

# Evaluering – fra rituellet til reelt.

## Om å se i bakspeilet for å komme seg videre

Knut Samset

*Det er ikke populært at noen kikker en i kortene. Man kan saktens se at det kan være nyttig at noe blir evaluert, men vi vil ikke bli evaluert selv. Evalueringskontoret i en av FN-organisasjonene var derfor vant til at de avdelingene det gjaldt protesterte når noe skulle evalueres. Men de kom ikke langt med protestene, for begrunnelsen var ofte usaklig, til dels fantasifull. For å komme klientene i forkjøpet laget man derfor et standardformular med overskriften «Begrunnelse for hvorfor vi mener at prosjektet ikke bør evalueres» som fulgte med som vedlegg.*

*Avdelingene kunne nå krysse av på den eller de begrunnelsene de ville påberope seg. Listen var lang, med punkter som «Det blir for dyrt», «Det er unødvendig fordi prosjektet er meget vellykket», «Vi kan ikke se nytten av det», «Prosjektet er for sensitivt», «Det er for tidlig», «Det er for sent i prosessen», etc. Nederst stod det med liten skrift at evalueringskontoret antakelig uansett ikke ville ta hensyn til protesten. Det virket.*

## Evalueringens formål

Internasjonal bistandsvirksomhet har alltid ligget et hestehode foran når det gjelder å evaluere nytten av utviklingsprosjekter. Det skyldtes et standhaftig (og til dels berettiget) rykte om at det meste går galt, en sterk opinion som krevet innsyn, og politikere som krevet begrunnelser for nye investeringer. Organisasjoner som FN og Verdensbanken gikk i bresjen med å utvikle metodikken og anvende evaluering som hjelpemiddel. De enkelte landenes bistandsmyndigheter diltet etter.

Etter hvert har dette grepet om seg, og i dag brukes evaluering i utstrakt grad som ledelsesverktøy i alle sektorer og på alle nivåer i samfunnet. Bare i Norge er per i dag registrert om lag 2500 evalueringer på det statlige nettstedet «Evalueringsportalen.no», hvor rapportene kan leses og lastes ned. Erfaringen er at evaluering blir stadig viktigere i evidensbasert planlegging og beslutning, og er blitt en industri med betydelig omsetning for en rekke konsulentselskaper og forskningsinstitutter. I 2009 fikk vi Norsk Evalueringsforening, og i 2015 feiret vi det internasjonale evalueringsåret.

*.. avdelingene det gjaldt protesterte når noe skulle evalueres. Men de kom ikke langt med protestene, for begrunnelsen var ofte usaklig, til dels fantasifull*

det

Da kan det saktens være på sin plass å reflektere tilbake til bistandsevalueringene og tenke litt fremover: Hva skal vi bruke evalueringer til i dag?

## Rituelle evalueringer

Begrunnelsen for å evaluere er enten å *kontrollere* hva som er blitt gjort, bedre *styringen* noe som er i gang, eller *lære* av det som har skjedd. Men erfaringen er at resultatene og anbefalingene fra evalueringer ikke alltid blir brukt. Enkelte skeptikere hevder derfor at en evaluering ofte er en rituell handling mer enn en sannhetssøkende prosess. Det har de nok rett i, for i enkelte tilfeller gjennomføres evalueringer for å skape legitimitet for noe som har skjedd, og dersom man er

bekymret for at konklusjonen vil være negativ, blir det viktig å ha kontroll på den som leder teamet og i tillegg personer som ikke stiller vanskelige spørsmål. Såkalt nyttige idioter.

Dette var nokså vanlig praksis innenfor enkelte av FN-organisasjonene, i alle fall i tidligere tider. Man kalte det «Tri-partite» evaluering, som skulle ivareta interessene til de tre sentrale partene, det vil si giverlandet som finansierte det hele, FN-organisasjonen som var mellomleddet med ansvar for gjennomføringen, og mottakerlandet. Partene fikk velge hver sin deltaker i evalueringsteamet. I prinsippet var det fornuftig, men i praksis ble det allikevel feil. Det vi så var at FN-representanten, som også ledet teamet ofte var en pensjonert tjenestemann med lojalitet til sin tidligere arbeidsgiver og med håp om flere lukrative oppdrag senere. Giverlandets representant ble gjerne valgt fra en liste konsulenter som tidligere hadde vist seg å være vennligsinnete, og mottakerlandet ble oppfordret til å sende med en yngre medarbeider, helst en kvinne. Det sier seg selv at utfallet av disse evalueringene ble nokså forutsigbart og at det meste derfor tilsynelatende lot til å være i orden med FN-prosjektene. Men i dag er uavhengighet og habilitet urokkelige krav til evalueringsteam, så verden går fremover.

