet ikke blot på sædvanlig fagpolitik angående for eksempel socialt arbejde, uddannelse, bolig, transport, byplanlægning, miljø og energi, men også forvaltningspolitik, det vil sige den offentlige sektors organisering inklusiv mål- og resultatstyring, samarbejdsstyring samt udliciteringsaftaler og anden inddragelse af private og civile områder i det offentlige gennemførelsesforløb (Vedung 1998: 20 ff). Definitionen dækker med andre ord både programevaluering og organisationsevaluering.

Definitionen som sorteringsinstrument

Denne minimaldefinition af evaluering kan siden hen benyttes som sorteringsinstrument. Forestil dig, at vi står foran en stor bunke med rapporter, som vi har en mistanke om, har med evaluering at gøre. Den skulle kunne deles op i to bunker. I den ene bunke lægger jeg evalueringer, i den anden alt det andet. Først lægger jeg alle de rapporter, der ikke handler om den offentlige sektor, i den anden bunke. Blandt de rapporter, der er tilbage, finder jeg en del, der indeholder bedømmelser af forslag til organisationsændringer og virksomheder, og som man bare har tænkt sig at indføre, men som endnu ikke er blevet indført. Også disse lægger jeg i den anden bunke. En del af de tilbageværende rapporter henviser til organisationsændringer og virksomheder, som er vedtaget og foregår, eller som er afsluttede. De indeholder også bedømmelser af enten resultater, slutpræstationer, forvaltning eller beslutningsindhold. Hvis de desuden er omhyggeligt udført, og det kan påvises, at de er lavet med den hensigt, at de skal komme til praktisk anvendelse, så siger jeg: "Dette er evalueringer," og dem lægger jeg i bunken med evalueringer.

Nu står jeg så der med en bunke rapporter, som jeg kan pege på og sige: "Dette er evalueringer." Dermed har definitionen opfyldt sin opgave. Hvis jeg skal granske denne bunke med evalueringer yderligere i den hensigt at sige noget interessant om dem, så har jeg ikke længere brug for definitionen. Den har udspiilet sin rolle, nemlig at være et instrument til udvælgelse. Men denne bunke af rapporter, som jeg kalder

for evaluering, er jo kendetegnet af utrolig meget andet, der ligger uden for det minimum, som blev taget op i definitionen. For at sige noget om dette andet har jeg brug for andre begreber end dem, der indgik i definitionen. Til disse begreber hører klassificeringer af evalueringers overordnede formål. Denne opgave har jeg tænkt mig at tage fat på nu.

Evaluerings overordnede formål

På det højeste niveau handler al evaluering om læring om praktisk anvendelse. Evaluering er en aktivitet, der reflekterer over andre aktiviteter for at uddrage viden, som kan være til gavn i fremtidige forbedringer af disse andre aktiviteter. Læring er al evalueringens mest overordnede pointe.

Næst efter dette mest overordnede formål (engelsk: purpose) kan evaluering stille mod udvikling, kontrol, grundviden.

Læring

- 1.1 Udvikling (fremme (engelsk: "promotion, development"))
- 1.2 Kontrol (regnskab, ansvarlighed (engelsk: "accountability"))
- 1.3 Grundviden (engelsk: "basic knowledge")

Jeg har tænkt mig først at skrive mere udførligt om udvikling som formål med evaluering. Kontrol og grundviden vil jeg kort tage op til sidst.

Udvikling

Formålet med evaluering kan også være at udvikle. Det kan handle om at forandre, rette og bevare. Men det kan også handle om at afvikle. Formålet med at udvikle kaldes også fremme eller på engelsk "promotion". Udvikling handler om at gøre ting bedre.

Udvikling er en meget værdig bevæggrund for evaluering. Mange hævder, at udvikling er hovedpointen. Ved evaluering for udvikling er rapporterne ægte og betingelsesløst rettet mod på en solid base af data at fremstille sandheden om det offentliges indgriben for derigennem at levere et velforberedt grundlag for forbedring af denne indgriben. Der findes ingen andre skjulte eller bagvedliggende hensigter.