Men evalueringen må ikke bare være etterrettelig, den må også bli brukt. Et eksempel på det motsatte så vi i det sørlige Afrika der giverlandet hadde vært inne over en periode på mer enn 20 år for å bore etter vann til lokalbefolkningen på landsbygda. Prosjektet ble evaluert flere ganger og jeg hadde ansvar for to av dem. Allerede i den første evalueringen slo evaluator fast at dette var et av stedene i Afrika hvor det var minst behov for brønnboring, fordi grunnvannet lå så høyt at folk flest lett kunne grave brønner selv. Man anbefalte derfor å avvikle prosjektet.

Men det skjedde ikke, og neste evaluering 3-4 år senere registrerte at en nå hadde etablert en svær prosjektorganisasjon med avansert og kostbart utstyr, noe landet selv ikke ville ha råd til å videreføre. Anbefalingen var igjen å avvikle. Den tredje evalueringen noen år senere fant at svært mange brønner var ute av drift, de var ødelagt, ikke vedlikeholdt, plassert nær gravsteder, forurenset, osv. Samme anbefaling én gang til: å avvikle. Den fjerde evalueringen etter mer enn 20 år fant at de samlede investeringene som nå var kommet opp i omtrent én milliard kroner var bortkastet. Det ville vært billigere å holde hele befolkningen med flaskevann fra Europa gjennom alle årene, stod det i rapporten. Anbefalingen var på nytt å avvikle.

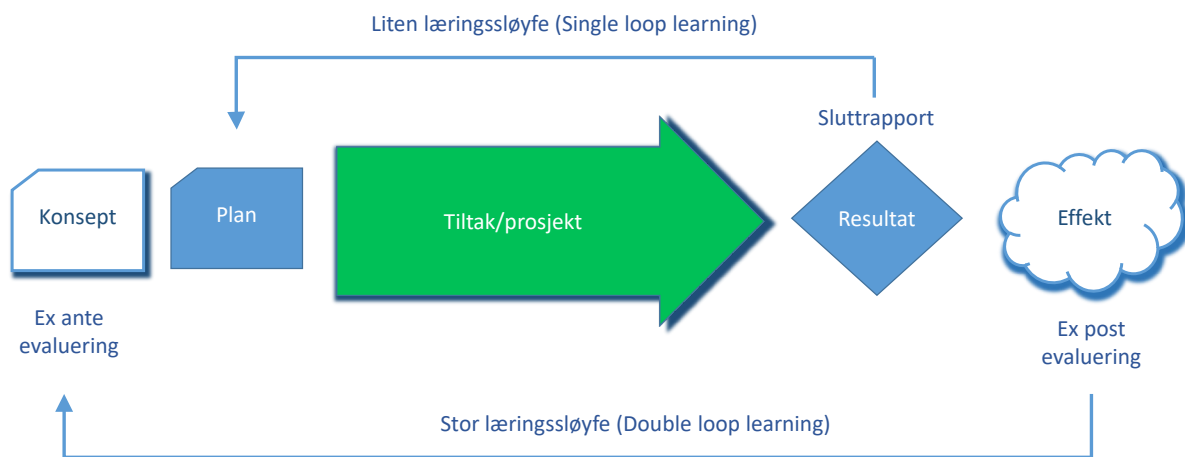
Nå skulle man tro at giverlandet hadde fått nok informasjon til å kunne sette et punktum, men nei, man valgte å evaluere på nytt, men nå med et smalere perspektiv som bare gjaldt selve leveransen, det vil si antall brønner som var produsert. Denne gangen fikk man en evalueringsrapport som konkluderte at alt var i orden, fra et lite firma som ikke hadde erfaring fra evalueringer, og bare hadde hatt en kort befarig på stedet. Prosjektet kunne fortsette.

Kunne noe slikt skjedd i Norge i dag? Lite trolig vil de fleste si. Men det finnes eksempler. I disse dager har man etter snart 30 år med utredninger, derav åtte samfunnsøkonomiske analyser, satt i gang forprosjekt av Stad skipstunnel for antatt byggestart i 2018. Lokale interessenter med politisk støtte fra berørte fylker og kommuner har greid å holde prosessen i gang til tross for gjentatte avslag fra Staten, som ikke har klart å stoppe prosessen til tross for at alle evalueringene viste at investeringen ikke er lønnsom.