Formativ udvikling

Det udviklende formål kan være formativt eller summativt. Ved *formativ udvikling* er pointen via en nøjagtig kortlægning og bedømmelse af en *igangværende aktivitet* at blive klogere for at kunne forbedre den selvsamme aktivitet. Formativ udvikling bygger på *formativ evaluering* af en virksomhed, i hvilken man ikke bringer virksomhedens mål på bane, men koncentrerer sig om at granske



midlerne (metoderne, strategierne, handlingerne) til at nå disse mål. Formativ evaluering er en proces for fejlopdagelse, hvilket inden for de bestående strukturelle rammer skal føre til modificerede strategier og handlinger for bedre at nå de givne mål. Målene flyttes ikke, de bringes ikke på bane. Er midler, metoder, strategier og handlinger overensstemmende med formålet for at opnå de givne mål? Det er hovedspørgsmålet i en formativ evaluering. Idéen er at opnå single loop-læring. Single loop-læring indebærer, at organisationen lærer noget om, hvordan midlerne fungerer, hvilket fører til, at den ændrer og forbedrer midlerne uden på mindste måde at ændre på målene.

Det er her, de små, kontinuerlige forbedringer, som Kai-zen også taler om, kommer ind i billedet. At foretage små, små forbedringer er ikke noget at skamme sig over. Alt i den offentlige sektor behøver ikke være storslået og imponerende. Skønt der også findes et sådant formål med udvikling. Det kaldes summativ udvikling.

Summativ udvikling

Ved *summativ udvikling* er pointen via en nøjagtig kortlægning og bedømmelse af en igangværende aktivitet at lære noget for senere at kunne søsætte lignende, men bedre aktiviteter. Summativ udvikling bygger på *summativ evaluering*. Summarisk evaluering bringer ikke bare midlerne, men også målene og måske også det underliggende problem på bane. *Summativ evaluering* indebærer double loop-læring. Ved double loop-læring diskuteres ikke blot midlerne (metoderne, strategierne, handlingerne) til at nå givne mål og løse problemer; hele programmet med problemer, mål og midler studeres og diskuteres. Summativ udvikling er dermed mere storslået og omfattende end formativ udvikling.

Former for udviklingsorienteret evaluering

Udviklingsorienteret evaluering og især den formative slags med små, små forbedringer kan have et utal af former. Den kan udføres som *selvvurdering*. Alle mulige interessegrupper i en offentlig virksomhed eller en organisation kan beskæftige sig med selvvurdering. Lad os tage professionelle socialarbejdere som eksempel. Hver enkelt socialarbejder skulle nu og da

kunne reflektere lidt mere ihærdigt og systematisk over forskellige aspekter ved, hvad hun udretter, og hvordan det skulle kunne forbedres (Schön 1991). Disse refleksioner skulle også kunne ske i fokusgrupper.

Selvvurdering kan *kombineres med bedømmelse af kolleger*. Man indkalder da en gruppe ligemænd (peers) fra anden side, kyndige og erfarne kolleger, som på basis af selvvurdering og observation på stedet, tekstlæsning samt samtaler og interview kommer med forslag til forbedringer (Haraldson m.fl. 2005).

Der findes mange andre modeller. For eksempel bedømmelse af samarbejdspartnere eller Hanne Krogrstrups BIKVA-model. Eller brugerorienteret evaluering, hvorved kunderne eller brugerne får mulighed for at vælge et omfang at evaluere i samt at udføre selve bedømmelsen.

Vanskeligheder med udvikling:

Potemkin-kulisser og sociale fælder

En af vanskelighederne med evaluering for udvikling i den offentlige sektor hænger sammen med tilbøjeligheden til at bygge *Potemkin-kulisser* og have i *sociale fælder*.

I Sverige og Finland berettes gerne historien om Katarina II's yndling og elsker, fyrst Grigorij Potemkin, som under kejserindens visitationsrejse i 1787 i det nyligt erobrede Krim lod opføre tilfældige byer som kulisser for at bilde hende ind, at provinsen under hans statholderskab var blomstret hurtigt op. Potemkin – den underordnede, udføreren – som havde erobret og et stykke tid forvaltet provinsen, forsøgte at udnytte sit enorme informationsovertag over for kejserinden – den overordnede, bestilleren – som langt borte i det fjerne St. Petersborg ikke havde mulighed for at sætte sig ind i landsdelens problemer. Motivet bag forsøget på at vildlede Katarina var egeninteressen i at bevare hendes gunst og redde sit eget skind. Fyrsten foregreb kejserindens sorg og måske vrede, hvis hun fik indtryk af, at provinsen var varetaget dårligt, og handlede på grundlag af denne forestilling for at undgå hendes negative reaktion.