Når politiske krefter slipper løs kan det meste overstyres. Og da er det ikke sikkert at evalueringresultatene kommer til nytte. Dette så vi eksempel på i et annet afrikansk land der giverlandet bidro til å bygge opp en teknisk etat på nasjonalt nivå og finansiere en del kostbare

anleggsmaskiner. Da programmet ble evaluert så det hele tilforlætelig ut og godt organisert. Utstyret var ankommet og i bruk. Men så oppdaget vi at det også var andre giverland inne i sektoren. Etter en runde på de forskjellige ambassadene gikk det opp for oss at mottakerlandets myndigheter hadde sendt dokumentasjon for anskaffelsen av det samme utstyret til flere giverland og fått oppgjør fra alle, uten at de satt igjen med tilsvarende anskaffelser. Store beløp var med andre ord blitt borte.

Evalueringsrapporten var godt dokumentert og helt klar på hva som hadde skjedd. Saken ble øyeblikkelig løftet opp på øverste nivå hos giverlandet, og evalueringens teamleder ble innkalt til direktørens kontor sammen med etatens ledere, hvor spørsmålet var «Hva gjør vi med dette?» Det kom flere forslag: «kreve etterforskning?», «kreve tilbakebetaling?» «kutte bistanden?». Etter en kort diskusjon konkluderte direktøren med et skjevt smil og sa, «nei, jeg tror vi legger den i skuffen, jeg». Og så la han den i skuffen. Omkostningene i form av negativ publisitet ved å gjøre noe med saken ble for høye. Og i mottakerlandet ville man ikke kommet noen vei med det hele, den var antakelig forankret alt for høyt opp i regjeringen.



## Reelle evalueringer

Eksemplene ovenfor illustrerer at evalueringer kan brukes rituellet, og uten at de er sannhetssøkende eller at resultatene blir brukt. Dette er antakelig fremdeles mer vanlig enn de fleste tror. Men det er antakelig mindre vanlig hjemme enn i bistandsvirksomheten. Statsforvaltningen i Norge driver en omfattende evalueringsevirsomhet. I perioden 2005-2011 ble det levert omtrent hundre evalueringsrapporter årlig (Askim m.fl. 2013). De omfatter de fleste politikkområder. Men det er først og fremst underliggende etater og ikke departementene selv som er oppdragsgiver. Og evalueringer gjennomføres i økende grad av konsulenter og i mindre grad av forskningsinstitutter, noe man tror kan ha sammenheng med den lave terskelen for anbud på offentlige anskaffelser i Norge.

Det er grunn til å tro at behovet for evaluering her hjemme i de fleste tilfellene er godt begrunnet og at evalueringene gjennomføres av godt kvalifiserte team. Det har skjedd en omfattende profesjonalisering med årene, og kvaliteten på evalueringsrapportene er gjennomgående god. Men en kan ikke på det grunnlaget anta at de også blir lest eller brukt. Det

Eksemplene illustrerer at evalueringer kan brukes rituellet, og uten at de er sannhetssøkende eller at resultatene blir brukt. Dette er antakelig fremdeles mer vanlig enn de fleste vil innrømme

dreier seg om tykke rapporter med tungt tilgjengelig stoff som gjelder komplekse saksforhold, og en målgruppe som nødvendigvis er smal. En metaevaluering om nytten av evalueringer viste dette med all tydelighet. (Samset m.fl., 1993). Forfatterne intervjuet blant annet saksbehandlere og beslutningstakere i bistandsforvaltningen om hvilke læringsarenaer som var viktigst. Av en liste på 19 arenaer kom evalueringsrapportene på 15. plass. Ikke overraskende er det slik at folk lærer først og fremst ved egen medvirkning, og i mindre grad ved kommunikasjon. De viktigste læringsarenaene var befaringer på stedet, kontakt med andre organisasjoner, uformell informasjonsutveksling med kolleger, og den løpende rapporteringen. Ikke evalueringsrapportene.

Men selv om situasjonen i dag er at vi er blitt bedre som bestillere og gjennomførere av evalueringer, at kvaliteten er god og at svarene er relevante, er det som nevnt ikke gitt at resultatene blir brukt. For evalueringene griper inn i en virkelighet hvor det handler om både å gjøre rasjonelle og politiske valg, som ikke alltid er sammenfallende.

Men det viktigste argumentet mot evaluering handler ikke om dette, men at de fleste evalueringene gjøres på et alt for sent tidspunkt. Det kalles Ex post eller sluttevaluering. Evalueringen blir dermed bare en *post mortem*-undersøkelse, fordi prosjektet er avsluttet - pasienten er allerede død. Dersom hensikten er å styre prosessen, er det for sent fordi denne er avsluttet. Dersom hensikten var kontroll, så måtte det i så fall være for å finne den skyldige. I en drapssak blir det viktig å finne morderen, men det er ikke like sentralt i prosjekter og programmer. Da gjenstår *læring* som begrunnelse, og her blir evalueringsvirksomheten legitim. Men igjen, bare dersom læringen anvendes.

Nå er det bare det at de færreste tiltak og prosjekter blir gjenstand for formell evaluering for å fastslå effekten etter at de har vært i drift noen år. I de fleste tilfellene nøyer man seg med en sluttrapport som oppsummerer det som gjelder gjennomføringen av tiltaket. Muligheten for læring er da begrenset til det som gjelder, nettopp, gjennomføringen av tiltaket.

Man skiller mellom to typer læringsløyper, som vist i figuren (Argyris et Schön, 1978). Den lille læringsløyfen (single loop learning) bidrar til å forbedre en prosess uten å stille spørsmål ved selve prosessen. Den store læringsløyfen derimot (double loop learning) tar tak i de grunnleggende spørsmålene om hvorfor, som kan føre til at hele prosessen endres eller erstattes av noe annet. For at slik læring skal være mulig må man inn med evaluering på et senere tidspunkt hvor man kan fastslå effekten og nytten av tiltakene. Men hvis denne innsikten virkelig skal gi en læringseffekt, må man også inn med disse erfaringene på et tidlig tidspunkt før planleggingen av nye tiltak, der det er mulig å komme inn og korrigere for det mer grunnleggende og prinsipielle.

Et interessant eksperiment med akkurat dette er Statens ordning med ekstern kvalitetssikring av store statlige investeringsprosjekter. Den har innført en formell konseptfase for alle nye prosjekter som innebærer at departementer og etater må gjennomføre en evaluering i forkant (ex ante) for å identifisere og evaluere de mest lovende konseptuelle løsningene på et gitt problem. Resultatet blir deretter kvalitetssikret ved hjelp av eksterne rådgivere, før det eventuelt legges frem for regjeringen som foretar konseptvalget, og deretter Stortinget som fastsetter budsjettet.

Læring forutsetter at man kan fastslå effekten og nytten av tiltakene. Og dersom denne innsikten virkelig skal gi en læringseffekt, må man inn med disse erfaringene på et tidlig tidspunkt før man planlegger nye tiltak

I løpet av de første 17 årene siden ordningen kom i gang i år 2000, er omtrent 270 prosjekter kvalitetssikret. Ettersom det tar mange år med planlegging og gjennomføring før disse store prosjektene kan evalueres så er i skrivende stund (2015) bare omtrent 30 prosjekter kommet så langt. Omtrent halvparten av disse er evaluert, og antallet vil øke etter hvert.

Evalueringene legges ut på anbud for å sikre uavhengighet og habilitet, og resultatene er åpent tilgjengelig for allmennheten via en nettside der alle rapporten kan lastes ned. Forholdene er dermed lagt til rette for at resultatene blir brukt. Det vil altså fremdeles ta noen år før man kan vurdere læringseffekten av statens ordning, men både forvaltningsmessig men ikke minst evaluingsfaglig er dette et interessant eksperiment som forsøker å sette evaluering i system.

Evaluering er ikke lenger rituelt men reelt. Målet er ganske enkelt å trekke ut erfaringer i etterkant og bruke dette i forkant for å forbedre nye tiltak. Logikken er enkel. Det blir spennende å se om det virker.

## Kilder

Argyris C. og Schon, D. (1978). *Organizational Learning*. San Fransisco: Jossey-Bass.

Askim J, Døving E, Johnsen Å, (2013). Evalueringspraksis i den norske statsforvaltningen 2005-2011, i Halvorsen A, Lier Madsen E, Jentoft N (red.) *Evaluering. Tradisjoner, praksis, mangfold*, Fagbokforlaget.

Samset, K., Forss, K. og Hauglin, O. (1993). *Learning from experience. A study of the feedback from evaluations and reviews in Norwegian bilateral aid*. The Norwegian Ministry of Foreign Affairs, Evaluation report no. 1.93.