Selv i nutidig evaluering for udvikling kan en tilbøjelighed til at bygge Potemkin-kulisser snige sig ind. For at den udviklende idé med evaluering skal fungere optimalt, bør for eksempel professionelle socialarbejdere, som foretager en selvvrurdering, åbent og ærligt kunne vride og vende alt, der angår deres arbejde og arbejds-situation. Alt skal der gøres rede for og bedømmes, for at det kan blive forbedret og udviklet. Men socialarbejderne indgår i en organisation. Dette indebærer, at de har overordnede over sig, hvilket lægger bånd på dem. Institutioner, siger man i dag. Institutioner indebærer, at der findes regler, skrevne og uskrevne, som det kan være hensigtsmæssigt at følge. Et eller andet sted findes der en kejserinde med magt til at forringe deres vilkår, hvis hun hører, at virksomheden varetages dårligt, selv om formålet med at drage dette frem i lyset er udvikling og forbedring. Dette gør, at socialarbejderne ikke kan være helt ærlige og åbne i deres redegørelse. De tvinges til at *spille komedie*, tilbageholde information, frisure virkeligheden, male med lysere farver, end de burde. De havner i den *so-ciale fælde*. Dette hæmmer evaluering for udvikling.

Potemkin-kulisser og sociale fælder opstår i magt-sammenhæng på grund af *strategiske* udregninger. Mennesker, som handler strategisk, agerer på grundlag af, hvad de *tror*, at *andre* agter at gøre. Strategisk handling er baseret på *anticiper*ing, dvs. at man foregriber andres fremtidige handlinger og deres sandsynlige konsekvenser og agerer på grundlag af, hvad man tror, at dette kommer til at betyde for en (Vedung 1998: kap 6 og Vedung 2009: kap 8, undervejs).

Spørgsmålet om, hvordan strategisk ageren af typen Potemkin-kulisser skal kunne undgås ved udviklende evaluering, må jeg holde uden for denne lille historie.

Kontrol

Meningen med ordentligt at kortlægge og bedømme en virksomheds resultater og processerne frem til dem kan være at gøre rede for dem over for en overordnet. På grundlag af blandt andet den redegørelse er det siden tænkt, at den overordnede skal kunne kontrollere, hvordan den underordnede har varetaget sit hverv, og stille denne til ansvar. Dette kaldes *kontrol*. Kontrol er det andet grundlæggende formål med evaluering ved siden af udvikling. Man kan også kalde det indberetning, ansvarlighed eller, på engelsk, *accountability*.

Kontrol omfatter to parter, arbejdsgiveren (hovedmanden, bestilleren) og den, der tager opgaven (arbejdstageren, udføreren), og to funktioner, arbejdstagerens redegørelse opad og arbejdsgiverens bedømmelse nedad. Arbejdsgiveren kan ikke tage sig af alt selv, men uddelegerer hvervene til en arbejdstager; sidstnævnte binder sig ved en overenskomst til at udføre disse hverv. Efter et stykke tid forventes arbejdstageren at redegøre for sine handlinger og sine opnåede resultater til arbejdsgiveren. På grundlag af denne indberetning kan arbejdsgiveren kontrollere, at arbejdstageren har fulgt overenskomsten, tage beslutning om bebrejdelse, ros og rettelser samt bevilge regnskabsgodkendelse.

Den offentlige sektor i et repræsentativt demokrati som Danmark er sammensat af et stort antal relationer mellem arbejdsgivere og arbejdstagere, principaler og agenter. Folket, demos, er politikernes arbejdsgivere, og politikerne folkets arbejdstagere. Politikerne er til gengæld den højere forvaltnings arbejdsgivere, og den højere forvaltning er politikernes arbejdstagere. Den højere forvaltning er igen den lavere forvaltnings arbejdsgiver, og den lavere forvaltning er den højere forvaltnings arbejdstager osv. Den offentlige sektor i demokratier som Danmark, Sverige, Tyskland, USA og andre lande består af et stort antal relationer mellem arbejdsgivere og arbejdstagere. Ved evaluering som kontrol lægges der vægt på, at arbejdsgivere i alle led i den offentlige sektor har behov for evaluerende redegørelser fra sine arbejdstagere for at kunne stille dem til ansvar og evt. bevilge dem ansvarsfrihed.

Grundviden

Grundviden (engelsk: "basic knowledge") er i øjeblikket ikke det fremtrædende formål med evaluering. Dog synes dette tredje formål med evaluering at være på fremmarch, for eksempel inden for bevægelsen for evidensbaseret ageren i den offentlige sektor og dens favorisering af systematiske vidensoversigter. Ved grundviden skal evaluering bidrage til overlappende, mere generel viden om institutioner, organisationer, processer, handlinger, begivenheder, ledelsesstrategier, ledelsesmidler eller myndigheders måde at fungere på. Dette formål med evaluering er ikke begrænset til læring inden for det program eller den organisation, der evalueres. Pointen er meget mere generel. Grundviden som formål med evaluering minder om pointen med grundforskning ved universitetet, som udføres uden praktisk anvendelse for øje. I grundviden i evaluering findes imidlertid tanken om en uspecificeret praktisk læring meget langt fremme i tiden. Selv grundviden handler altså om udvikling, men på langt sigt.

Formålet med grundviden titter frem, når bestilleren lægger vægt på, at evalueringsrapporter skal spredes i vide kredse, langt borte fra den aktivitet, der er blevet evalueret. Endnu mere tydeligt er formålet med grundviden, når der propaganderes for sammenfattende meta-evalueringer af, hvad et antal enkelte evalueringer på et område er kommet frem til. Pointen med sådanne informationssammensætninger (systematiske vidensoversigter) er at trække det generelle frem af en mængde specialrapporter ved at indordne det i et mere almengyldigt begrebssystem, så det bliver lettere tilgængeligt som beslutningsgrundlag (Rieper og Foss Hansen 2007).

Epilog om evaluering for udvikling

Denne artikel har diskuteret *udvikling* som et grundlæggende formål med evaluering. Al evaluering handler om feedback, dvs. tilbageføring af information om styring og organisering med henblik på at stimulere til forbedringer af denne styring og organisering. Ordet "evaluering" må absolut ikke afgrænses, så denne idé om feedback falder uden for begrebet. Udviklingsori-

enteret evaluering kan have mange former som eksempelvis selvvurdering, bedømmelse af kolleger og samarbejdspartnere samt brugerevaluering. Den kan være formativ og summativ. To problemer med udviklingsindrettet evaluering er tilbøjeligheder til Potemkin-kulisser og forekomster af sociale fælder.

Evert Vedung er professor emeritus i statsvidenskab, især boligpolitik ved Uppsala Universitet. Han er en af grundlæggerne af den svenske evalueringsforening. Han er adjungeret professor ved Københavns Universitet og har gæstet RUC, Odense Campus samt Århus og Aalborg Universitet. Han er æresmedlem af Dansk Evaluerings Selskab.

Litteratur

Haraldsson, Jennie m.fl. (2005): *Tid för gemensam reflektion: Självvärderingar inom storstadssatsningen i Göteborg*. Göteborg: Göteborgs stad, Stadskansliet och Göteborgs universitet, Institutionen för arbetsvetenskap. 54 sidor.

Kaizen, se <http://sv.wikipedia.org/wiki/Kaizen> og <http://en.wikipedia.org/wiki/Kaizen> 20081002.

Krogstrup, Hanne Kathrine (2003): *Evalueringsmodeller: Evaluering på det sociale område*. Århus: Systime.

Rieper, Olaf og Hanne Foss Hansen (2007): *Metodedebatten om evidens*. København: AKF-Forlaget. www.akf.dk/udgivelser/2007/pdf/metode debat_evidens.pdf/ d. 28. oktober 2008.

Schön, Donald (1983): *The Reflective Practitioner: How Professionals Think in Action*. Aldershot: Ashgate.

Vedung, Evert, 1998/2009: *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur, 2. oplag. 3. oplag ventes at udkomme i foråret 2009.